The Use Of Raman Spectroscopy And Microfluidics In Determining The Biochemical Progression From Barrett's Oesophagus To Oesophageal Adenocarcinoma

Alisha Farooq

Submitted in accordance with the requirements for the degree of *Doctor of Philosophy*

November 2024

School of Molecular and Cellular Biology

School of Physics and Astronomy

University of Leeds

I dedicate this work to my beloved grandparents.

There isn't a single day that I don't acknowledge how blessed I've been to have you, and how much I've achieved through your prayers. Abu, if love could've saved you, you'd have lived forever. I can't wait to reunite in Jannah and tell you how your "Aaliya" finally got her doctorate. InShaAllah.

میں یہ کام اپنے پیارے نانا نانی کے نام کرتا ہوں۔
ایک دن بھی ایسا نہیں گزرتا جب میں یہ نہ سوچوں کہ
میں آپ کے ہونے پر کتنا شکرگزار ہوں اور آپ کی دعاؤں کی بدولت
میں نے کتنا کچھ حاصل کیا ہے۔
ابو، اگر محبت آپ کو بچا سکتی تو آپ ہمیشہ زندہ رہتے۔
میں جنت میں دوبارہ ملنے کا بے حد انتظار کر رہی ہوں تاکہ آپ کو بتا سکوں
کہ آپ کی "عالیہ" نے بالآخر اپنی ڈاکٹریٹ مکمل کر لی۔
ان شاء اللہ۔

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The work in Chapter 3 (page 74) of this thesis was published in Scientific Reports [*Farooq, A., Wood, C.D., Ladbury, J.E. et al. On-chip Raman spectroscopy of live single cells for the staging of oesophageal adenocarcinoma progression. Sci Rep 14, 1761 (2024). https://doi.org/10.1038/s41598-024-52079-3*]. I was responsible for the cell preparation, all data acquisition, Raman spectra analysis, multivariate analysis for PCA/LDA, and generation of the manuscript. Stephen D. Evans was responsible for writing the codes for Raman spectral manipulation and PCA/LDA. Stephen D. Evans, John E. Ladbury and Christopher D. Wood contributed with data discussion, manuscript corrections and final review.

The work presented in Chapter 5 (page 108) was done in collaboration with Mr Henry Aldridge, who was responsible for the writing and processing of the LDA code and CNN analysis.

Acknowledgements

I'd like to begin by thanking my supervisors Professor John Ladbury and Dr Christopher Wood. I am deeply grateful for the guidance, support, and expertise you have generously provided throughout my doctoral journey. To Professor Stephen Evans, thank you for being the most exceptional supervisor. For always finding time in your incredibly busy schedule for me to laugh or cry (sometimes both) about my science, and for sending me across the planet to yap about it to others. Your support and guidance has been invaluable, and I'm very blessed to have been moulded as a scientist in such an encouraging environment.

Since I'm already thanking the "adults", I'd like to give a shoutout to Dr Ben Johnson for tussling with IT on my behalf, and for not kicking me out of his office during my spontaneous crying moments (#3 on the list of best places to cry in Bragg btw). I'm pretty sure you've completed the hours to qualify as a licensed therapist by my departure. A big thank you to Dr Julia Gala de Pablo, for taking the time to answer my dumbdumb questions re. Raman, and for allowing me to use and adapt her analysis code for my data. The warmest regards to my office grandad, the late Professor Richard Bushby. For his constant pearls of wisdom, his never-ending enthusiasm for science, and for filling a personal and emotional void I didn't realise I held at the time. I'm saddened by his absence, but also grateful that I had the opportunity to say a proper goodbye.

I'd like to thank all my MNPeers, I can't wait to hear about how productive you've all been now the office noise is restored back to a reasonable decibel. Big up the Evans Group, thank you for being the most entertaining office family. To my desi sis Anjali, for encouraging my many unnecessary side quests, and for making daily life tolerable. To Damien, for his patience with my endless barrage of questions and for our mutual love of all things Princess Diana, Britney Spears and Shrek (very much wishing I could add cows to this list). To Foxy Joe, for his humour. To Aileen, for showing me that there is indeed light at the end of the seemingly endless tunnel #RamanSurvivors. To Nikki, for her support not only with the final stages of my PhD but also all my post-PhD mania, I owe you a big fat lemon cake!

To Benard, for reminding me to stay hydrated, even whilst residing in the southern hemisphere. To Kalila, for affirming that I am special but *in a good way.* To Holly, for the potato lore and talking sense when I most needed it. To Christa, the other Brown, for taking the time to give feedback on this thesis despite living 7 timezones away. To Bioelec, for informally adopting me as part of their group in both office and social situations. A special mention to Joel's receding hairline; thesis writing and viva preparation were much less stressful knowing that if those three hairs could survive this hardship, so could I.

To my non-academic friends!! Thank you for understanding my ridiculously busy schedule and not holding it against me. There are absolutely too many of you to name individually but I've missed you all, prepare to be absolutely sick of me now that the concept of free time exists in my life again. To Zainab, I genuinely wouldn't have made it through to the end without you, there are no words, I love and appreciate you always. My older sis Nayab, for bringing the true essence of Pakistan (Southall Travel WHO?). To Dionne, almost 15 years of friendship and higher education later, I can't wait to clear my schedule and spend our

weekends back together. To Josie, for keeping me sane throughout the writing stage, and for entertaining my post-thesis plans knowing fully well that I had £2 to my name and no employment lined up.

The biggest thank you to my family, who have supported me throughout this ridiculously long period of studying, and multiple life crisis', I love you all <3 A special mention to Khala Baj and Zeenat for being so understanding and tolerant, especially on the days I was going through it (every day). I can imagine the relief from no longer trying to explain what it is that I'm studying when people ask.

This thesis marks the end of a long and challenging journey, but it also stands as a testament to the love, guidance, and unwavering support I've been lucky enough to receive over these past few years. Here's to hoping all future endeavours will be a piece of cake (preferably chocolate).

Abstract

Oesophageal adenocarcinoma (OAC) is a highly aggressive cancer, often arising from Barrett's oesophagus (BO), a condition where the normal squamous epithelium is replaced by intestinal-like columnar epithelium due to chronic acid and bile reflux. Disease progression typically follows a continuum from non-dysplastic BO to dysplasia, eventually leading to OAC. Identifying the biochemical and mechanical changes at each stage is crucial for early diagnosis and intervention. This study utilised advanced single-cell techniques, including Raman spectroscopy and deformation cytometry, to characterise these changes and develop novel diagnostic tools.

Single-cell Raman spectroscopy, a non-invasive biophysical technique, was applied to detect biochemical signatures in healthy oesophageal tissue, non-dysplastic and dysplastic BO, and OAC. This revealed distinct spectral features linked to lipid, protein, and nucleic acid content, reflecting cellular changes during disease progression. For instance, cancerous cells exhibited increased nucleic acid content, indicating higher metabolic activity, while dysplastic cells displayed early alterations in lipid profiles. These biochemical variations highlight potential biomarkers for early detection of oesophageal malignancies.

To complement the biochemical profiling, deformation cytometry was used to measure the mechanical properties of individual cells with microfluidic devices. This revealed significant decreases in cell stiffness in dysplastic and cancerous cells compared to healthy and non-dysplastic BO cells, likely due to cytoskeletal remodelling, a hallmark of tumour progression and metastasis. By integrating Raman spectral data with mechanical measurements, this study created a multi-modal framework capable of accurately distinguishing disease stages.

The microfluidic platforms facilitated high-throughput single-cell analysis, enabling rapid, automated measurement of biochemical and mechanical properties while improving reproducibility. This system also allowed real-time monitoring of cellular changes, demonstrating potential applications for time-sensitive studies like drug response evaluations. Together, these findings highlight the promise of combining Raman spectroscopy and deformation cytometry as diagnostic tools for early cancer detection and clinical applications.

Abbreviations

A.I. (Artificial Intelligence)

ABSS (Acidic Bile Salt Solution)

BO (Barrett's Oesophagus)

CNN (Convolutional Neutral Network Analysis)

D.I. (Deformation Index)

DC (Deformation Cytometry)

EMSC (Extended Multivariate Signal Correction)

GORD (Gastro-Oesophageal Reflex Disease)

HDI (Human Development Index)

LDA (Linear Discriminant Analysis)

ML (Machine Learning)

OAC (Oesophageal Adenocarcinoma)

OSCC (Oesophageal Squamous Cell Carcinoma)

PC (Principal Component)

PCA (Principal Component Analysis)

PDMS (Polydimethylsiloxane)

PPIs (Protein Pump Inhibitors)

RS (Raman Spectroscopy)

RT – DC (Real-Time Deformation Cytometry)

SP (Stagnation Point)

List of Figures

List of Tables

Table of Contents

1 Introduction

1.1 Overview

Oesophageal cancer is the seventh leading cause of cancer death worldwide, accounting for just under half a million deaths annually. The most recent GLOBOCAN statistics estimate around 511,000 new cases every year, with 2040 estimations predicting an increase in both incidence cases to (957,000), and patient mortality (880,000) (14). Countries with a lower human development index (HDI) are burdened with 80 % of cases, whilst 70 % of diagnosis develop in men; a 2-to-3 fold difference in both mortality and incidence rate between the sexes (15). This cancer burden is increasing rapidly on a global scale, with two predominant histological subtypes: oesophageal squamous cell carcinoma (OSCC) and oesophageal adenocarcinoma (OAC).

OSCC accounts for 85 % of global cases, developing through the malignant transformation of the oesophageal squamous epithelium, with 79.7 % of new cases detected in Asia in 2020. China, Japan, Pakistan, India and Bangladesh were amongst the most affected, with the incidence in China almost five times higher than in America (16). Despite this, the incidence for OAC has now surpassed that of OSCC in higher HDI countries following increases in age-standardised incidence (17), representing two thirds of cancer diagnosis'. With OAC incidence rates rising, key contributors for cancer development are gastroesophageal reflux disease (GORD), Barrett's oesophagus (BO) and excess body weight (17, 18). Unfortunately, the five-year survival rate for both subtypes remains discouraging, at less than 20 % for patients at advanced stages (19). Accompanied by global aging and population growth, GORD affects 10-30 % of the adult population in HDI countries, with higher prevalence with Northern Europe and 5 in 1000 people affected in the UK (20, 21). Pre-cancerous condition BO currently affects less than 1 % of the UK population, with 3-13 % of these patients going on to develop oesophageal adenocarcinoma. Whilst incidence figures appear low, mortality rates are steadily increasing (22, 23).

Advanced endoscopic techniques have the potential to detect the progression of pre-malignant lesions to cancer and vastly enhance patient outcomes. However, this effectiveness is hindered by low endoscopy rates due to its invasive nature. Despite extensive

research on cancer risk predictors, a large number of cases remain undiagnosed, highlighting the requirement for improved screening techniques at earlier stages of disease progression. Cost-effective, minimally invasive, and well-tolerated non-endoscopic technologies are crucial. Implementing such devices in primary care settings within high-risk populations should aid detecting in early disease progression and highlighting patients at risk for cancer development. The use of modernised minimally invasive techniques and faster detection will in turn provide a substantial improvement in both morbidity and mortality rates. A comprehensive understanding of each stage of disease progression is important for delivering these improvements in prevention, with an emphasis on minimal costs and intrusion, to give maximum throughput and accuracy.

1.1.2 Prevalence and Risk Factors

The *'OAC epidemic'* (24) has seen soaring incidence rates since the 1950s, with the highest incidence rates seen in Northern Europe, particularly the UK and the Netherlands (17, 25). Accompanied by global aging and population growth, incidence rates for oesophageal cancer have risen by 38,000 since 2017 (15). The primary risk factors for OAC are closely linked to the disease progression from BO. GORD is currently the most significant risk factor for OAC, with rising cancer incidence reflective of increasing GORD prevalence, as shown in Figure 1.1. A systematic review and meta-analysis of five population-based studies demonstrated that experiencing GORD symptoms at least weekly elevates the risk of OAC nearly five-fold (OR 4.92, 95% CI 3.90–6.22), and daily symptoms increase the risk approximately seven-fold (OR 7.40, 95% CI 4.94–11.1) (26). Nonetheless, 40 % of OAC patients do not report GORD symptoms (27).

Obesity is another significant risk factor for both GORD and BO, contributing to the rising incidence of OAC, which parallels the growing prevalence of obesity. A systematic review and meta-analysis of six studies conducted through March 2013 found that central obesity is associated with an increased risk of OAC (OR 2.51, 95% CI 1.54–4.06) (28). Tobacco smoking also correlates with a higher risk of OAC. A pooled analysis of ten population-based case-control studies and two cohort studies, comprising 2990 OAC cases and 9453 controls, indicated that smoking doubles the risk of OAC (OR 2.08, 95% CI 1.83–2.37) with a strong dose-response relationship linked to the number of pack-years of smoking (29).

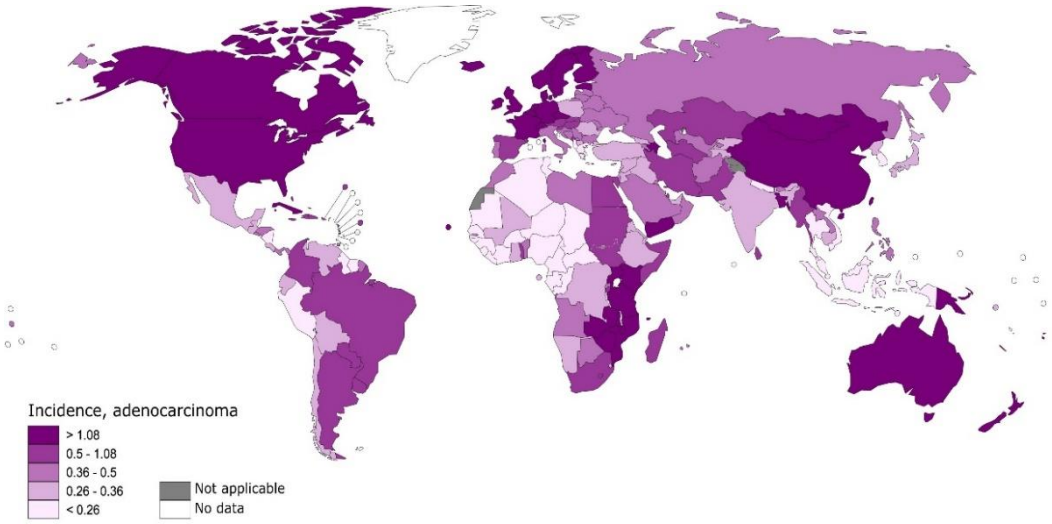

Figure 1.1: Worldwide oesophageal adenocarcinoma rates (age-adjusted according to the world standard population, per 100,000) in 2020. Reprinted with permission (10)

Dissimilar to GORD, OAC exhibits a notable male predominance, with a global male-to-female ratio of 4.4, ranging from 1.7 to 8.5 (30). This disparity cannot be fully attributed to higher BO rates in men, as suggested by a meta-analysis of the sex ratio for BO (31). Possible explanations include greater severity of GORD due to increased abdominal obesity and higher smoking rates among men, as well as potential contributions from sex hormone factors (32).

Contrasting to its role in OSCC, alcohol consumption is not a significant risk factor in the progression from GORD to BO to OAC, although some studies indicate a weak association. A meta-analysis of 24 studies, including 5500 OAC cases, found no substantial link between alcohol consumption and OAC risk, even at higher consumption levels (33). Additionally, infection with Helicobacter pylori appears to reduce the risk of OAC. A systematic review and meta-analysis of studies up to February 2007 found an inverse relationship between Helicobacter pylori prevalence and OAC risk (pooled OR 0.52, 95% CI 0.37–0.73) (34).

In 2021, > 7800 people in England were diagnosed with oesophageal cancer, predominantly affecting those aged 70 to 74, with 1,054 diagnoses in men and 371 in women. This distribution is illustrated below in Figure 1.2. 13 % of men and 10 % of women in England were regular smokers, and many men exceeded the recommended weekly alcohol intake of

14 units. In 2018/19, 30 % of women and 25 % of men in England reported eating at least five portions of fruit and vegetables per day, while 9 % of men and 7 % of women consumed none (35).

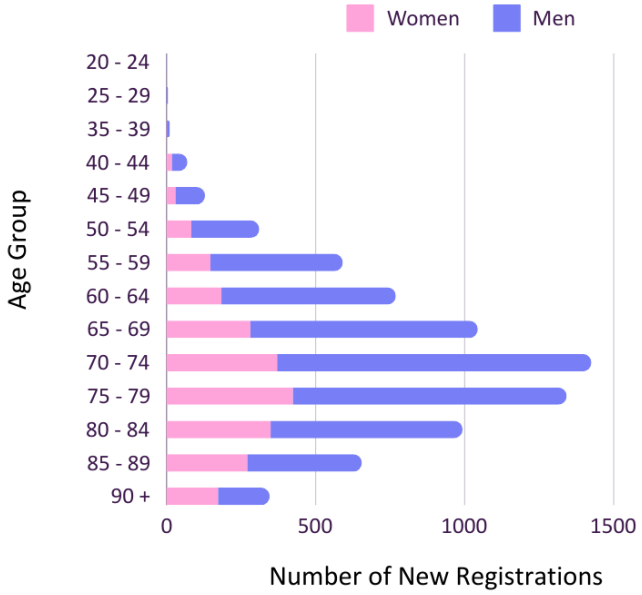

Figure 1.2: Registrations of newly diagnosed cases of oesophageal cancer in England by 2021, by age group and gender. Reproduced with permission from the Office for National Statistics (UK). © Stastica 2024

Assessing the exact prevalence of BO in different populations is challenging due to the asymptomatic nature of the condition, which often remains undiagnosed until an endoscopy is performed. Consequently, the actual prevalence may be significantly higher than currently predicted. In European studies, BO is found in approximately 1 % to 2 % of the unselected general population, whilst in the USA, the prevalence is around 5 % to 6 %. However, amongst individuals with GORD, the prevalence can rise to as high as 15 % (36). A Swedish population-based study from the early 2000s found BO with intestinal metaplasia in 1.6% of 3,000 randomly selected individuals (95% CI 0.8–2.4) (37). An Italian study from 2000 to 2004 reported a prevalence of 1.3% in 1,033 individuals (38). A systematic review and meta-analysis of 51 studies from Asian countries up to September 2014 found a pooled prevalence of 1.3% (95% CI 0.7–2.2) and noted an increasing trend over time from 1991 to 2014 (39).

BO is more common in men and is notably more prevalent in Caucasians compared to those of African ancestry (40). The condition also becomes more frequent with age and is more

likely to affect individuals with a family history of the condition. UK studies indicate that the prevalence increases by 7.4 % for each additional year of age between 20 and 59 years in males, with a similar pattern observed in females, though with a 20-year delay. It is important to note that the development of metaplasia to columnar epithelium does not directly correlate with the degree of oesophageal acid reflux, suggesting that other genetic and environmental factors play a significant role (36).

Risk factors associated with progression to adenocarcinoma include male gender, advanced age, extensive segment involvement (greater than 8 cm), presence of intestinal metaplasia, prolonged history and/or early onset of GORD, duodeno-gastro-oesophageal reflux, mucosal damage (such as ulceration and stricture), and a family history of the condition (41, 42). Sex-specific differences in the progression of BO may be influenced by hormonal factors. Evidence suggests that males develop BO earlier than females, resulting in a longer duration of risk for progression (43). Familial clustering of both BO and oesophageal adenocarcinoma (OAC) has been observed, with linkage and genome-wide association studies indicating a complex genetic trait (44, 45). While no monogenic mutations have been identified for BO or OAC, the heritable components are estimated at 35 % and 25 %, respectively (46). The characteristics of the BO lesion, such as longer lesion length and the progression from non-dysplastic BO to low-grade dysplasia, also influence the likelihood of developing OAC (47).

Characterised by the reflux of acidic and bile salt-containing duodenal and gastric contents, GORD is a primary factor in the development of the initial metaplastic BO lesion and is among the strongest risk factors for OAC (27, 48, 49). 10 – 15 % of patients suffering from GORD will develop BO, with 1 – 10 % of those BO patients developing OAC over a period of 10 – 20 years. The vast majority of BO patients die of natural causes (50, 51). A 2021 systematic review and meta-analysis of 44 cross-sectional studies revealed that 7.2% (95% CI, 5.4%-9.3%) of 26,521 individuals with GORD had histologically confirmed BO (52). There is greater progression towards OAC with long-segment BO (> 3cm) than short-segment BO (< 3cm). A meta-analysis of 4097 BO patients, who weren't presenting with dysplasia, reported annual OAC progression rates of 0.06% (95% CI, 0.01%-0.10%) for short-segment BO and 0.31% (95% CI, 0.21%-0.40%) for long-segment BO (53). GORD evaluation via upper endoscopy was analysed for 378 patients as part of a cohort study, with short-segment BO having a higher prevalence (8.5 %) than long-segment BO (4.8 %) (54). This outcome has also been reported in other studies (53, 55).

Evidence suggests that GORD significantly promotes dysplastic progression, especially in BO lesions located in the right oesophageal hemisphere, which is most commonly exposed to gastric refluxate (56, 57). Additionally, medications that relax the lower oesophageal sphincter, increasing the frequency and severity of gastric reflux, are associated with the development of OAC (47, 58-63). These medications include anti-cholinergic drugs, calcium channel blockers, and theophylline. Proton pump inhibitors (PPIs) have shown a dose-response relationship in preventing dysplastic BO progression (64), with high doses being more protective when combined with aspirin. Patients with BO often exhibit elevated basal acid production, while Helicobacter pylori infection may reduce the risk of OAC by decreasing gastric acid secretion (65-67). Studies indicate that 36 % of OAC cases are directly linked to frequent symptomatic acid reflux, with daily GORD symptoms increasing the risk of OAC seven-fold compared to less frequent symptoms (68, 69), with a significant proportion of OAC cases likely arise in individuals with silent reflux disease (70).

1.1.3 Clinical Characteristics

The majority of patients with BO have a known history of GORD, with symptoms such as heartburn after eating and acid regurgitation. Less common symptoms include dysphagia and the sensation of obstruction within the throat when there is none. Rarely, patients may present as asymptomatic however most patients will have GORD symptoms with no other specific physical examination findings (71).

Unselected population screening for BO is not recommended due to the relatively low risk (1 – 2 %) (52, 54, 72-75) with an annual risk of progression to OAC at 0.3 – 0.8 % (76-78). The cost-effectiveness of BO screening programmes has therefore been disputed (79-81), with suggestions that it only be offered to high-risk populations. A recent systematic review and analysis (82) assessed the prevalence of BO in individuals with known risk factors. Prevalence was low (0.8 %) for patients not exhibiting GORD symptoms. Higher prevalence was detected in patients presenting known risk factors for BO such as a family history of BO/OAC (23 %), male sex (6.8 %), age > 50 years (6.1 %), GORD (2.3 %), and obesity (1.9 %). Individuals with GORD and one other risk factor had an overall increased risk of 14.5 %, with each additional risk factor increasing BO prevalence by 1.2 %, shown as a positive linear relationship. The minority of patients who don't suffer from GORD (40 – 50 %) (83, 84) would be missed when

adhering to current suggested screening strategies, all which require GORD symptoms as an pre-requisite, as illustrated in Figure 1.3.

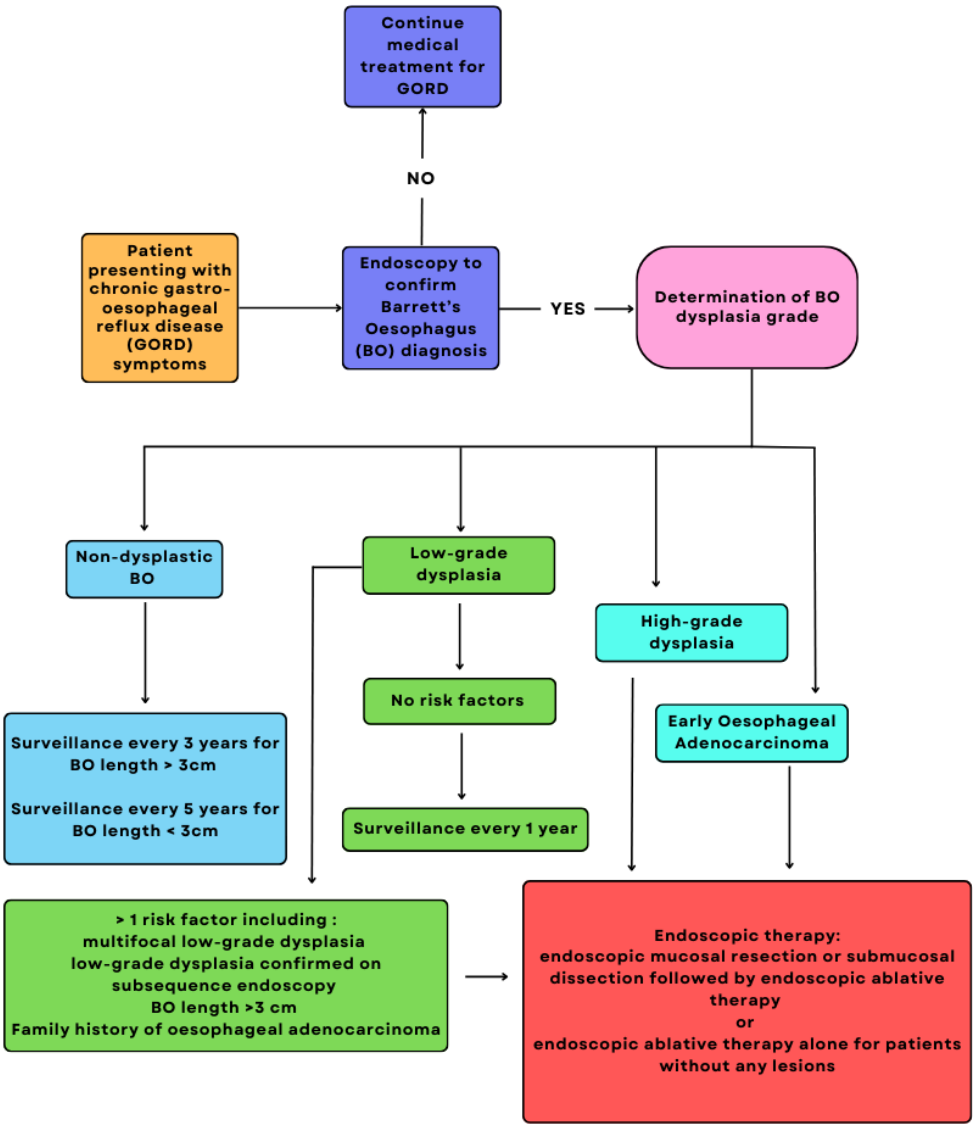

Figure 1.3: Treatment algorithm for patients with Barrett's oesophagus.

1.1.4 Diagnostic Challenges

Despite modest improvements over recent decades, survival rates for OAC remain poor. Overall, less than 20 % of patients survive for five years, and approximately half succumb to their cancer within a year of diagnosis (58, 85). A recent analysis using USA SEER data showed that 73 % of patients with stage 0 and 37 % with stage 1 disease were alive at five years (58).

Strategies promoting the prevention and early diagnosis of OAC are essential for improving overall outcomes. These include screening to detect patients with BO, surveillance to identify those at risk of or demonstrating progression, effective chemoprevention to limit dysplastic progression, and effective treatments for cases where dysplasia or invasive disease has occurred. It currently remains unclear whether community surveillance programs translate to improved survival (83, 86-88). This uncertainty may be partly due to a lack of sensitive and specific biomarkers, which hampers current screening and surveillance approaches from identifying patients most at risk of neoplastic progression (64). Intensive and often invasive follow-up regimes are required, placing pressure on healthcare resources and exposing patients to potential harm. Chemo-preventive options for BO are limited beyond anti-reflux treatments, with no definitive chemotherapeutic option for patients who progress to dysplasia. In such cases, invasive treatments such as ablative therapy or endoscopic or surgical resection are generally required (84).

The cost-effectiveness of screening programmes may be improved by using prediction tools which incorporate multiple risk factors to select patients for screening. This strategy would still require patients to undergo time-consuming invasive endoscopies for the collection of multiple Barrett's lesions for analysis. New screening modalities that would reduce both surgery time and required sample volume would be more cost-effective than standard patient sedated endoscopy. With the implementation of such modalities, the prospects of BO / OAC screening may be expanded in the near future.

1.2 Development of the Barrett's Lesion

1.2.1 The Oesophageal Barrier

The formation of the initial BO lesion results directly from chronic exposure of the lower oesophagus to gastric refluxate (9,22,115), comprising of hydrochloric acid-rich gastric acid, bile salts, ingested foods, bacteria, proteolytic pancreatic enzymes, and pepsin (116). The lower oesophageal sphincter, along with the diaphragmatic sphincter and the phrenoesophageal ligament, create a high-pressure zone in the lower oesophageal region. In conjunction with gastric emptying and oesophageal peristaltic movement, this increased pressure minimises the frequency and duration of refluxate exposure to the oesophageal mucosa (117,118). Despite this, reflux episodes are typically common in healthy populations, with a recent ambulatory impedance-pH study reporting a median of 44 episodes per 24 hours (119).

To minimise contact with harmful gastric stimuli, the oesophageal epithelium must maintain a protective barrier, as illustrated in Figure 1.4. This barrier is created through the combined actions of multiple layers: a pre-epithelial (luminal), an intrinsic epithelial, and a post-epithelial (stromal) compartment (120). A mucous buffer layer originates from within the lumen, formed using bicarbonate, water and mucous from swallowed saliva, produced by the submucosal glands (118, 121). This mechanism neutralises hydrogen (H^+) ions and prevents diffusion of harmful stimuli to the apical cellular membrane, with additional protection via a layer of unstirred water flowing laminar along the apical epithelial surface (120,122).

The swallowed saliva also contains various tissue repair mediators, including transforming growth factor α (TGF-α), prostaglandin E_2 (PGE_2), and epidermal growth factor (EGF) (123). These defences are less effective than their gastric counterparts, leaving the oesophageal mucosa more vulnerable to damage (120). The intrinsic non-squamous epithelial barrier plays a crucial role in preventing harmful molecules from penetrating deeper mucosal layers. Key to this function are the intracellular glycocalyx and the apical junction complex, which includes tight junctions, adherens junctions and desmosomes (120). The transmembrane proteins claudin and occludin are essential for maintaining the integrity of the epithelial barrier (124,125). Junctional adhesion molecules, which support occludin assembly, and the

late epidermal differentiation protein filaggrin, which helps maintain the intercellular glycoprotein matrix, are also important (126,127). H^+ ions that bypass these defences are neutralised by a neutral intercellular glycoconjugate or removed through diffusion into the bloodstream, which increases in response to mucosal acid exposure. This hyperaemia also facilitates the removal of lactic acid and carbon dioxide, and delivers bicarbonate ions to neutralise acid (128,129).

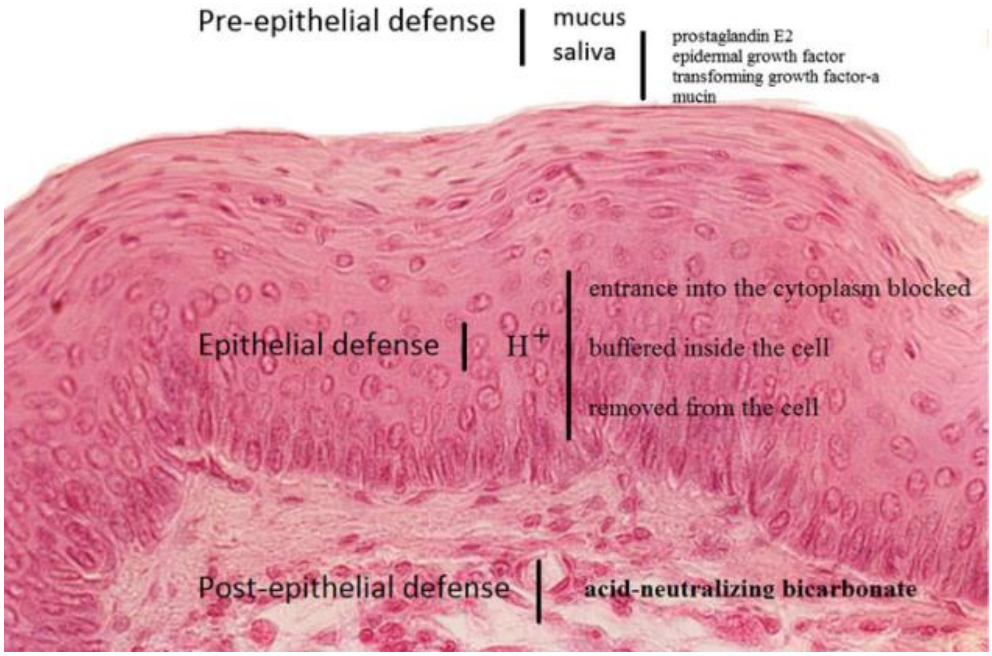

Figure 1.4: Illustration detailing the protective mechanism of the oesophagus against noxious refluxate H+ ions. Reprinted with permission. (4)

Despite these protective mechanisms, prolonged exposure to harmful stimuli results in damage to tissues and cell death. Previous research has determined that the resulting cellular injury and necrosis is a result of cytokine-mediated inflammatory processes (130). Initiated during damage to the apical junction complex, this results in increased para-cellular permeability, shown morphologically via the presence of dilated intercellular spaces. Consequently, this leads to a greater H^+ concentration along basolateral epithelial cell membranes, affecting their sodium-independent chloride/bicarbonate exchange (133). Acute and chronic inflammation arises due to the build-up of intercellular acidification, concluding with DNA damage and oxidative stress (120,132). Whilst regeneration of

damaged cells is possible (135), the epithelium is able to repair itself through the migration of healthy adjacent cells to the damaged area within 30-60 minutes of initial injury (134), as detailed in Figure 1.5.

Increased intragastric pressure is a feature of obese GORD patients, resulting in an increased gastro-oesophageal pressure gradient and transient relaxation of the lower oesophageal sphincter, thereby leading the gastric contents to flood the lower oesophagus. This also occurs when intra-abdominal pressure is increased. Additionally, obesity enhances the spatial separation of the lower oesophageal sphincter and the crural diaphragm, predisposing individuals to a hiatal hernia, resulting in increased exposure of the oesophagus to acid (7).

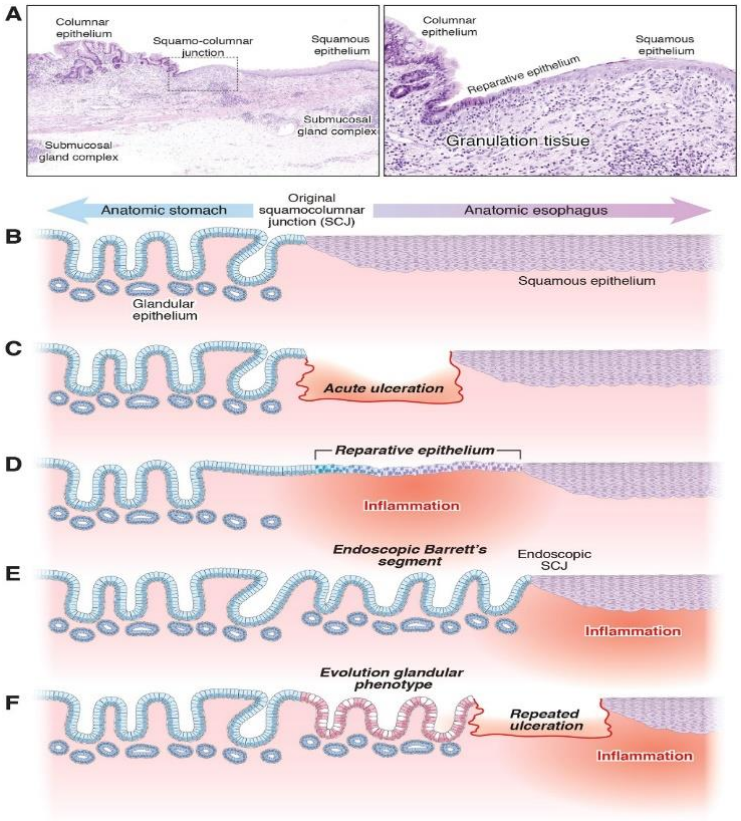

Figure 1.5: Models of wounding and competitive replacement in Barrett's oesophagus development. (A) The squamo-columnar junction in a Barrett's oesophagus segment shows granulation indicative of recent epithelial denudation. (B) This junction aligns with the stomach-oesophagus boundary. (C) Chronic acid-biliary reflux erodes squamous epithelium, forming an ulcerative defect. (D) The defect is covered by undifferentiated epithelium. (E) Progenitor cells mature into mucous glands, which (F) differentiate into intestinal or gastric types based on the microenvironment. Recurrent reflux and ulceration lead to progressive replacement of squamous epithelium by columnar epithelium. Reproduced with permission (13)

1.2.2 Molecular Origins of Barrett's Oesophagus

The development of oesophageal cancer has a large number of mutations akin to cancers such as melanoma and lung cancer. Non-dysplastic BO has a remarkably lower number of mutations in contrast to oesophageal cancer; whilst having a larger number of mutations than other malignancies such as breast cancer and multiple myeloma (89). The mutation, inactivation, or loss of regulatory genes are responsible for controlling the cell cycle, which leads to changes in the oesophageal lining, into BO (90). The most important tumour suppressor gene in BO is p16, also known as CDKN2A, with 80 % of BO associated with p16 anomalies (91). Inactivation of this gene results in genomic instability and resultant uncontrolled cell multiplication. The alteration of p53 can be found in both high-grade BO and OAC, however it is not a feature in low-grade BO. Both p16 and p53 affect the G1 to S-phase of the cell cycle, through either induction of apoptosis or blocking cell division processes should DNA repairs fail. Faults in both genes lead to a lack of cell death, excess cell division and progression to malignancy.

Epigenetic governors such as Homeobox [Hox] genes code for DNA-attached proteins which determine the fate of various cell types. Caudal-type Hox [Chx] -1 and -2 have been linked with to the transformation of cell architecture from squamous to columnar cells (92), heralding the development of metaplasia in the oesophagus (93) via hyper-methylation of promoter sequences on tumour suppressor genes, leading to the inactivation of the latter. Other cell cycle regulators include tumour suppressor genes such as adenomatous polyposis coli gene involved in cell cycle control, and proto-oncogenes such as Myc genes.

Significant alterations occur in gene regulation and protein expression during the metaplastic process. This involves the loss, diminished function, or downregulation of transcription factors critical for the normal development of oesophageal squamous epithelium, such as tumour protein 63 [p63], transcription factor SOX2, and paired box protein Pax-9 [PAX9]. Concurrently, there is an upregulation of transcription factors linked to intestinal development, including SOX9, nuclear factor kappa-light-chain-enhancer of activated B cells [NF-κB], GATA binding protein 6 [GATA6], homeobox proteins CDX1 and CDX2, protein analog homolog 1 [ATOH1], hypoxia-inducible factors [HIFs], mini-chromosome maintenance protein (MCM2) (Figure 1.6), and hepatocyte nuclear factors [HNF] 1α, 3α, 3β, 3γ, and 4α (70, 94-97).

Upstream of these factors, the combination of pro-inflammatory cytokines produced from the lower oesophageal stroma and the GORD-induced inflammatory environment leads to activation in numerous different pathways (98, 99). These include signalling cascades such as the Hedgehog signalling pathway, the highly-conserved Wnt signalling pathway, and the TGFβ superfamily member bone morphogenetic protein BM4 (100-106). Downregulated pathways associated with both transdifferentiation and transcommitment include the Notch signalling cascade (107-109).

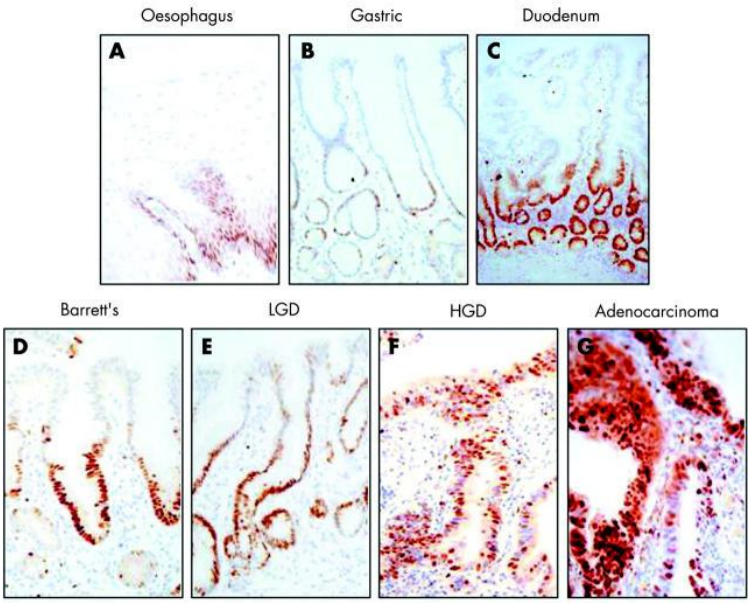

Figure 1.6: Abnormal proliferation compartments of Barrett's oesophagus and associated dysplasia. Expression of a proliferation marker mini-chromosome maintenance protein (Mcm2) is shown by immunohistochemistry. Whereas the proliferative compartment is confined to the basal layers (normal squamous oesophagus, (A)) and in the crypts and the glands (normal stomach and duodenum, (B) and (C)), in Barrett's metaplasia proliferation extends towards the surface (D). With increasing dysplasia the number of proliferating cells increases with expansion of the proliferative compartment (E-G). Reproduced with permission from BMJ Publishing Group Ltd (9).

1.2.3 Transdifferentiation, Translocation & Transcommitment

The cellular origin of BO is debated with controversial and conflicting opinions, partly due to the inability to observe the metaplastic process *in vivo* and the lack of reliable physiological animal models. The initial hypothesis was that BO arose as a consequence of migrating transitional zone cells from the gastro-oesophageal junction to the distal oesophagus and gastric cardia as a response to acid refluxate-induced tissue damage (110). However, even if this were to be the case, the differing epithelial cell lineages present in BO still need to be accounted for.

When studying the oesophageal and gastric cardiac mucosae distribution across oesophagectomy specimens, it was inferred that exposure of the cardiac glands to the luminal surface could result in their transition to columnar epithelial islands which could give rise to BO through clonal expansion (111). The theory of upward migration from the gastro-oesophageal junction was discarded after animal studies revealed the cell of origin was intrinsic to the oesophagus (112).

The predominant theories of the last few decades suggest that the metaplastic epithelium arises from transdifferentiation; a direct conversion of squamous cells to columnar cells. Transdifferentiation is defined in this context as the irreversible metaplastic conversion from one fully differentiated state into another. This is supported by previous research detailing the conversion of murine epithelium from columnar to stratified squamous, during embryonic oesophagus development. This correlated with the co-expression of squamous (cytokeratin 14) and columnar (cytokeratin 8) differentiation markers during the conversion, suggesting that these changes can occur independent of cell division or apoptosis (113).

In theory, the reverse transformation could explain the phenotypic change seen in BO, although this switch is yet to be determined in adulthood. Additionally, evidence of newly developed squamous epithelium post-BO endoscopic removal (provided all differentiated epithelium was removed) would further weaken this hypothesis (114). Another related hypothesis suggests that there may be an intermediate “multi-layered epithelium”; composed of multiple cellular layers, appearing squamous in their basal portion and columnar more superficially, with both layers expressing cytokeratins (115).

Despite this, recent data has strengthened the argument that the upward migration of columnar cells from the gastro-oesophageal junction leads to BO development (116). Examination of early oesophageal development in wild-type mouse embryos led to suggestion that the columnar cells lining of the embryonic oesophagus, composed of columnar cells, was sloughed off as squamous epithelial cells migrate through the oesophagus, undermining and replacing the columnar cells (117). They hypothesised that the squamo-columnar junction in adult mice, and the gastro-oesophageal junction in human adults, contain a residual population of embryonic columnar cells. These cells later give rise to BO via migration after refluxate damage to the proximal squamous epithelium (117). The concept of a oesophagus-gastric junction cell of origin may also explain the origin of gastric-type columnar metaplasia in the distal oesophagus, however the debate as to whether these are linked or separate entities is ongoing (118). This indicates that there are multiple cells of origin for BO, and that the progression from BO to OAC is linked to the particular cell type that BO developed from.

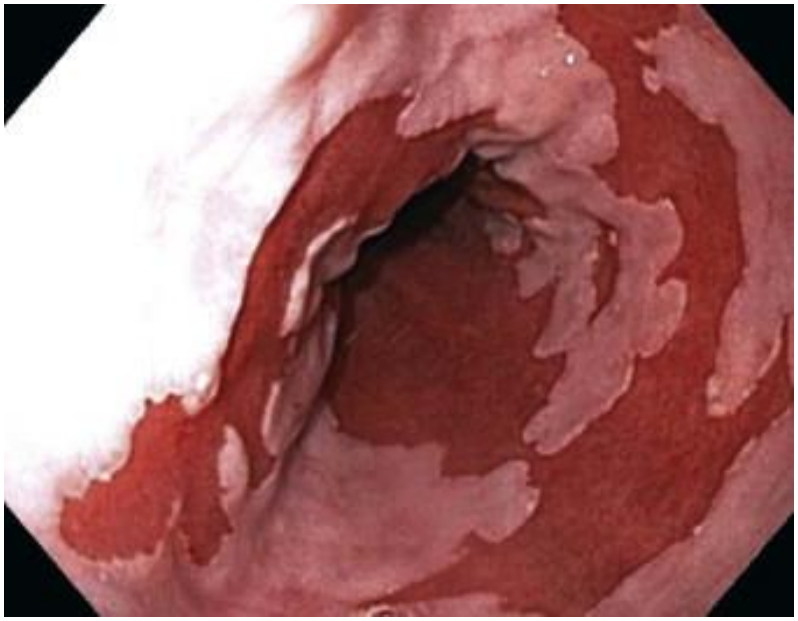

Figure 1.7: Endoscopic appearance of Barrett's oesophagus. Note the white- appearing normal squamous mucosa displaced above the true end of the oesophagus. The intervening mucosa appears salmon-pink. Reposted with permission (7)

Alternate theories include changes in the commitment (transcommitment) of pluripotent stems cells responsible for refreshing of the oesophageal epithelium, leading to metaplastic development (119). The oesophagus has multiple sites from which these cells can arise,

including the sub-mucosal glands and the lining of their ducts, and the stratified squamous basal layer epithelium. This hypothesis suggests an induced change in the commitment of multipotent stem cells due to environmental stressors such as acidic refluxate. Whilst correlating with the evidence that the BO cell(s) of origin are intrinsic to the oesophagus, it would also explain the various cell phenotypes present in BO, as well as the regeneration of squamous epithelium post-BO lesion removal (114).

The first stem cell population to be involved in this theory are the squamous epithelial cells in the basal layer. Observation of highly characterised squamous epithelium in the interfollicular epidermis, which contains features similar to the oesophagus, saw the presence of multiple different cell linages within adult tissue, from a multipotent stem cell population (120). The glandular neck region of the oesophageal submucosal gland ducts has been suggested to contain a second stem cell population, possibly responsible for the origin of the columnar epithelium (121). This prefigures a migration of surface glandular cells, derived from new ducts and glands via stem cells in the lamina propia, similar to ulceration observed in the gastrointestinal tract due to the near-residing ulcer-associated cells (122).

Retinoic acid has previously been adapted as a stimulant for squamous cell differentiation, with the results suggesting that the submucosal gland ducts within the stromal department provide the columnar cell source (123). Bone marrow-derived progenitor cells have also been hypothesised as possibly contributing to BO development (124). There may be more than one route to BO development (70, 125), however, given the presence of multiple cell linages, transcommitment can be considered the most likely.

1.2.4 Adenocarcinoma Progression

The genomic progression of BO to OAC has a baseline association of increased clonal diversity reflective of high levels of multi-generational chromosomal instability (126, 127). Associated with mitotic slippage, this is characterised by tetraploidy, aneuploidy, and early copy number changes, possibly allowing for pre-malignancy detection of OAC (126, 128-130). Malignancy transformation can be detected at high levels of sensitivity and specificity using measurements of genomic instability such as flow cytometric DNA analysis, single nucleotide polymorphism arrays, and shallow whole genome arrays (130, 131). The main complication when evaluating clonal diversity is the requirement of a wide sampling field, which, given the

high risk of complications during traditional endoscopic procedures, may favour the use of non-endoscopic pan-oesophagus sampling modalities such as capsule delivered sponges (129).

Contrastingly, high levels of heterogeneity result in no effective biomarkers from gene mutation panels (132). The burden of mutations in BO is reported at 5.6 – 6.8 SNVs/Mb, however differences between non-dysplastic and dysplastic cases is uncertain (129, 133). Immuno-histological confirmation of p53 aberrance correlates with high-grade dysplasia and OAC development, with its use recommended as part of UK guidance for histopathological assessment (41, 132), with emerging evidence for the association between the rapid accumulation of copy number heterogeneity and p53 loss (127). Recent data is suggestive of the theory that extrachromosomal DNA almost exclusively arises due to TP53 alterations in high-grade dysplasia, amplifying a variety of oncogenes and immunomodulatory genes (134). Additional meta-analysis revealed a higher genetic correlation between GORD and BO, than GORD and OAC (135) (136).

Once developed, OAC can be considered a structurally unstable, heterogeneous, C-type malignancy, characterised by intensive large scale structural rearrangements and frequent copy number changes (129, 137, 138). These structural rearrangements are also present in high-grade dysplasia, including fragile site deletions, chromothripsis, breakage-fusion-breakage cycles and somatic mobile element insertions (137). The mutational burden of OAC is much higher than that of BO, at 7.1 – 25.2 SNVs/Mb, averaging 9.9 SNVs/Mb (139). In a reflection of BO colonal diversity, little overlap is seen between the mutational patterns of OAC and BO. Over 60 mutational driver genes can be recognised in OAC development, with a median of 5 driver gene events per tumour (140), with these genes exherbated and promoted by assisting patient-specific helper genes (141). A commonly known driver, SMAD4 mutation or deletion, is common in one third of OAC cases, and is associated with significantly higher mortality and poorer prognosis (hazard ration 0.6, 95 % confidence interval 0.42 to 0.84; $p = 0.003$) (140). This trend is reflected in *GATA4* amplification (0.64, 0.38 – 0.78; $p < 0.001$) and activation of the Wnt pathway, which has previous association with well-differentiated tumours (140). Later stages contain an increased number of APOBEC signatures indicative of worsened outcomes (142).

Three relevant treatment groups can be categorised from mutational OAC signatures (138). One-fifth of patients display a DNA damage repair impairment signature, which could

improve susceptibility to irradiation treatment or a combination of other DNA damaging agents with poly ADP ribose polymerase (PARP) inhibition. The second group enriches for a mutational pattern associated with ageing, C>A/T dominant, whilst the final group, comprising half of all patients, has a T>G dominant mutational pattern, associated with a high mutational load (138). This hyper-mutation is indicative of the loss of immune signalling genes and Wnt dysregulation. An example gene would be β2 microglobulin, explaining the poor sensitivity of OAC to checkpoint inhibition (140). The immune landscape is further influenced through the consistent amplification of targetable receptor kinases, also present during dysplastic progression of BO (138, 143).

1.3 A Model System

1.3.1 Available Models for Research

In order to determine the most effective experimental models for studying OAC progression from BO, it is imperative to assess the advantages and limitations of current models. Murine models enable transgenic studies, allowing for lineage tracing and selective gene manipulation. These models help researchers explore the interaction between genetic factors and environmental stressors, such as diet or acid exposure, in cellular dysregulation. However, differences between mouse and human anatomy, like the absence of oesophageal submucosal glands in mice, must be considered. A review of mouse models for BO and OAC highlighted the IL-1β model (144), which induces inflammation, abnormal NF-kB signalling, and progression from metaplasia to dysplasia and neoplasia (116, 145). The model showed that high-fat diets (145, 146) and bile salts (147) exacerbate dysplasia, whilst manipulating genes like CDX2 (148), telomerase RNA component (149), and Notch2 (150) further influences disease progression. Additionally, combining the IL-1β model with hyper-gastrinemic mice revealed that increased gastrin levels (151, 152), potentially linked to proton pump inhibitor use (153), promote dysplasia and cancer stem cell expansion (151). The development of squamous dysplasia and oesophageal squamous cell carcinoma was seen in older mice (154).

Murine xenograft models allow the study of tumour formation and metastasis via implantation of genetically engineered human oesophageal cell lines into immunocompromised mice (155). These models can be used for assessing the effects of drug targeting pathways such as TGF-β, JNK and MEK signalling (156, 157), however a key limitation would be that tumour development is anatomically distinct from the oesophagus (158, 159). Further limitations of animal models are details in Table 1. The rat surgical model was developed in the 1960s, consisting of a connection between the oesophagus and the small intestine, resulting in the development of a columnar epithelium and adenocarcinoma via acid exposure (160, 161). Recent refinements have improved operative mortality (162), however key issues such as reduced acid exposure and concerns regarding the origin of the columnar epithelium limit the use of such models for physiological anatomy (154, 163).

ANIMAL MODEL	STRENGTHS / WEAKNESSES	RESULT
Mouse	Established research approach ✓ Transgenic technology ✓ Anatomic differences from humans ✗ Technically difficult surgery ✗ Oesophageal tumours may not mimic OAC ✗	Many resources and murine-appropriate reagents available for analysis ✓ Lineage tracing is possible Targeted gene silencing / overexpression ✓ Integration of controlled exposures Keratinisation, lack of oesophageal submucosal gland cultures, and forestomach lined with squamous epithelium ✗ Small areas of metaplasia in the IL-1B mouse limited to the squamocolumnar junction ✗ High mortality ✗ Both ESCCs and EACs form in the IL-1B ✗
Rat	Relatively affordable surgical model which can be combined with controlled exposures ✓ Anatomic differences from humans ✗ Technically challenging surgery with high mortality and variable approaches ✗ Limited endoscopic assessment ✗ Direct anastomosis of small intestine with oesophagus ✗ Housing, diet and other factors can affect microbiota leading to subsequence effects ✗	Microenvironment and effects of exposures can be determined ✓ Keratinisation and lack of oesophageal submucosal gland cultures and markers differ from humans ✗ Technical skills important and acidification of refluxate varies with surgical technique ✗ Difficult to obtain biopsies ✗ Post-surgery metaplasia may not be oesophageal in origin ✗ Role of some factors in variable rates of neoplasia is not well understood ✗
Dog	Surgical model with intact area of squamous mucosa ✓ Anatomic similarity to human oesophagus with the presence of oesophageal submucosal gland cultures ✓ Spontaneous oesophageal tumours are rare ✗	Results in columnar mucosa in the oesophagus ✓ Oesophageal submucosal gland culture-derived wound healing ✓ Canine research focused on natural history not feasible for OAC ✗
Pig	Anatomic similarity to human oesophagus with presence of oesophageal submucosal gland cultures ✓ Animals reach a large size with rapid growth ✗ Atypical animal model ✗	Oesophageal submucosal gland culture-derived wound healing ✓ Long-term housing & partnership with a veterinary collage may be needed ✗ Porcine reagents, genomic characterisation, and transgenic technology may be limited ✗

Table 1: An overview of animal models used in Barrett's oesophagus and oesophageal adenocarcinoma research. Reproduced with permission under a Creative Commons Attribution-NonCommercial-No Derivatives License (CC BY NC ND) (1)

Developed in the 1980s, the canine model induced oesophageal metaplasia via acidic reflux through a hiatal hernia and myotomy, forming oesophageal columnar mucosa (164, 165). The latest model supports squamous healing and reverses metaplasia using a proton pump inhibitor and porcine-derived hydrogel applied via endoscopic mucosal resection (159). While useful for comparative oncology (166), oesophageal tumours in dogs are rare and typically non-adenocarcinomas (167).

Similarly, the porcine model, with its oesophageal submucosal glands, mirrors human OAC development (168, 169). The CAV3 protein, linked to familial BO and OAC, suggests a role in wound healing (170). Limitations of this model include their large size and appropriate housing.

Human oesophageal biopsies can be cultured short-term, but immortalised cell lines offer a practical, consistent, and modifiable alternative. Risks such as phenotypic drift, contamination, and misidentification necessitate careful verification (171). Although rare in preclinical research, *in vivo* studies provide critical insights into human-relevant disease processes, identifying label-retaining cells in basal squamous epithelium and BO, unlike mouse models. Organotypic culture models mimic epithelial structure and barrier function by incorporating immune cells and fibroblasts. These models, built using oesophageal fibroblasts, collagen type I, and Matrigel, support epithelial cell layers for studies of morphology, differentiation, and columnar marker induction (107, 172-174). Organoid cultures, developed in 2011 (175), are valuable for analysing BO and OAC heterogeneity, genomic changes, and personalised treatments (176-179). Advances such as “organ-on-a-chip” technologies further enable complex multi-cellular structures and mechanical forces (180, 181).

Most BO and OAC research relies on cell lines to study metaplasia, inflammation, genomic instability, and mutations like TP53 (95, 132, 155, 182-190). Single-cell analysis is increasingly used to study gene expression and clonal mutations via mitochondrial cDNA amplification (191, 192). Multi-omics single-cell studies reveal cancer cell evolution and diversity but face challenges like limited non-polyadenylated transcript analysis and data sparsity due to small sample sizes in DNA proteomics, proteomics, and epigenomics.

1.3.2 Single Cell On-Chip

Single-cell analysis offers significant potential for uncovering cellular diversity and identifying abnormal cells that may emerge during cancer and disease development. Cell phenotype and genotype comparisons typically rely on averaged measurements from a population of cells. Whilst these results provide valuable insights, they fail to capture the responses of individual cells due to the inherent heterogeneity within cell populations. This arises from the dynamic nature of cellular processes, including cell signalling, differentiation, motility and the cell cycle, all driven by biochemical reactions. Additionally, “biological noise” leads to stochastic gene expression, even amongst genetically identical cells in identical environments, resulting in variations in phenotype (12). Therefore, whilst the average study may reveal significant differences between multiple populations, single-cell analysis can provide further insights into the heterogeneity and presence of sub-populations within mixed samples.

Single-cell technologies have already been applied across varied biomedical fields such as early disease diagnosis (193, 194), stem cell biology (195), neuroscience (196), and drug screening (197, 198). High-throughput screening is required for precise determination of any underlying causes of heterogeneity, with classic methods including optical (199) and magnetic (200) tweezing, flow cytometry (201), atomic force microscopy (202), and single-molecule fluorescence spectroscopy (203). Each of these methods has its limitations; with flow cytometry requiring the use of expensive instruments and complex cell surface markers, long pre-treatment times, sample contamination risks and a reduction in cellular viability (204). The other methods listed suffer from complex operation, high cost, and low detection flux problems.

Microfluidics provides a high-throughput option for biophysical characterisation. By confining fluid flow to the micron scale, microfluidic systems guide fluids through channels measuring tens of microns, allowing precise fluid manipulation via devices that require low sample volumes. Due to the advantages of high detection sensitivity, low cost and reagent consumption, and large-scale sub-system integration, the use of microfluidics in the life sciences is advancing rapidly. An efficient tool for single-cell analysis, the ability to manipulate single cells for capture has led to a wide range of strategies such as micro-traps (205), micro-droplets (206), micro-patterns (207), and micro-wells (208). On-chip measurements keep cells viable, enabling subsequence destructive genetic sequencing.

Recent research on oesophageal adenocarcinoma development has utilised chip-based models, albeit not at the single cell level. The development of the EndoScreen Chip by Webster *et al.,*(209) led to detection of specific serum biomarkers via the use of a lectin immunoassay, in particular Jacalin-lectin binding complement C9 (JAC-C9). Increased levels of JAC-C9 were shown to correlate with patients with oesophageal adenocarcinoma, underscoring the potential of microfluidic chips in enhancing diagnostic endoscopy.

Shimshoni *et al.,*(210) created BO organoid cultures for modelling the tissue repair processes post-gastric refluxate exposure, via integration of patient-derived epithelial cells and fibroblasts with BO cells on-chip. This replicated the epithelial-stromal interactions characterised with BO, with the model demonstrating how interactions between BO epithelial cells and fibroblasts from cancerous regions led to increased epithelial cell proliferation and pre-cancerous changes.

Ohnaga *et al*., (211) developed a polymetric microfluidic device for the isolation of circulating tumour cells (CTCs) from blood samples to determine progression and prognosis of different cancers, including oesophageal. The study concluded that the microfluidic device was able to effectively capture CTCs from oesophageal cancer cell lines, with performance influenced by anti-epithelial cell adhesion molecule antibodies (EpCAM) expression levels.

1.4 Raman Spectroscopy

Mapping tumours from sub-cellular to whole-organ scales is a significant challenge in cancer research, especially for linking biological changes to pathology. Much of our understanding of how genotype refers to phenotype comes from population-wide cell measurements (212). Whilst useful, these average measurements fail to capture the unique behaviour of individual cells in heterogeneous populations. For instance, population data may show a gradual concentration gradient, but single-cell analysis could reveal a sharp *all-or-nothing* switch. To fully understand a system, its individual components must be studied, though single-cell techniques are time-consuming and require analysing large numbers of cells for statistical significance.

High-content screening has become critical in applications such as drug delivery, relying on fluorescence-based methods (213, 214). Also known as high-content analysis, this technique combines high-throughput screening with automated microscopy and image analysis to study cells in a high-throughput manner. This has its limitations – cytotoxicity, prone to photobleaching and limited stains due to spectral overlap, making it costly and complex (215). Other single-cell techniques, such as chemical cytometry, are low-throughput and destructive. There is currently limited reliable high-throughput, high-content, non-destructive, label-free technique to analyse cell chemical composition (216).

Raman spectroscopy (RS), based on inelastic light scattering, was first predicted by Smekal in 1923 (217), and experimentally confirmed by C.V. Raman in 1928 (218). RS is a vibrational spectroscopy method where laser light interacts with a sample to produce a spectrum. When photons from the laser encounter a vibrating molecule, most scatter elastically (Rayleigh scattering) without changing energy. However, about one in a million photons undergo a shift in frequency, known as a Raman scattering (219-222). This shift can either result in energy loss (Stokes shift) or gain (anti-Stokes shift), reflecting changes in the molecule's vibrational state. The technique employed here is Stokes Raman confocal micro-spectroscopy, which reduced out-of-focus contributions to the spectrum.

Though inherently weak, RS became more feasible with the advent of high-power lasers in the 1960s (223). RS has since been applied to biological research as a label-free method offering high-content information from a single measurement. It can be used on living cells, providing biochemical data at sub-cellular resolution and differentiating between cell types (213, 224-227). This makes RS particularly valuable for studying cell heterogeneity, drug response, tissue mapping and cancer detection in microfluidic systems (228). To conduct high-throughput single-cell analysis, precise manipulation of small fluid volumes and immobilisation of cells are necessary during measurement (229). Microfluidic devices, with their microscale channels and low reagent consumption, provide an ideal platform for enhancing the throughput of RS (216). By integrating RS with microfluidics, techniques like Raman-activated cell sorting (RACS) could offer label-free, high-throughput insights into cell composition, holding promise for advanced biological and cancer research (230, 231).

1.4.1 Raman Spectroscopy of Biological Systems

Cells consist of thousands of different molecules at varying concentrations, all of which contribute to their Raman spectrum (232). For a molecule with N atoms, the number of possible vibrational bands equals 3N – 6 (for non-linear molecules), though not all are Raman active (233). CH_2 as a non-linear molecule, exhibits three fundamental vibrational modes due to its atomic configuration and degrees of freedom. For CH_2, with three atoms, this results in three vibrational modes: symmetric stretching (v_1), bending (v_2), and asymmetric stretching (v_3), as illustrated in *Figure 1.8*. Each of these modes influences the polarisability of the molecule in unique ways, determining their activity in Raman spectroscopy (234).

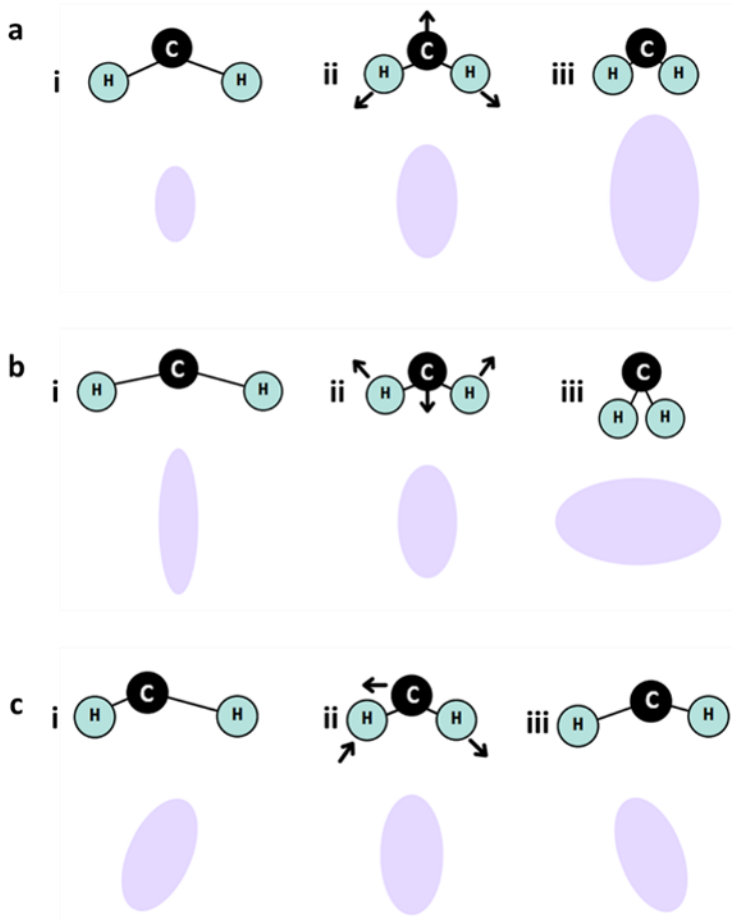

Figure 1.8: Illustration of the three fundamental vibrational modes of the CH_2 molecule : (a) symmetric mode, (b) bending mode, and (c) asymmetric mode; with their associated polarisability ellipsoid during vibration in their: (i) equilibrium state, (ii) stretched state, and (iii) compressed state.

In the symmetric stretching mode, Figure 1.8, both C-H bonds in the molecule expand and contract simultaneously, resulting in significant changes to molecular polarisability. When the bonds are stretched, the electrons are less firmly attached to the nucleus and increasing the polarisability, whilst during compression, the electrons are more tightly bound leading to a decrease in polarisability and a larger ellipsoid size. The magnitude and direction of polarisability change in this mode make it strongly Raman active, producing distinct and measureable spectral lines.

The bending mode involves changes in the angle between the two C-H bonds of the molecule. This mode causes the molecule to adopt a bent shape, altering the polarisability ellipsoid, which becomes smaller and thicker, and the molecule behaves as though it were diatomic. Although it is Raman active, the intensity of this mode is significantly lower compared to symmetric stretching due to the relatively minor changes in polarisability.

In the asymmetric stretching mode, one C-H bond lengthens whilst the other shortens, leading to complex changes in the direction of the major axis of the polarisability ellipsoid. Unlike the symmetric stretching mode, this vibration maintains the overall size and shape of the ellipsoid but results in a significant directional shift. The polarisability changes in this mode are less pronounced, making this vibration challenging to observe in Raman spectra (234).

While CH_2 is a non-linear molecule with three vibrational modes, CO_2 is a linear molecule with a centre of symmetry, resulting in unique Raman and infrared activity patterns governed by the rule of mutual exclusion (Figure 1.9). The vibrational modes of CO_2—symmetric stretching, bending, and asymmetric stretching—can be directly compared to those of CH_2 to highlight how molecular geometry and symmetry influence the polarisability changes that determine Raman activity. In symmetric stretching, both C=O bonds expand and contract simultaneously, causing significant changes in polarisability, making this mode Raman active but infrared inactive. The bending mode involves changes in the bond angle, leading to shifts in the dipole moment but minimal polarisability changes, rendering it infrared active and Raman inactive. In asymmetric stretching, one C=O bond elongates while the other compresses, resulting in notable dipole moment shifts but only minor polarisability changes, making it infrared active but Raman inactive. These behaviours highlight the rule of mutual exclusion, which dictates that vibrational modes in molecules with a centre of symmetry are typically either Raman active or infrared active, but not both (234).

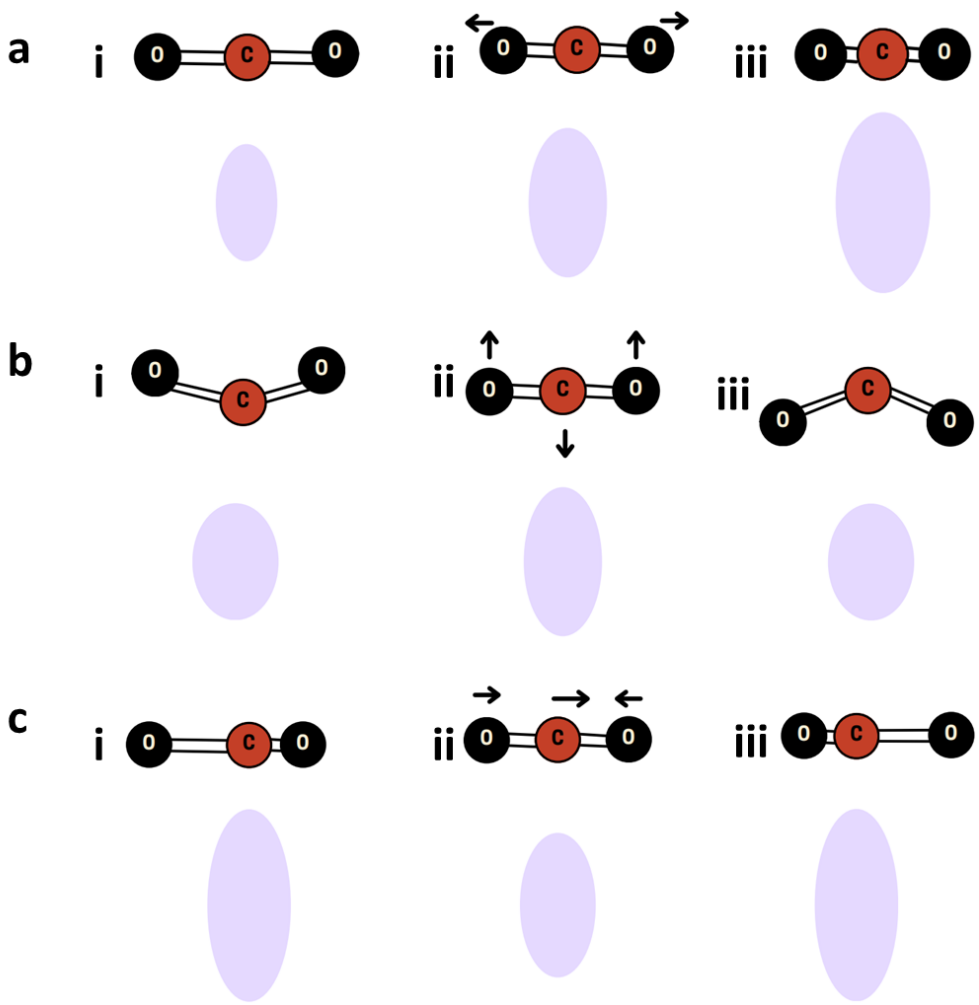

Figure 1.9: Illustration of the three fundamental vibrational modes of CO_2: (a) symmetric mode, (b) bending mode, and (c) asymmetric mode; with their associated polarisability ellipsoid during vibration in their: (i) equilibrium state, (ii) stretched state, and (iii) compressed state.

Despite the vast diversity of molecules in a cell, their abundance may be limited, with approximately 70 % of a cell's mass comprising of water, with the remainder consisting of macromolecules such as proteins, nucleic acids, and polysaccharides. RS detection sensitivity depends on a molecule's total area (223), with UV-RS able to detect proteins at lower concentrations such as 1 mg/ mL (235). This capability highlights RS's potential as a diagnostic tool, given the rich spectral content of cells. Whilst many of the main spectral features observed in cells have been reported (225, 236-238), interpreting these spectra is challenging because different molecular vibrations can contribute to the same spectral regions. Peak identification aids in understanding cellular biochemistry but is less useful for distinguishing

between cell populations. Differences between cell types are often subtle, necessitating multivariate analysis to classify them accurately (239). Chemometric techniques have been widely used for this purpose (239, 240).

A typical cell spectrum features two key regions for cellular analysis: the fingerprint region (below 1800 cm^{-1}) with many smaller peaks and the high wavenumber region above 2800 cm^{-1}, marked by a strong broad band, as illustrated in Figure 1.10 (11). Lipids contribute significantly to the C – H stretching region around 3000 cm^{-1}, whilst proteins contribute through the Amide I, II, and III bands, located around 1650, 1550 and 1300 cm^{-1}, respectively. The shape and intensity of these amide bands provides insights into protein structure. Additionally, nucleic acids (e.g. thymine, cytosine, uracil, and adenine), and disulphide bonds adds important features to the Raman spectra

A typical Raman spectrum of a cell is characterised by two key regions that are essential for cellular analysis: the fingerprint region and the high wavenumber region. The fingerprint region, which lies below 1800 cm^{-1}, contains a dense array of smaller peaks corresponding to the vibrational signatures of various molecular components within the cell. This region is highly specific, making it invaluable for identifying and distinguishing different biomolecules. In contrast, the high wavenumber region, located above 2800 cm^{-1}, is dominated by a strong, broad band associated with C-H stretching vibrations, which provide important information about the lipid and protein content in the cell (241). This is illustrated in Figure 1.10.

Lipids contribute to the C-H stretching region around 3000 cm^{-1}, as their symmetric and asymmetric stretching vibrations from CH_2 and CH_3 groups are prevalent in this region, as are proteins. Protein C-H stretching vibrations originate from side chains of amino acids, such as aliphatic groups, and from the peptide backbone, which overlap with lipid signals. These overlapping contributions provide insights into the relative proportions of lipids and proteins and their structural organisation, such as in cellular membranes or protein-lipid complexes (233, 242).

Proteins also play a prominent role in the fingerprint region, where their characteristic vibrational modes provide detailed structural and functional insights. The Amide I band, located around 1650 cm^{-1}, is primarily due to C=O stretching in the peptide bond and is highly sensitive to the secondary structure of proteins, such as α-helices and β-sheets. The Amide II band, around 1550 cm^{-1}, arises from N-H bending coupled with C-N stretching, while the Amide III band, around 1300 cm^{-1}, results from C-N stretching and N-H bending vibrations.

The shape, position, and intensity of these amide bands offer valuable information on protein conformation and interactions, making them critical markers in cellular studies (243, 244).

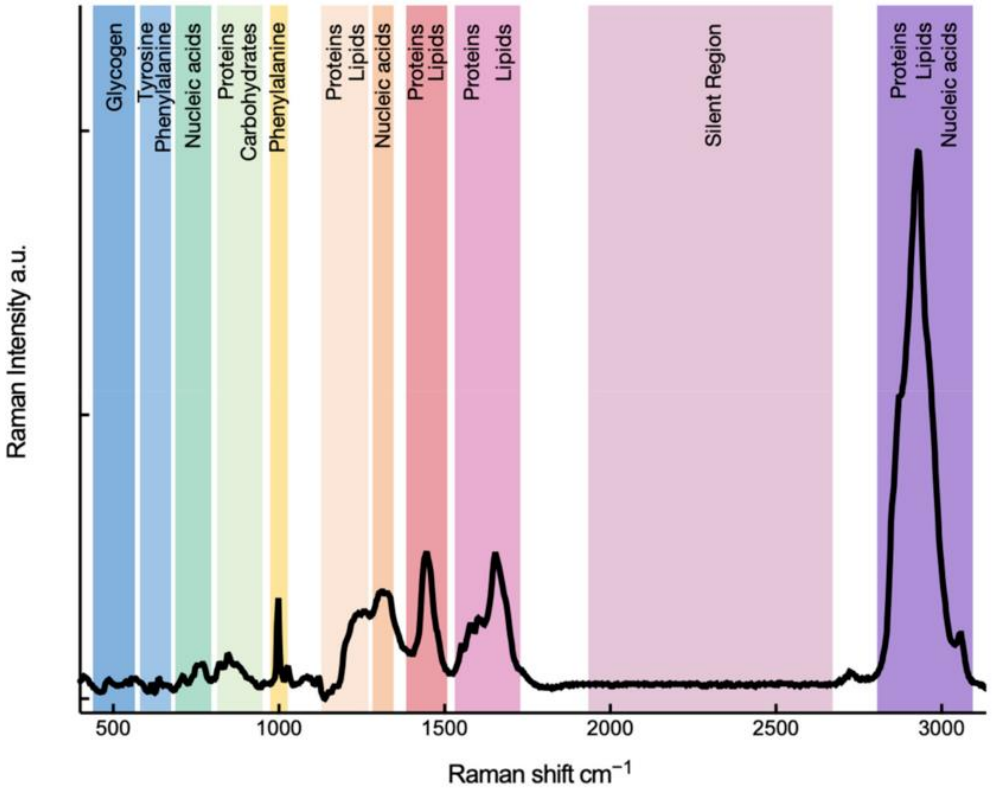

Figure 1.10: The Raman spectrum of a single cell of human primary glioblastoma U87 cell line, demonstrating various bands representative of cellular constituents. Reprinted with permission (11).

In addition to proteins and lipids, nucleic acids also contribute distinct features to the fingerprint region. Their vibrational modes, including C-N stretching, N-H bending, and ring breathing motions, are essential for identifying and analysing DNA and RNA in cells. Furthermore, disulphide bonds, characterised by S-S stretching vibrations around 500 cm^{-1}, add specific signals to the Raman spectrum, reflecting the presence of stabilising structures within proteins. (245, 246). Further details on common biological Raman bands are detailed in *Table 2* (223, 238).

Raman Shift (cm^{-1})	Biological Component	Molecular Vibration / Assignment
400-500	Proteins	S-S stretching (disulphide bonds)
620-650	Proteins (tyrosine)	Ring breathing modes of tyrosine
720-730	Lipids (choline)	C-N stretching in phospholipids
800-850	Nucleic acids	O-P-O stretching (phosphate backbone)
850-880	Proteins (tyrosine)	C-H bending or ring modes
1000-1030	Proteins (phenylalanine)	Ring breathing modes of phenylalanine
1080-1100	Lipids/DNA	Symmetric PO_2^- stretching in DNA or phospholipids
1240-1260	Proteins (amide III)	C-N stretching and N-H bending (α-helix)
1300-1350	Lipids	CH_2 twisting and wagging
1440-1450	Lipids/Proteins	CH_2 and CH_3 deformation
1550-1580	Proteins (amide II)	N-H bending and C-N stretching
1600-1650	Proteins (amide I)	C=O stretching (secondary structures)
1650-1700	Unsaturated lipids	C=C stretching
2850-2880	Lipids/Proteins	CH_2 symmetric stretching
2880-2940	Lipids/Proteins	CH_2 asymmetric stretching
3000-3050	Aromatic amino acids	C-H stretching in phenylalanine, tyrosine, tryptophan

Table 2: Table of common Raman shifts (in cm^{-1}) associated with biological molecules, detailing their corresponding molecular components and vibrational modes, such as S-S stretching in proteins, C-N stretching in lipids, and CH_2 twisting in lipids, highlighting the specificity of Raman spectroscopy in identifying biomolecular structures.

When performing RS on single cells, long acquisition times are required to obtain high-quality spectra, often raging from tens to hundreds of seconds per cell. This can limit the number of cells analysed in each experiment. For instance, Swain *et al.,* (247) required 40 seconds per spectrum over four different spots on the cells, restricting the sample size. Similarly, Krishna *et al.,* (236) used 90-second acquisition times with spectra averaged three times. In live cell applications, RS has been demonstrated with various cancer cell lines. Gala de Pablo *et al.* (248) conducted a study using label-free live single-cell Raman spectroscopy to obtain chemical fingerprints of colorectal cancer cells, specifically focusing on the SW480 and SW620 cell lines. Principal component analysis combined with linear discriminant analysis achieved an accuracy of 98.7% in cell classification. The study found that SW480 cells exhibited higher content of disordered secondary protein structures in the amide III band, while SW620 cells showed larger α-helix and β-sheet band content. Additionally, SW620 cells displayed higher levels of nucleic acids, phosphates, saccharides, and CH_2 content. Oshima *et al.,* (249) conducted an RS study using a 532 nm laser on one healthy and four lung cancer cell lines, revealing that cytochrome C (bands 747, 1127, and 1583cm^{-1}) was the primary source of variability. However, they only performed five measurements per culture dish with a 1 µm laser spot, yielding highly specific signals rather than an average over the whole cell. Pully *et al.,* (250) used RS for time-lapse imaging of human lymphocytes from healthy donors with a 647.1 nm laser. After acquiring 10 images, each with a 2-minute acquisition time, the

cells showed signs of photo-damage, including blebbing and loss of cytoplasmic and nuclear material, due to the imaging conditions in buffer and at room temperature. They observed a decline in the 1524 and 1154 cm^{-1} bands, associated with carotenoids, but saw no significant changes in protein or nucleic acid bands. Laser trapping RS (LTRS), which involves optical trapping and Raman acquisition in live cells, was used by Harvey *et al.,* (251) to classify bladder and prostate cancer cell lines, analysing spectra from 20-30 cells per line. Protein-related changes were observed in the amide I (1650 cm^{-1}) and amide III (1300 cm^{-1}) regions, reflecting variations in protein content and structure. Nucleic acid differences were evident around 785 cm^{-1}, associated with nucleic acid components. Lipid-related variations were noted in the CH_2 and CH_3 stretching regions (2850–2950 cm^{-1}), indicating differences in lipid content.

Similarly, Casabella *et al.,* (252) employed LTRS in a Poly (methyl methacrylate) microfluidic device to study live prostate cancer cells and lymphocytes, enabling the analysis of hundreds of cells. Significant variations were observed in the amide I and amide III bands, indicative of differences in protein content and secondary structure between the two cell types. Nucleic acid-associated bands, particularly around 785 cm^{-1}, displayed notable differences, highlighting variations in DNA and RNA content. Lipid composition contributed to the spectral variability, with changes detected in the CH_2 and CH_3 stretching regions.

1.4.2 Raman for Medical Diagnosis

RS is a sensitive non-invasive method suitable for analysing living cells. However, its diagnostic application presents several challenges. Firstly, cells are heterogeneous, so an average spectrum must be captured from the entire cell (220). Secondly, RS often requires long acquisition times, as the signal is weak, and high laser powers can damage the sample (253, 254). Thirdly, the Raman signal can be easily obscured by other signals, such as auto-fluorescence or background substrate (220). To reduce background noise, substrate materials like quartz or magnesium fluoride are recommended (254, 255).

Alternate Raman techniques such as Resonance Raman and Surface-Enhanced Raman Scattering, can increase signal intensity. Resonance Raman uses a light source that matches the absorbance band of the target molecule (220, 256), whilst Surface-Enhanced Raman

Scattering amplifies the signal with the help of nanoparticles or structured metal surfaces (231, 257, 258).

RS has potential biomedical applications, from monitoring cell composition to classifying cell types (220, 224). It has been used for imaging organelles (259), studying cell cycles (247), and even identifying exosomes (260). Its diagnostic use, particularly in cancer research, has shown promise, with studies focusing on cervical (255), lung (249), breast cancer (237, 253, 261, 262) amongst others (263-265); most notably for biomarker detection across oesophageal tissues. Maitra *et al.* analysed human serum samples, revealing significant changes in lipid and protein profiles, with key biomarkers including CH_2 and CH_3 stretching vibrations (2850–2950 cm^{-1}) indicative of lipid composition, and Amide I vibrations (1650 cm^{-1}) reflecting protein secondary structures (266). In their tissue Raman mapping study, Maitra *et al.* identified nucleic acid content as a critical biomarker, with peaks around 785 cm^{-1} associated with nucleobase breathing modes and phosphate backbone vibrations (1090 cm^{-1}), along with protein conformational changes observed in the Amide I (1650 cm^{-1}) and Amide III (1300 cm^{-1}) bands (267). Hao *et al.* corroborated these findings in a systematic review and meta-analysis, highlighting consistent increases in nucleic acid signals (785 cm^{-1}) and protein-related bands (Amide I and III regions) in oesophageal cancer tissues compared to healthy samples (268). Various reviews and studies provide insights into the biological and diagnostic applications of RS, including its integration with other imaging techniques (220, 224, 269).

1.4.3 On-Chip Raman Spectroscopy

Current Raman-activated cell sorting systems often suffer from low throughput because of the intrinsically weak Raman signal. These systems frequently require manual operation, which adds inefficiency. Given the extended acquisition times in RS, first-generation sorting systems commonly rely on laser trapping, such as laser tweezers Raman spectroscopy, to confirm cells for analysis. Alternatively, laser-induced forwards transfer is sometimes used to eject cells from a substrate for further investigation, however this method presents challenges in maintaining high-quality acquisition, achieving precise cell isolation, and aligning cells accurately with the laser spot for analysis (231).

Microfluidic platforms allow for precise control of minute fluid volumes, making them highly advantageous for single-cell analysis. In terms of microfluidic device fabrication, durable materials like silicon and glass are commonly used when chemical resistance and longevity are priorities. These materials offer excellent performance but require the use of hazardous chemicals during fabrication. Quartz is an excellent material for minimising background interference in RS but is costly. Cheaper alternatives like polydimethylsiloxane (PDMS) and Perspex can be used due to their ease of fabrication, though they produce significant background signals that complicate analysis. While polymer-based devices are cost-effective, their strong out-of-focus Raman signals, particularly in shallow PDMS chambers, limit their compatibility with RS. Consequently, glass or quartz microfluidic chips are preferred in Raman-based systems despite their higher cost (228).

The field has seen several advancements in integrating RS with microfluidics. Early work by Huang *et al.,*(270) and Lao *et al.,*(271) demonstrated the potential of laser trapping RS in glass and PDMS devices, successfully classifying leukaemia cells with moderate throughput (dozens to hundreds of cells per session). Another significant milestone came from the integration of microfluidic devices with optical fibres, eliminating the need for microscopes in some set-ups and allowing for improved Raman signal collection (272). Studies have since advanced to the use of silicon nitride micro-hole arrays, which provide a low Raman background signal (273). This technology was tested on system containing leukocytes, breast cancer cells, and leukaemia cells, achieving high classification accuracy. Additionally, advancements in the use of nonlinear Raman techniques such as Raman scattering and coherent anti-Stokes Raman scattering (CARS), have addressed the need for enhanced signal intensity and higher throughput. SRS has been used to image lipid droplets in single cells within PDMS microfluidic chips, probing specific molecular vibrations with background removal techniques (274). The development of dual-comb CARS systems (275) further pushed the boundaries of throughput, reaching speeds of up to 2000 events per second in Raman-based cell cytometry. This marked the fastest Raman-based cell analysis system to date, offering unprecedented capabilities for real-time cell sorting and biochemical analysis.

Whilst the integration of RS with microfluidics has advanced considerably, significant challenges remain. These include reducing background interference, optimising material choices and enhancing signal intensity to increase throughput and precision in real-time and

single-cell analysis. Further innovations, particularly in material engineering and signal processing, will be crucial for the future of this field.

1.4.4 Raman Theory

The “Raman Effect” was first theoretically predicted by Smekal in 1923, followed by quantum mechanical formulations from Kramers and Heisenberg in 1925, and Dirac in 1927. The effect was experimentally observed by Raman and Krishnan in 1928, leading to its name. Raman’s experimental setup was relatively simple: he used sunlight as the light source, with a violet filter to isolate violet light, and a green filter to detect scattered green light. When violet light scattered through different liquids, Raman noticed the appearance of a green component, marking the first observation of what is now known as Raman scattering. The Raman Effect can be understood through two theoretical frameworks: classical electrodynamics and quantum mechanics, each offering different insights.

1.4.4.1 Classic Theory

In the classical description, molecules are modelled as collections of bonded atoms that vibrate at characteristic frequencies. These vibrational motions, which are often approximated as perfectly harmonic for simplicity, are intrinsic to the molecular structure. For small symmetric molecules like CO_2 and CCl_4, sub-modes such as stretching and bending can be clearly identified and analysed (223). When light interacts with a molecule, the electric field component of the incident electromagnetic wave distorts the electron cloud, inducing a dipole moment (μ). This induced dipole moment is proportional to the electric field (E), as expressed in the relationship:

$$\mu = \alpha E \qquad \textbf{1.4.4.1}$$

where α represents the polarisability of the molecule. The polarisability, a measure of how easily the electron cloud can be distorted, is a tensor described by components (α_{ij})which depend on the molecule's orientation relative to the electric field. For simplicity, many

models neglect geometric factors such as orientation and spatial arrangement, focusing instead on the bulk behaviour of the polarisability tensor (222).

Light, as an electromagnetic wave, primarily interacts with matter through its electric field component *E*, which oscillates with time. The term E_0 is the amplitude of the wave, indicating the maximum strength of the electric field, with larger values corresponding to more intense light. The cosine function, $\cos(2\pi\nu_0 t)$, defines the oscillatory nature of the field. Within this term, ν_0 is the frequency of the wave, measured in hertz (Hz), determining how many cycles the wave completes per second, and *t* represents time. The factor 2π converts the frequency into angular units (radians per second), aligning with the periodic nature of the cosine function. This field can be described mathematically at a phase of zero and a frequency (ν_0):

$$E = E_0 \cos(2\pi\nu_0 t) \qquad \textbf{1.4.4.2}$$

When a molecule vibrates, any change in the polarisability of a bond or group of atoms can be expressed as a Taylor series around the equilibrium position. For small vibrations, this is simplified by considering only the first-order change in polarisability with respect to the vibrational displacement (q_v) along the vibration axis:

$$\alpha = \alpha_0 + \left(\frac{\partial\alpha}{\partial q_v}\right)_0 \cdot q_v \qquad \textbf{1.4.4.3}$$

Here, α_0 represents the equilibrium polarisability, which is the intrinsic polarisability of the molecule when it is at rest and not vibrating. The term $(\partial\alpha/\partial q_v)_0$ represents the rate of change of polarisability with respect to the vibrational displacement (q_v), evaluated at the equilibrium position. This indicates how sensitive the polarisability is to changes in the molecular structure as it vibrates. The vibrational displacement (q_v) describes how far the molecule's atoms move from their equilibrium positions due to vibrational motion. For small vibrations, this displacement can be approximated as a harmonic motion, mathematically expressed where q_0 is theamplitude of vibration, ν_v is the vibrational frequency, and *t* is time.

$$q_v = q_0 \cos(2\pi\nu_v t)$$ **1.4.4.4**

Substituting these equations into the expression for the dipole moment, μ=αE, results in a time-dependent dipole moment:

$$\mu(t) = \alpha_0 E_0 \cos(2\pi\nu_0 t) + \left(\frac{\partial\alpha}{\partial q_v}\right)_0 \cdot E_0 q_0 \cos(2\pi\nu_0 t)\cos(2\pi\nu_v t)$$

1.4.4.5

The first term $\alpha_0 E_0 \cos(2\pi\nu_0 t)$ represents Rayleigh scattering, where the molecule's equilibrium polarisability (α_0) interacts with the electric field (E_0) of the incident light, oscillating at its original frequency (ν_0). This corresponds to elastic scattering, with no energy exchanged between the light and the molecule. The second term accounts for the contribution of molecular vibrations to the dipole moment, leading to Raman scattering. Here, $(\partial\alpha/\partial q_v)_0$ is the rate of change of polarisability with respect to the vibrational displacement (q_v), and the amplitude of vibration (q_0). The molecular vibration, oscillating at frequency ν_v, modulates the induced dipole moment, resulting in frequency shifts in the scattered light. These shifts include the Stokes ($\nu_0 - \nu_v$) and Anti-Stokes ($\nu_0 + \nu_v$) components, corresponding to energy being transferred to or absorbed from the vibration, respectively. This equation highlights the connection between the light's electric field and the molecule's vibrational motion, forming the basis of Raman scattering and its application in identifying molecular vibrational modes (223).

Using the trigonometric identity *2cos(A)cos(B) = cos(A+B) + cos(A-B)*, the dipole moment can be defined. This is used to break down the product of two oscillating cosine terms into a sum of two new cosine terms with frequencies corresponding to the sum (*A+B*) and difference (*A−B*) of the original frequencies. The two oscillating cosine terms in this equation are $\cos(2\pi\nu_0 t)$, representing the light's electric field, and $\cos(2\pi\nu_v t)$, representing the molecular vibration (222). The dipole moment can be expressed as:

$$\mu(t) = \alpha_0 E_0 \cos(2\pi\nu_0 t) + \frac{1}{2}\left(\frac{\partial\alpha}{\partial q_v}\right)_0 q_0 E_0 \left[\cos(2\pi(\nu_0 - \nu_v)t) + \cos(2\pi(\nu_0 + \nu_v)t)\right]$$

1.4.4.6

This equation reveals three components of the dipole moment:

1. Rayleigh scattering (ν_0): Elastic scattering at the same frequency as the incident light.
2. Stokes scattering ($\nu_0 - \nu_v$): Inelastic scattering at a lower frequency, where energy is transferred to the molecular vibration.
3. Anti-Stokes scattering ($\nu_0 + \nu_v$): Inelastic scattering at a higher frequency, where energy is absorbed from the vibration.

For Raman scattering to occur, the polarisability must vary during molecular vibrations. Only vibrational modes for which $(\partial\alpha/\partial q_v)_0 \neq 0$ are considered Raman active, with examples previously discussed in Figure 1.8 and Figure 1.9.

The vibrational frequency (ν_v) of a molecule can be approximated using the ball-and-spring model, where atoms are connected by "springs" representing chemical bonds (223, 233). The frequency of harmonic vibrations depends on the bond strength (spring constant, k) and the reduced mass (μ_m) of the bonded atoms:

$$\nu_v = \frac{1}{2\pi}\sqrt{\frac{k}{\mu_m}}, \quad \mu_m = \frac{m_1 m_2}{m_1 + m_2}$$

1.4.4.7

Stronger bonds (e.g., double or triple bonds) have higher spring constants, leading to higher vibrational frequencies. Conversely, heavier atoms increase the reduced mass, resulting in lower frequencies. Raman spectroscopy is particularly effective in analysing organic compounds rich in carbon and hydrogen, such as those found in biological systems. Bonds with large polarisability changes, including C=C, C=S, S=S, and aromatic structures, produce

strong Raman signals, whereas highly polarised bonds like N=O or O–H have a weaker scatter (256).

However, classical Raman theory involves approximations. Molecular vibrations are anharmonic, meaning overtones and combination bands appear but are weaker than fundamental vibrations. Higher-order effects, ignored in this theory, can lead to non-linear phenomena such as Stimulated Raman Scattering, Hyper Raman Scattering, and Coherent anti-Stokes Raman Scattering (CARS), which extend the utility of Raman spectroscopy in advanced analytical applications (276).

1.4.4.2 Quantum Theory

Quantum theory works on the assumption that a molecules energy is quantised, meaning it can only occupy specific energy levels. These energy levels consist of electronic, vibrational and rotational components, represented as $E = E_e + E_v + E_r$. Here, electronic energy levels are the most widely spaced, followed by vibrational levels, while rotational energy levels are the most closely spaced. This hierarchy of energy spacings underpins the distinct transitions observed in molecular spectroscopy.

Vibrational energy levels can be approximated using a harmonic oscillator model, which describes the vibrational energy as:

$$E_v = h\nu_v \left(v + \frac{1}{2}\right) \qquad \textbf{1.4.4.9}$$

where h is Planck's constant, ν_v is the vibrational frequency, and v is the vibrational quantum number (v = 0,1,2, …). This equation shows that vibrational energy levels are evenly spaced in the harmonic approximation, with the lowest energy state (v = 0) representing the ground vibrational state. The quantised nature of these levels explains why molecules exhibit discrete spectral features in vibrational spectroscopy.

In Raman scattering, when an incident photon interacts with a molecule, the molecule temporarily enters a virtual state. This state is not one of the molecule's true quantised

energy levels and exists only fleetingly. The transition to this virtual state is classically forbidden but is allowed in quantum mechanics due to Heisenberg's uncertainty principle, which permits such transitions within extremely short timeframes (typically less than one picosecond). The virtual state allows the molecule to scatter the incident photon either by losing energy (Stokes scattering) or gaining energy (Anti-Stokes scattering) (221).

Unlike processes such as fluorescence, where a molecule absorbs light and transitions to a real excited state, Raman scattering involves the molecule remaining in the virtual state only briefly before returning to a different quantised energy level. The scattered photon's energy reflects the difference between the initial and final molecular energy states, which corresponds to a vibrational mode.

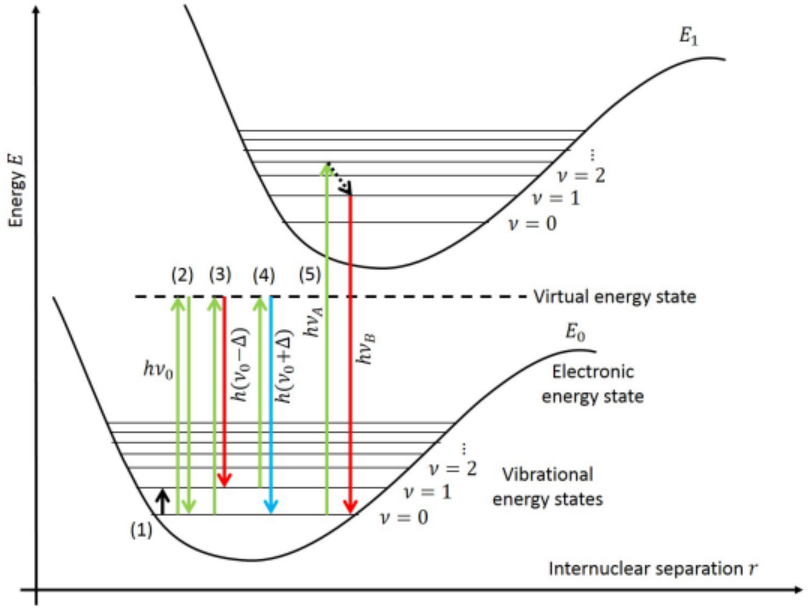

Figure 1.11: A Jablonski Energy Diagram illustrates the distinctions between (1) IR absorption, (2) Rayleigh scattering, (3) Stokes Raman scattering, (4) anti-Stokes Raman scattering and (5) fluorescence. The energy levels E_0 and E_1 represent different vibrational states at varying frequencies. A dashed line indicates a virtual state. Colours highlight the light shifts, either red or blue, with green representing the incident light.

At room temperature, most molecules occupy their ground vibrational state (v = 0), with the population of higher energy levels following a Boltzmann distribution:

$$\frac{n_j}{N} = \frac{e^{-h\nu_j/kT}}{\sum_i e^{-h\nu_i/kT}} \qquad \textbf{1.4.4.10}$$

Where n_j is the population of a specific vibrational level, N is the total number of molecules, $h\nu_j$ *is the energy of the* j-th vibrational level, k is Boltzmann's constant, and T is the absolute temperature. This distribution explains why Stokes transitions, where molecules start in the ground state and scatter to a higher vibrational state, are more common than Anti-Stokes transitions, which require the molecule to already be in an excited vibrational state.

1.5 Deformation Cytometry in Cancer Research

Single-cell techniques are often slow and require large sample sizes for statistical relevance. Flow cytometry offers an alternative approach, analysing thousands of cells per second to measure physical and chemical traits, albeit requiring sample labelling (277). Cells pass through a laser beam in single file, with scattered light providing data on properties like size and shape. The addition of labelling with fluorescent labels, provides further chemical and molecular information. Deformation cytometry is a relatively new form of cytometry that measures the mechanical properties and is thus complimentary to conventional cytometry modalities.

Biological materials exhibit mechanical properties across different scales, from protein filaments to whole organs. In cancer research, one challenge is linking biological changes to pathology whilst finding suitable methods to map tumours across various length scales. Cell deformability arises from a complex interaction of its subcellular components. Disease can alter these components, including the cytoskeleton, nucleus, and cytoplasm, making deformability a useful marker for identifying diseased cells (278, 279). When applied to Barrett's oesophagus development, it can be used to mechanically profile cells to identify diseased populations whilst tracking changes in their mechanical properties throughout disease progression to cancer.

1.5.1 Structure- and Fluid-Induced Deformation

Cell deformation via microfluidic techniques can be classified into two categories: structure-induced and fluid-induced deformation (280). Structure-induced deformation occurs when a cell is passed through a constriction channel narrower than the cell diameter ($W < D$). This method measures parameters such as cell elongation, transit time, and recovery to determine cell stiffness. It is particularly useful for studying red blood cells, as the narrow channels mimic *in vivo* capillaries. Shelby *et al.,* first used this technique to deform malaria-infected red blood cells (281) and Zheng *et al.,* paired it with impedance measurements, eliminating the need for image analysis (282). This enabled higher throughput (~ 100 cells / s) whilst analysing parameters like transit time and impedance ratios. However, cell volume and adhesion to the channel walls can affect results, and the smaller channels are prone to clogging, which limits throughput.

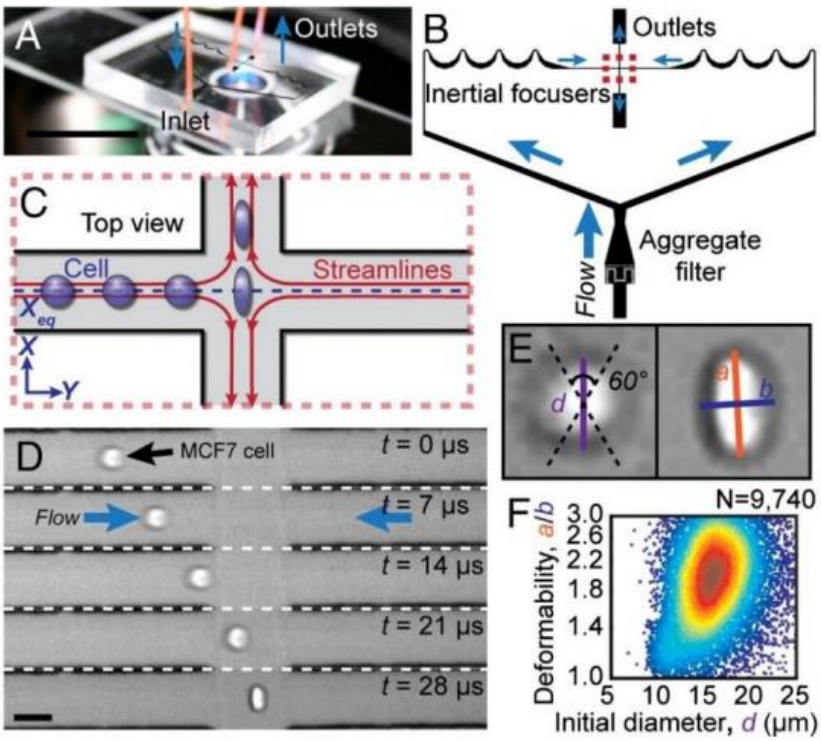

Figure 1.12: The principles of deformation cytometry are illustrated as follows: (a) device set-up; (b) a schematic of the cross-slot device featuring inertial focusers provided; (c) deformation at the stagnation point; (d) a high-speed image captures a deformation event; (e) shape parameters used to quantify deformability are defined; (f) a density scatter plot presents deformability measurements. Reprinted with permission (8).

Fluid-induced deformation, on the other hand, deforms cells through shear or extensional hydrodynamic forces rather than structural confinement (283, 284). This approach avoids the friction and adhesion issues present in structure-induced deformation. Constriction channels with widths slightly larger than the cell ($W > D$) can apply shear fluid stress, as demonstrated by Zhang *et al.,* who showed that red blood cells undergo stretching, tumbling and recoiling (285) – with these cells naturally deformable, this method is less effective for other cell types. Extensional flow, where a velocity gradient is created between a larger chamber and a contraction area, has been shown to induce more significant deformation. Lee *et al.,* demonstrated that extensional flow deforms cells more efficiently than shear flow (286).

1.5.2 Microfluidic Geometry

A cross-slot geometry can be used to generate extensional flow, using two opposing inlets and outlets to deform cells at the stagnation point (SP) in the centre of the junction. This method can achieve high strains (> 50 %) (8) but requires cells to be focused along a central streamline, which can be achieved using inertial (287) or viscoelastic (288) focusing. Deformability cytometry (DC), introduced by the Di Carlo Group (8) utilises inertial focusing and a cross-slot geometry to mechanically phenotype cells. This method significantly increased throughput, reaching up to 20,000 cells /s (289, 290). The applied stress field depends on the cells initial size and shape, so these properties are measured independently for classification (291). Deformability is defined as the ratio of the cells major and minor axes, with density scatter plots used to visualise the data.

Microfluidic techniques such as DC utilise high-speed imaging to capture thousands of cell deformation events, generating extensive datasets that require significant computational time for analysis – often resulting in a delay of around 15 minutes before results are available. To overcome this limitation, the Guck group developed Real-Time Deformation Cytometry (RT-DC), which allows for continuous acquisition of deformation data with real-time analysis (292).

RT-DC employs constriction deformations where the channels width exceed the cell diameter ($W > D$) and suspends cells in high-viscosity buffers to enhance shear forces induced by proximity to the channel walls. This setup causes cells to deform into a bullet-like shape due to the strong velocity gradients within the channel. Whilst RT-DC operates at a throughput

of 100 cells / s, lower than DC, it has analysed total cell populations exceeding 100,000, compared to DC's processing of only a few thousand despite its faster rate (280). The deformations in RT-DC are relatively small and anti-symmetric, making them more challenging to characterise than those in DC.

Mietke *et al.,* developed an analytical model for RT-DC that relates elastic parameters to cell deformation by combining theory with simulations and experiments (293). By assuming steady-state deformations, they calculated surface stresses on an isotropic, incompressible sphere within a cylindrical channel and used these as boundary conditions to solve linear elasticity equations. The model, valid for small deformations, allowed them to calculate the elastic modulus E of cells, effectively decoupling size from deformation. Although their results agreed well with experiments using rectangular channels and agar beads of known stiffness, the elastic modulus measured for HL60 cells (E = 1.48 ± 0.51 kPa) was around nine times higher than values obtained using atomic force microscopy. They suggested that the shorted timescale of RT-DC (~1 ms) or the dominant influence of the cell membrane and cytoskeletal cortex might explain this discrepancy.

Guillou *et al.,* used a cross-slot microfluidic device to deform cells at low strain ($0.01 \leq \varepsilon \leq 0.18$) and low *Re* numbers (< 0.2), allowing them to apply an analytical model to extract mechanical properties (294). They employed a viscoelastic two-parameter power-law model to predict the shear modulus of cells, validating their results with micropipette aspiration and comparisons to dextran particles. Due to the requirement that cells pass through the stagnation point of the device, the number of cells recorded for each flow condition was relatively low ($10 \leq n \leq 30$), and the model was only applicable for small deformations.

Integrating traditional precise techniques with microfluidics has the potential to create high-throughput systems capable of outputting mechanical properties. For instance, Guck *et al.,* combined continuous cell flow through a microchannel with optical stretching measurements (295). This method requires precise alignment of laser fibres on either side of the microchannel, adding complexity. Micropipette aspiration can also be integrated with microfluidics to increase throughput and extract properties like elastic modulus (296), but the rectangular shape of microchannels may reduce the applicability of traditional micropipette aspiration models. The elastic modulus measured using different techniques can vary due to several factors. Firstly, materials often exhibit viscoelastic behaviour, meaning they possess both elastic (G') and viscous (G'') characteristics. As a result, the

measured modulus can be rate-dependent, varying with the speed or frequency of the applied force. Secondly, variations can arise due to the depth of indentation or the extent of deformation. At shallow indentations, the mechanical response may primarily reflect the properties of the cell membrane, whereas deeper indentations engage underlying structures such as the actin cytoskeleton, resulting in different modulus values. Similar considerations apply to whole-cell deformation. These factors may contribute to the discrepancies reported in the literature (290). Additionally, throughput remains limited compared to other microfluidic techniques like DC and RT-DC. Further detailed discussions on the advantages and disadvantages of various biomechanical assays can be found in literature (280, 290, 297, 298).

1.5.3 Shear and Inertial Forces

Hydrodynamic cell stretching results from a combination of shear and inertial forces, with the *Re* number indicating whether shear or inertia dominates. The flow regime is influenced by device geometry, cell suspension viscosity and flow rate, with this configuration allowing for the induction of both low and high strains across various flow conditions.

The shear force acting on a spherical particle is derived from the viscous stresses exerted by the surrounding fluid. This force is proportional to the shear rate ($\dot{\gamma}$) and the fluid's dynamic viscosity (μ). For a spherical particle with radius *r*, the shear force (F_s) is given by:

$$F_s \approx \dot{\gamma}\mu(4\pi r^2) = 2\pi U\mu r \qquad \textbf{1.5.1.1}$$

The shear rate ($\dot{\gamma}$), quantifies the rate of deformation of the fluid layers, whilst the fluids resistance to deformation is quantified as (μ). *U* represents the relative viscosity between the fluid and the particle. Shear forces dominate in low Reynolds number (*Re*) regimes, where viscous effects are significant, such as in microfluidic devices or Stokes flow.

Inertial forces arise due to the particle's interaction with the flow's momentum. For a spherical particle moving in a flow with velocity *U*, the inertial force (F_c) is given by:

$$F_c \approx \frac{1}{2}\rho U^2 C_D A_p \qquad \textbf{1.5.1.2}$$

The density of the fluid is represented using (ρ), with C_D the drag coefficient, dependent on the *Re*. The projected cross-sectional area of the particle (πr^2 for a sphere), is denoted using A_p. This equation shows that the inertial force scales with the square of the relative velocity (U^2) and is influenced by the drag coefficient (C_D), which varies based on the flow regime.

The dominance of shear (F_s) or inertial (F_c) forces depends on the flow regime:

- Shear-dominated flow: At low *Re*, viscous forces govern particle motion, with F_s proportional to the viscosity (μ) and shear rate ($\dot{\gamma}$).
- Inertia-dominated flow: At high *Re*, inertial forces become significant, scaling with the fluid density (ρ) and the square of the velocity (U^2).

1.5.4 Microfluidic Theory

Microfluidics revolves around controlling and manipulating fluids in channels which have dimensions typically on the micrometre scale. At this scale, the behaviour of fluid flow is dictated by the interplay between inertial and viscous forces, with the Reynolds number (*Re*) serving as a dimensionless parameter to quantify their relative importance (299). *Re* can be expressed as:

$$Re = \frac{\text{inertial forces}}{\text{viscous forces}} = \frac{\rho v D_H}{\mu} \qquad \textbf{1.5.4.1}$$

Where *Re* can be determined using the fluid density (ρ), mean flow velocity (v), hydraulic diameter of the channel (D_H) and fluid dynamic viscosity (μ). *Re* provides critical insight into flow characteristics, distinguishing between laminar (*Re* < 2300) and turbulent (*Re* > 4000) regimes. In microfluidics, the extremely small dimensions of channels ensure that *Re* remains low, often < 1, resulting in laminar flow, where fluid moves in smooth, predictable streamlines with negligible turbulence (300).

The hydraulic diameter (D_H) quantifies the effective diameter of non-circular channels, accounting for the geometry of the cross-section. For rectangular micro-channels, it is calculated as:

$$D_H = \frac{4A}{P} = \frac{2wh}{w+h} \qquad \textbf{1.5.4.2}$$

The cross-sectional area of the channel is defined by ($w \cdot h$), the perimeter of the channel by ($2w + 2h$), along with the width (w) and height (h) of the rectangular channel. This parameter ensures that equations developed for circular pipes, such as the Reynolds number, can be applied to micro-channels with arbitrary shapes.

At low *Re*, viscous forces dominate, leading to laminar flow, where particle velocities are stable and predictable. This contrasts with turbulent flow, seen at high *Re*, characterised by chaotic, unpredictable particle paths due to inertial instabilities like vortices and eddies. The transition between these regimes occurs at a critical Reynolds number, typically around *Re* = 2300 for circular pipes. However, this value varies with channel geometry and surface properties. In microfluidic systems, laminar flow dominates because of the small channel dimensions, and systems often operate under Stokes flow (*Re* < 1), where inertial effects are negligible. An intermediate regime, roughly 1 < *Re* < 100, allows for subtle inertial effects that can be exploited for particle focusing and sorting (301).

In microfluidic channels, flow resistance plays a significant role in determining the system's performance. The relationship between pressure drop (ΔP), flow rate (Q), and channel resistance (R_Φ) can be determined via the equation $\Delta P = QR_\Phi$, with the channel resistance for a rectangular channel given as:

$$R_\Phi = \frac{f(\beta)\mu(2w+2h)^2 L}{32(wh)^3} \qquad \textbf{1.5.4.3}$$

Where f(β) is a dimensionless function of the channel aspect ratio (β = h / w), calculated using a polynomial expansion, L is the channel length, and w and h refer to the channel width and height, respectively. For channels with high aspect ratios (β ≈ 0), the flow becomes nearly one dimensional, simplifying calculations and system design (302).

The aspect ratio (β = h / w) plays a crucial role in determining flow characteristics within rectangular micro-channels. Function f(β) accounts for these geometric effects, as shown below.

$$f(\beta) = 96(1 - 1.3553\beta + 1.9467\beta^2 - 1.7012\beta^3 + 0.9564\beta^4 - 0.2537\beta^5) \qquad \textbf{1.5.4.4}$$

The coefficients 1.3553, 1.9467, 1.7012, 0.9564, and 0.2537 are empirically determined through mathematical fitting to experimental data and numerical simulations. These coefficients ensure that $f(\beta)$ accurately represents the relationship between aspect ratio and flow resistance for rectangular channels. $f(\beta)$ modifies the general resistance equation for rectangular channels to make it more accurate for different channel geometries, ensuring that the flow resistance reflects the true fluid behaviour for a wide range of aspect ratios, from square channels ($\beta = 1$) to very flat channels ($\beta \rightarrow 0$). When $\beta = 1$, the channel is square, and the function simplifies to the case of equal width and height. When $\beta \rightarrow 0$, the channel becomes highly rectangular (flat), and the function accounts for the confinement of flow in one dimension (303).

1.5.5 Detection of Disease Progression

Mechanical phenotyping plays a key role in disease diagnostics by identifying diseased cells based on their mechanical properties. Changes in cell deformability have been linked to conditions like cancer, sepsis (304), diabetes (305), malaria (306), and sickle cell anaemia (307). Cancer is a frequent focus, as studies consistently show that malignant cells are less stiff than healthy ones (308). This reduced stiffness is thought to result from a disorganised cytoskeleton, which allows cancerous cells to invade tissues more easily and metastasise. Cell stiffness or softness is quantifiable by measuring strain from applied forces or calculating mechanical parameter like the elastic modulus, where a lower modulus indicates softer cells. For instance, Cross *et al.,* used atomic force microscopy to show that malignant cells are over 70 % softer than normal tissue cells (309).

High-throughput screening is often necessary in mechanical phenotyping, as diseased cells may make up only a small fraction of a sample. Microfluidic techniques have been successful in distinguishing cells with various diseases. In one study, Tse *et al.,* used DC to identify tumour cells in pleural effusion samples. Scatter plots of cell diameter and deformability created distinct regions for healthy leukocytes and malignant pleural cells. A scoring system (1 – 10) was developed to assess the likelihood of malignancy, achieving 100 % accuracy for most of the cell population, with cells in certain regions requiring additional techniques for precise diagnosis (291). Armistead *et al.* used a microfluidic device to measure the deformability and relaxation of colorectal cell lines, SW480, HT29, and SW620, representing

primary, advanced, and metastatic stages, respectively. The results showed that metastatic SW620 cells exhibited the highest deformability and lowest elastic modulus, indicating increased softness, while SW480 cells, from a primary tumour, were the least deformable. Relaxation analysis revealed that metastatic cells underwent plastic deformation, whereas primary tumour cells recovered their original shape, highlighting structural cytoskeletal differences. Linear discriminant analysis achieved 69% classification accuracy, distinguishing cell lines based on deformation parameters, though complete separation required multi-parametric analysis (310).

Similarly, Che *et al.,* combined microfluidic vortex trapping and DC to isolate circulating tumour cells from blood samples. Cancerous samples showed a population of larger, more deformable cells, confirmed as tumour cells through immunofluorescence (311). Microfluidic techniques have also been used to study deformability across different cancer types and stages, including breast (295), pancreatic (312), brain (313), prostate (314), and bone (315) cancers. Whilst the mechanical stiffness of some Barrett's oesophagus cell lines has been previously investigated using atomic force microscopy (316), there is yet to be any published DC results regarding the progression from Barrett's oesophagus to oesophageal adenocarcinoma.

1.5.6 Cellular Biomechanics

1.5.6.1 The Membrane

Cell membranes are primarily composed of a phospholipid bilayer, approximately 5-6 nm thick, held together by non-covalent interactions. These membranes are fluid and dynamic, allowing lipids to move freely within the membranes plane. In addition to phospholipids, membranes contain cholesterol and various proteins. The three key types of membrane proteins are integral, peripheral, and lipid-anchored proteins, each playing crucial roles in biological functions (317). Integral proteins span the membrane, with functions such as forming ion channels that regular ion flow via chemical or mechanical triggers. Lipid-anchored proteins are covalently attached to lipid molecules, securing them to either side of the membrane. Peripheral proteins are temporarily associated with integral proteins or the membrane surface.

Membrane proteins also link the membrane to the cytoskeleton, maintaining cell shape and integrity. Additionally, these proteins facilitate cell attachment to the extracellular matrix (ECM) and neighbouring cells, forming tissues. Whilst the membrane is fluid-like, contributing to the cells viscosity and resistance to bending, the cytoskeleton is considered the primary factor in determining cell mechanics. The cytoskeleton is responsible for maintaining cell shape, organising organelles, enabling cell division and facilitating movement. Being highly adaptable, the cytoskeleton constantly reorganises itself in response to environmental cues and signalling pathways, such as those active during the cell cycle when DNA replication and division occur. It also plays a crucial role in processes such as tissue development and would healing. Composed of three filament types, detailed in Figure 13: microtubules, intermediate filaments and actin filaments (also known as microfilaments), with each serving a distinct mechanical and functional role (318).

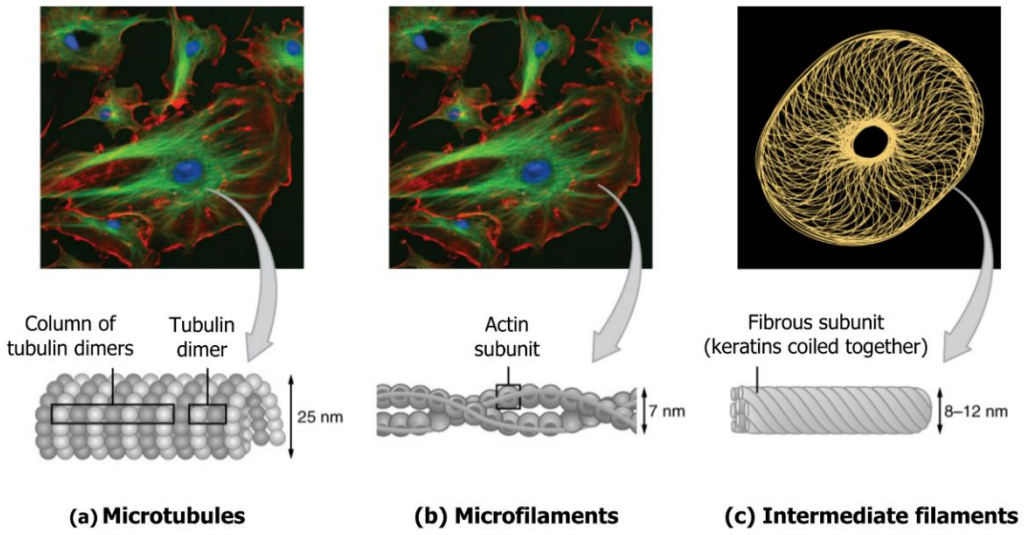

Figure 1.13: (a) Microtubules: Hollow cylindrical shape composed of many small spheres. Measures 25 nm tall. (b) Microfilaments: 2 strands of actin subunits, denoted by small spheres, twisted into a helix shape. Measures 7 nm tall. (c) Intermediate filaments: Fibrous subunit of keratins coiled together. Measures 8-12 nm tall (2)

1.5.6.2 Microfilaments

Microfilaments, also known as actin filaments or F-actin, are the thinnest of the three cytoskeletal filaments in eukaryotic cells, with a diameter of 7 – 10 nm. They are formed by the polymerisation of actin monomers (G-actin, ~ 42 kDa), which self-assemble into a double-helix structure with a pitch around ~ 72 nm. Actin filaments are polarised, with two distinct ends: the barbed (+) end, where the G-actin is preferentially added, and the pointed (-) end, where disassociation occurs (319). When the rates of polymerisation and depolymerisation at the respective ends are balanced, the filament exhibits "treadmilling", appearing to move forward.

The stiffness of cytoskeletal filaments is defined by their persistence length (l_p). When a filaments length is below its l_p, it behaves as a flexible rod, whilst lengths above l_p cause the filament to behave like a 3D random walk. F-actin has a persistence length of 3 – 17 µm, longer than intermediate filaments (200 nm – 1 µm) but shorter than microtubules (4 – 8 mm) (320). Actin filaments form bundles and networks that regulate cell shape and provide structural support. Various actin-binding cross-linkers control the orientation and spacing of

these structures, particularly at the cell's periphery, where actin filaments form a ~100 nm-thick network called the actin / cell cortex (321). This structure, consisting of actin filaments and myosin motors anchored to the lipid membrane via the membrane proteins, provides mechanical rigidity and helps maintain cell shape. Myosin motor proteins, which move along actin filaments in an ATP-dependent manner, generate tension and transport intracellular vesicles. Due to the fluidity of the membrane, the protein components of the cortex turn over rapidly, giving it a dynamic plastic nature. The composition of the lipid membrane and the actin network are closely linked and changes in one can lead to mechanical changes in the entire cell.

1.5.6.3 Microtubules

Microtubules, with a diameter of 25 nm and lengths ranging from 0.2 μm to 25 μm, are the largest of the cytoskeletal filaments. These are composed of α and β tubulin dimers that polymerise in a helical structure, forming filaments with a pitch of around 13 dimers, which self-assemble into cylindrical structures (322). They extend radially from the nucleus, helping maintain cell structure and enabling transport between the nucleus and other parts of the cell. The centrosome acts as the main organising centre for microtubule formation, promoting their characteristic radial growth. Two motor proteins, kinesin and dynein, facilitate transport across microtubules, with each protein moving in a specific direction determined by the filaments polarity, which is defined by the α or β subunits at the ends.

Microtubules are the stiffest of the cytoskeletal filaments, with a l_p of 4 – 8 mm (322). Despite their rigidity, microtubules contribute less to overall cell mechanics compared to microfilament, which, although individually less stiff, gain mechanical strength through extensive cross-linking.

1.5.6.4 Intermediate Filaments

Intermediate filaments, with a diameter of 10 – 24 nm, are named as so because their size falls between microfilaments and microtubules. Unlike microfilaments (made of G-actin) and microtubules (made of tubulin), intermediate filaments are composed of proteins encoded by 70 different genes, categorised into six subtypes based on their amino acid sequences and structures. Their final structure consists of eight protofilaments twisted into a rope-like

arrangement. As they assemble into anti-parallel tetramers, intermediate filaments lack the polarity (distinct plus and minus ends) seen in microfilaments and microtubules.

With more stability and less dynamic, there is no “treadmilling” behaviour observed. Whilst not observed in cell motility, their assembly and disassembly is regulated via phosphorylation. They serve various functions such as supporting the cell membrane and anchoring organelles within the cytosol. Cytoplasmic intermediate filaments include vimentin and keratin, whilst lamin is an example of a nuclear intermediate filament, providing structure and mechanical stability to the nucleus. These filaments undergo disassembly and reorganisation during mitosis.

The most flexible of the cytoskeletal components, intermediate filaments have a l_p of 200 nm to 1 µm, withstanding strains up to 200 % without breaking (323). This resilience is due to their hierarchical structure, allowing unfolding of subunits without rupturing. In the cell actin filaments form a peripheral cortex responsible for resisting mechanical stress. Above a certain strain threshold, cytoplasmic flow occurs, and intermediate filaments maintain cell integrity by allowing flexibility without making the cytoskeleton overly rigid or prone to excessive deformation (324).

1.6 Thesis Aims and Outline

The aim of this PhD project was to investigate and identify biochemical changes on live single cells from healthy oesophageal epithelium, through each stage of Barrett’s oesophagus, to oesophageal adenocarcinoma. These experiments were conducted on microfluidic devices, and conducted using biophysical techniques such as Raman spectroscopy and deformation cytometry. Both the current literature and theoretical background were covered in this chapter (**Chapter 1**). The next chapter of this thesis (**Chapter 2**) constitutes the methodologies that were optimised and utilised for this research.

Chapter 3 is the first results chapter and here I introduce the first set of single cell results obtained using Raman spectroscopy on untreated single cells from cell lines: HET-1A, CP-A, CP-B, CP-C, CP-D, and OE19. The presented analysis details the assigned cell peaks with most relevance for each cell line, whilst identifying differences between each sub-category through multivariate analysis. **Chapter 4** continues this story, with the Barrett’s oesophagus

cell lines: CP-A, CP-B, CP-C, and CP-D, exposed to fluctuating pH and the presence of acidic-bile salts in order to mimic the changing physiology seen in patients experiencing regular gastric reflux episodes. A combination of Raman spectroscopy and morphological analysis determines the effect of recurring changes to the extracellular environment, made possible through the use of microfluidics. Changes in the mechano-biology of each stage of disease progression are explored in **Chapter 5**, where deformation cytometry is performed to determine changes within the cellular cytoskeleton under different flow regimes. Cell lines HET-1A, CP-A, CP-B, CP-C, CP-D and OE19 were tracked and analysed using a combination of machine learning techniques as well as convolutional neural network (CNN) analysis. This work is summarised in **Chapter 6** where the conclusion for each chapter is cross-referenced with an overview on the next steps for the continuation of this work. All of the data presented in this thesis is novel in regards to single cell characterisation for all six presented cell lines using Raman spectroscopy and deformation cytometry, on a microfluidic device.

2 Experimental Methods

2.1 Cell Culture and Viability

Six cell lines were acquired for these experiments. Human oesophageal cell line HET-1A (ATCC® CRL-2692™) was used for the healthy control. The Barrett's oesophagus *in vitro* model comprised of one non-dysplastic cell line and three high-grade dysplastic cell lines: CP-A/KR-42421 (ATCC® CRL-4027™), CP-B/ CP-52731 (ATCC® CRL-4028™), CP-C/CP-94251 (ATCC® CRL-4029™) and CP-D/CP-18821 (ATCC® CRL-4030™), respectively. The cell line OE19/JROECL19 (ECACC 96071721) was used as a representative for oesophageal adenocarcinoma. All cell lines were authenticated and confirmed negative for mycoplasma infection.

Healthy (HET-1A) and Barrett's Oesophagus cell lines (CP-A, CP-B, CP-C and CP-D) were cultured in keratinocyte serum-free media (KSFM, Invitrogen), and the oesophageal adenocarcinoma cell line (OE19) was cultured in RPMI cell media containing 2 mM glutamine (Thermo Fisher, UK). Both cell media were supplemented with 10% foetal bovine serum (Merck Life Science Limited). Cells were washed and passaged using Dulbecco's Phosphate Buffered Saline (Merck Life Science Limited) and TrypLE Express Enzyme (Thermo Fisher, UK). Culture was maintained at conditions of 37 °C and 5% CO_2 with a maximum passage number of 20. Cells were passaged immediately before experimentation and re-suspended in their appropriate culture medium before flowing through the respective microfluidic device at a flow rate of 3 µL/min.

Confocal microscopy was performed on samples of cells using a Leica DMi8, allowing simultaneous capture of bright field images and fluorescent images. Cellular viability on-chip was determined by running both a live and dead stain through the device at 3 µL/min after each experiment (Viability/Cytotoxicity Assay Kit, Calcein AM, Sytox Red—L32250, Thermo Fisher, UK). A 488 nm optically pumped semiconductor laser aided sequential excitation and fluorescence emission was measured from 490 to 525 nm. Images were captured using a 10 × dry objective.

2.2 Acidic – Bile Salt Solution

An acidic bile-salt solution (ABBS) was chosen to mimic gastric reflux, with the pH acidified to pH4 via titration of 1M hydrochloric acid immediately prior to use. The bile salt solution comprised of taurocholic acid (TCA [T4009, Sigma]), glycochenodeoxycholic acid (GCDCA [#50534, Sigma]), glychocolic acid (GCA [#360512]), deoxycholic acid (DCA [D2510, Sigma]), and glycodeoxycholic acid (GDCA [sc-280755, Insight Biotechnology]), in a 100µM equimolar mixture. All salts were solubilised in DMSO prior to media addition, not exceeding a 0.1% (v/v) final concentration. An equal volume of DMSO was added to the untreated cell group as a control. pH test strips were used prior to, and periodically during, the experiment to confirm that the ABBS remained at pH 4.

2.3 Microfluidics

2.3.1 Single Cell Trapping

Whilst the integration of single-cell traps with microfluidics has been previously documented (325, 326), these traps exhibit inefficiencies for Raman spectroscopy applications due to their substantial structure, which results in a significant contribution to the background signal from PDMS material. The trap region comprised a single chamber featuring swallowtail-shaped designs that arranged single cells in an offset array by half a period. At the junction of the two “wings”, a gap of approximately 5 µm was created to accommodate single cells measuring between 10 – 13 µm. Originally designed for trapping liquid crystal droplets (327), the swallowtail-shaped hydrodynamic trap design was adapted for this research. This design offers several advantages over traditional trap structures. First, the configuration of the “wings” minimises the resistance of the gap to fluid flow. Secondly, this design alleviates the challenges associated with creating patterns with intricately small features (< 10 µm) within a relatively thick layer of SU8 (25 µm). Lastly, this trap structure is capable of capturing larger cells when the flow direction is reversed, enabling it to function as a dual-mode trapping system.

The device structure and trap design are further detailed in Figure 2.1. Prior to experimentation, cell clumps and debris were filtered out using a 20 µm cell strainer. Single cells were introduced to the device via the inlet port, as shown on Figure 2.1a, where they flowed through the trapping chamber at a flow rate of 3 µL/min, with excess single cells

removed off-chip through the outlet. Single cells were trapped in the swallowtail-shaped PDMS traps, illustrated in Figures 2.1b and 2.1c.

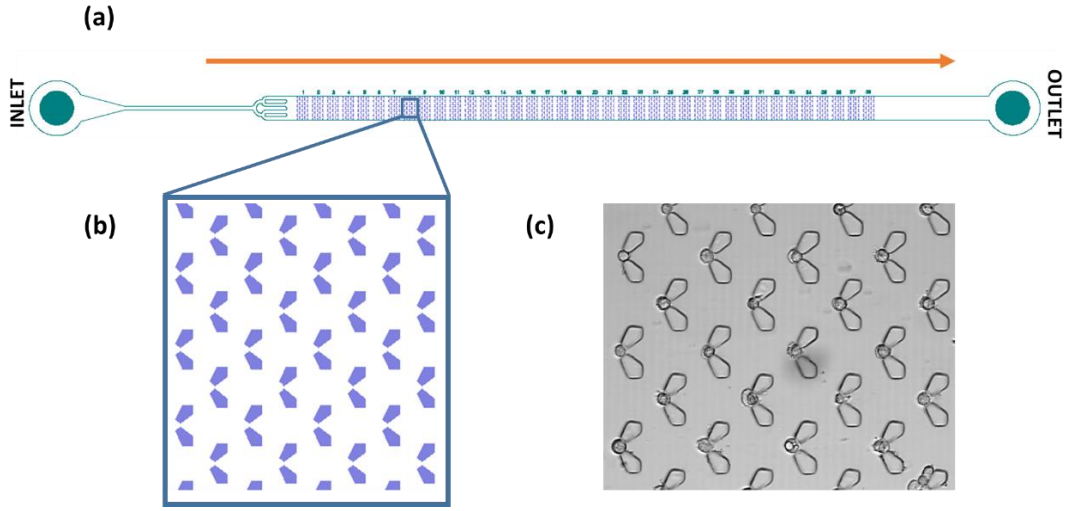

Figure 2. 1: (a) The single-chambered cell trapping device was designed using autocad and clewin. Both inlet and outlet have been appropriately labelled, with an arrow indicating the flow direction. (b) Schematic of the swallowtail-shaped single cell traps, with each row positioned offset to allow for optimal cell trapping. (c) Bright-field image of Barrett's oesophagus single cells (CP-A) trapped on-chip.

2.3.2 Deformation Cytometry

Microfluidic devices featuring a cross-slot configuration are widely utilised to create extensional flow, which is integral to this thesis for the deformation of single cells under varying flow conditions. Illustrated in Figure 2.2a is the cross-slot device, designed using AutoCAD, as previously published by Armistead *et al.,* (12, 310).

A suspension of single cells, at a concentration of 2-3 million cells/mL, was introduced through the inlet port depicted in Figure 2.2a. Prior to on-chip filtration (Figure 2.2ai) samples were passed through a 20 µm cell strainer to break up any large clumps. Figure 2.2b provides a closer look at an array of filtering pillars, spaced a minimum of 50 µm apart, and designed to prevent blockages downstream caused by cell debris.

Following filtration, the single cells were divided on-chip into two channels of equal size. These channels then converged at a junction where two opposing inflowing channels met two orthogonal outflowing channels (Figure 2.2aii). This junction was designed as the extensional-flow junction, with further details provided in the schematic shown in Figure 2.2c. It is at this location that the cross-slot configuration produces extensional flow, resulting in cell deformation at the stagnation point located at the centre of the junction. All channels within the extensional-flow junction measured 35 µm in width and 25 µm in height. Subsequently, the sample exits the device via one of the outlet ports.

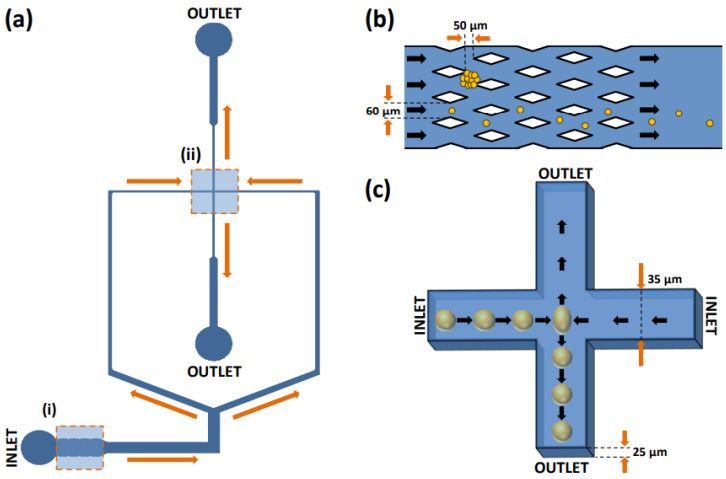

Figure 2.2: (a) The cross-slot device was designed using AutoCAD, featuring labelled inlets and outlets, with arrows indicating the flow direction. Key highlighted sections include (i) an on-chip filter and (ii) the extensional-flow junction. (b) A schematic depicts the on-chip filter positioned after the inlet to prevent larger particles from causing downstream blockages. (c) The extensional-flow junction, the critical feature of the cross-slot device, where cell deformation takes place. (6)

2.3.3 Master Fabrication

The fabrication of microfluidic devices typically consists of three key stages: designing the device geometry using AutoCAD and CleWin softwares, creating a master template, and finally, using soft lithography to produce devices from the master.

Masters for the devices were made using 3-inch silicon wafers coated with the negative photoresist SU-8 2025 (Microchem, Warwickshire, UK). The wafers were initially cleaned using a piranha wet etch (H_2SO_4 & H_2O_2) at temperatures exceeding 80 ˚C for 30 minutes.

Post-cleaning, they were rinsed and stored in deionised water, as shown in Figure 2.3a. Before use, the wafers were dried with a nitrogen gun and heated in an oven at 180 ˚C for approximately 20 minutes. Once cooled to room temperature, the wafers were placed on a spin coater (EMS 6000) to apply a uniform layer of SU-8. The acceleration, speed, and spin times were adjusted to achieve the required photoresist thickness (Figure 2.3b).

To solidify the SU-8 layer, a soft bake was performed, evaporating the solvent from the SU-8. The wafers were placed on a room-temperature hot plate (EMS 1000-1, Salisbury, UK), which was then set to 55 ˚C and allowed to heat over 5 minutes. The wafers were kept at 55 ˚C for 30 minutes before turning off the hot plate and letting the wafers cool to room temperature, which took approximately 1 hour and 30 minutes (Figure 2.3c). Direct-write laser lithography was then used to etch the microdevice patterns into the SU-8 layer with a 2 µm laser spot size (wavelength: 375 nm) using the MicroWriter ML™ system (Durham Magneto Optics) (Figure 2.3d). As a negative photoresist, the areas of SU-8 exposed to the UV laser became cross-linked, whilst the unexposed lasers remained soluble. The optimal exposure dose for the designs presented in this thesis was 400 mJ/cm^2, which provided fine detail for small channels (approximately 35 µm in width). Lower doses could improve resolution but led to adhesion problems between the SU-8 and the wafer, whereas higher doses risked overexposure, producing “bell-shaped” channels wider than intended.

A post-exposure bake was carried out, following the same temperature ramping procedure as the soft bake (Figure 2.3e). The wafers were developed in ethyl lactate ($C_5H_{10}O_3$) for approximately 2 minutes to dissolve the uncross-linked SU-8, followed by a quick rinse in isopropanol ($CH_3CHOHCH_3$) for 1 minute (Figure 2.3f). The earlier baking steps sometimes cause stress in the SU-8 layer, resulting in cracks. A hard bake at 180 C for 15 minutes, just below the glass transition temperature of SU-8 (210 C), effectively removed any cracks (Figure 2.3g). The resulting master could then be used for PDMS microdevice fabrication (Figure 2.3h).

The height of the SU-8 channels was determined using a Stylus Profiler (Bruker Dektak). This instrument employs a mechanical probe to scan the surface topography, and as the probe moves across the surface, a feedback loop maintains a constant force applied by the sample. The probe performed a line scan over an SU-8 channel, generating a height profile and determining the step height of the channel.

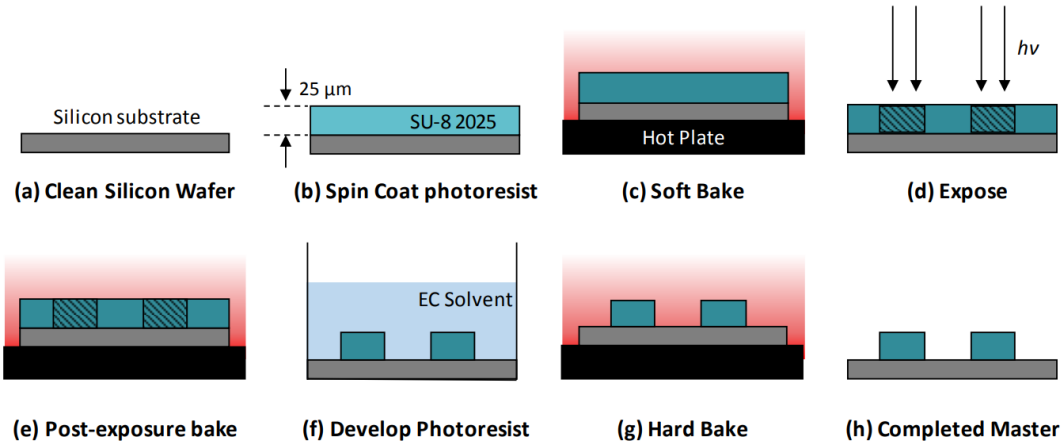

Figure 2.3: This schematic provides an overview of the production process for fabricating an SU-8-silicon master, with the key steps of the protocol displayed sequentially from (a) to (h) (6)

2.3.4 Soft Lithography

The channels of the microfluidic device were fabricated in PDMS (polydimethylsiloxane) by using the master as a mould to cast a negative replica of the SU-8 structures into the PDMS. A liquid mixture of the PDMS and the cross-linking agent SYLGARD 184, in a 1:10 ratio, was prepared and centrifuged at 40,000 rpm for 1 minute to eliminate air bubbles. This mixture was poured over the master in a petri dish (Figure 2.4a) and subjected to a soft vaccum for about an hour to remove any residual dissolved gas in the PDMS. Following this, the PDMS was cured in an oven at 75 ˚C for approximately one hour, solidifying into a hydrophobic elastomer. Once solidified, the PDMS was carefully peeled from the master (Figure 2.4b), and inlet and outlet holes for fluid access were punched using a 1 mm diameter biopsy puncher (Figure 2.4c).

To complete the device, the PDMS was oxygen plasma bonded to either a glass slide (deformation cytometry) or a glass cover slip (Raman spectroscopy), dependent on the experiment. Both the PDMS and the glass were placed in a plasma-cleaning chamber (Zepto, Diener Electronics, Germany) and treated with plasma for 30 seconds at 0.4 mbar of O_2 (Figure 2.4d). Plasma treatment oxidises both the PDMS and the glass surfaces, replacing hydrocarbon groups with hydroxyl groups (- Si – O – OH-), increasing surface activity. The PDMS and glass were then firmly pressed together to ensure bonding, forming strong

covalent bonds (-Si – O – Si -). Finally, the bonded microfluidic devices were baked at 75 ˚C for 30 minutes (Figure 2.4e).

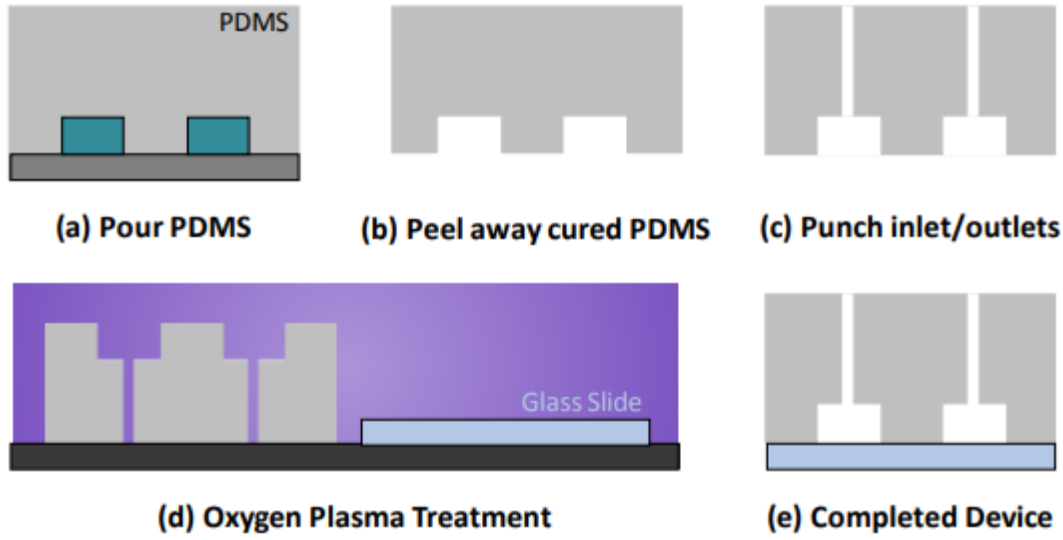

Figure 2.4: This schematic outlines the production process for fabricating microfluidic devices using PDMS, with the key steps of the protocol displayed sequentially (a – e) (6)

2.4 Raman Spectroscopy Protocol

2.4.1 Spectroscopy Facilities

The thesis utilised a Renishaw inVia Spectrometer System integrated with a Leica DMi8/SP8 laser scanning confocal microscope. Figure 2.5 illustrates the Raman spectrometer optics. The system had two excitation lasers: a 532 nm DPSS diode laser (50 mW, 22 mW at the sample) and a 785 nm near-IR diode laser (100 mW, 45 mW at the sample), both mounted with kinematic baseplates. The 532 nm laser had a plasma filter, with a pinhole to adjust the laser width. Motorised ND2 filters offered 16 power levels ranging from 0.00005 to 100%. A beam expander was used to adjust laser spot size (0-100% expansion) (Figure 2.6).

The light passed through a motorised shutter, mirrors, and a chosen objective to reach the sample in an inverted configuration. Biological samples were slotted into in-house 3D-printed holders which contained magnets to minimise movement during acquisition. These holders were then positioned on an automated XYZ stage (Z-movement is handled via objective adjustment). The high-speed encoded stage, had a step size of 0.1 µm. Renishaw's Wire

software controlled acquisition settings, except for detector and optical path changes, which were handled manually or via Leica fluorescence confocal software.

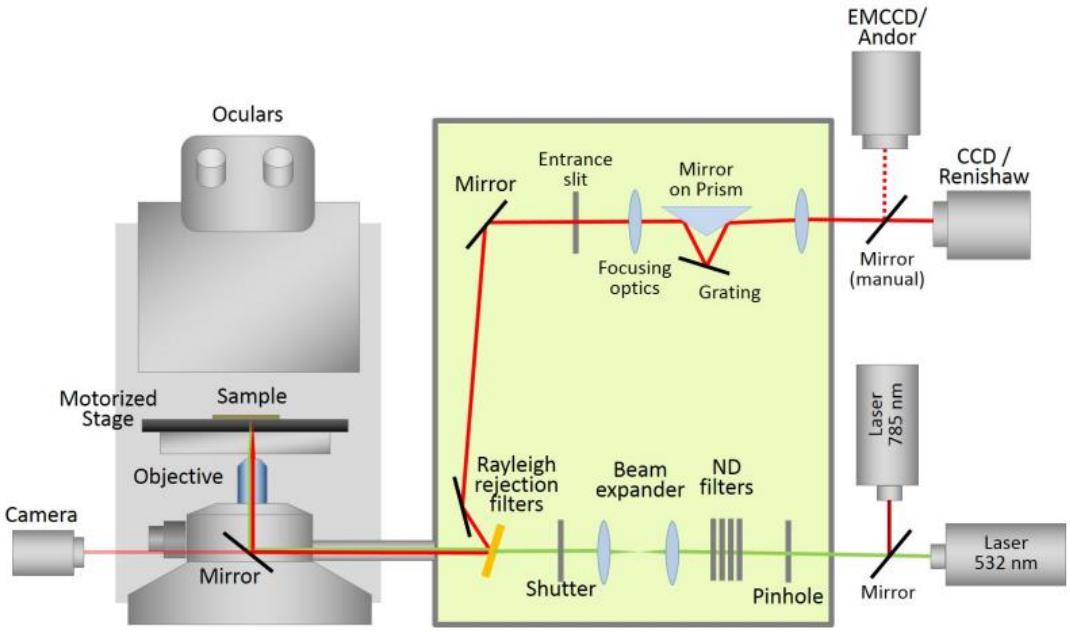

Figure 2.5: Key features of the optical configuration of the Renishaw inVia Raman microscope, simplified for clarity. Reprinted with permission (3)

Scattered light was partially reflected to a camera for laser spot observation and calibration, with most of it directed to Rayleigh reflection filters (rejecting intensities below 100 cm^{-1}). The entrance slit shapes the laser beam for grating, with standard and high confocality modes of 65 µm and 20 µm, respectively. Two gratings (1800 and 1200 lines/mm) separate wavelengths, spread across a CCD detector. Grating movement options include continuous (synchro-scan) for high-resolution spectra or step motion for spectral range selection.Two detectors were available: a Renishaw CCD array detector (infrared-enhanced, deep depletion, 1024 x 256 px, Peltier-cooled to -70°C) and an Andor EMCCD3 high-sensitivity back-illuminated detector (UV/VIS measurements, 1600 x 200 px, TE4-cooled to -100°C). These will be referred to as the Renishaw and Andor detectors.

The microscope was equipped with an EL6000 fluorescence light source (LQ-HXP 120 LEJ), a 120 W mercury short-arc reflector lamp with a wavelength range of 380-680 nm. Filter cubes with BP5 or LP6 filters allow standard fluorescence settings. BP filters transmitted light around a specific wavelength, indicated as center/bandwidth, while LP filters transmitted all

light above a certain wavelength. Available filter cubes included DAPI (exc. BP 360/40 nm, em. LP 425), FITC (exc. BP 470/40, em. LP 515), RHOD (exc. BP 540/45, em. LP 590), and TXR (BP 560/40, em. BP 630/75). An LED1 bright-field light source with a TTL shutter was also available for transmitted light.

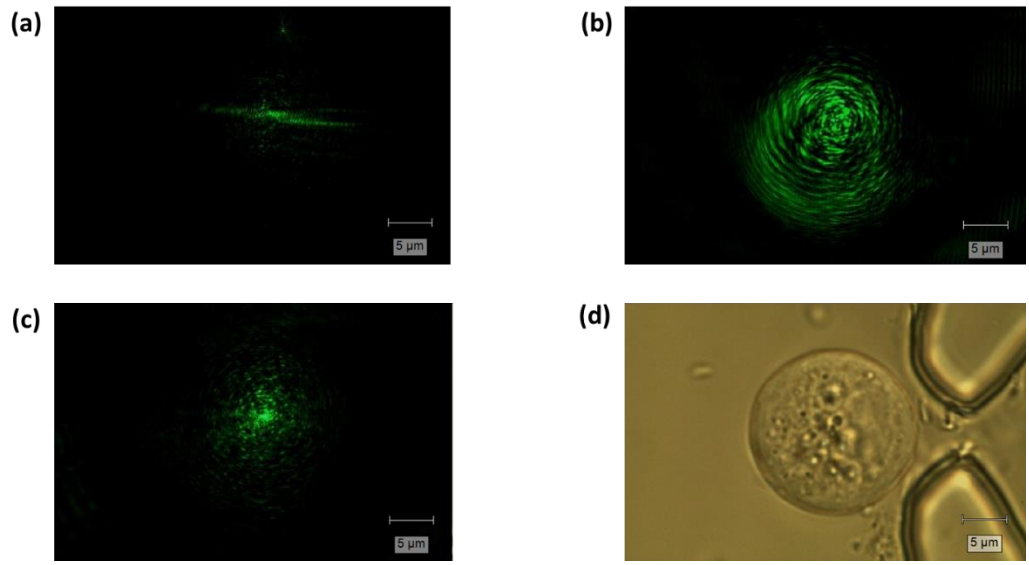

Figure 2.6: Images of the 532 nm Raman laser through the beam expander at defocus rates of (a) 0 %, (b) 20 %, and (c) 100 %. (d) Bright-field image of the single CP-A cell used for refraction of the Raman laser.

All Raman data presented in this thesis was collected via the 532 nm DPSS diode laser, supplying a laser power of 22 mW, at the sample surface, used with a grating of 1800 lines mm^{-1}. A near-infrared enhanced CCD array detector (1024 × 256 pixels, Renishaw) was used to collect light. The Raman system was calibrated to the Si band position (520.5 cm−1) prior to each experiment. All single-cell spectral data were obtained with the use of a 100 × oil objective (HC PL APO CS2 FWD 0.13 mm NA 1.4). A beam expander was positioned in the optical path between the excitation laser and the objective lens used to focus on the sample. A 20 % defocus, when used with a 100 × objective, produced a spot size of approximately 18 µm at the sample. The entrance slit was set to 20 µm. Spectral settings remained the same across all six cell lines, utilising a step configuration with a 5 s exposure time and 6 accumulations in two different windows (300–1800 cm^{-1} and 1800–3200 cm^{-1}), giving a total exposure time of 60 s per cell. To ensure reproducibility and account for motor drift over time, the Raman system was calibrated before each use with a standard silicon sample, which produced a sharp peak at 520.5 cm^{-1}. Each calibration was specific to the experiment, utilising

the same settings as expected for data collection. Background removal using Extended Multiplicative Signal Correction (EMSC) required the utilisation of both background spectra and biological reference spectra to accurately identify and isolate biological peaks. The background spectra were specifically obtained from the cell traps, ensuring that there was no contribution of any biological material, only PDMS. Meanwhile, the biological reference spectra were derived from a balanced dataset comprising various cell lines. This combination of cell lines ensured a representative range of biological variability, enabling more robust peak identification and reducing bias in the spectral analysis process.

2.5 Raman Analysis

2.5.1 Data Pre-Processing

A raw single-cell Raman spectrum can show variable background contributions depending on the cell's position in the chamber, baseline shifts from autofluorescence and scattering, intensity changes due to the cell size within the acquisition volume, cosmic ray interference, as well as intrinsic biological variation. To ensure reliable analysis across multiple experiments, consistent and robust pre-processing of the spectra is essential. The pre-processing steps are detailed below in a flow graph (Figure 2.7) using the sample spectra of 10 CP-A single cells for image purposes (Figures 2.8 - 2.9), excited under the conditions described in section 2.4.1. Both Matlab scripts were written by Dr Julia Gala de Pablo, and were adapted with permission, with the Python scripts written by Professor Stephen Evans. Example scripts and functions for both dataset creation and analysis are provided as supplementary materials in the e-thesis.

The initial dataset creation was performed using a single Matlab script, which saved the pre-processed data for efficient loading and further analysis (DatasetCreationScript_1.mat). The code iteratively processes selected datasets (`selectedDataSets`) to extract Raman spectral data. For each dataset, the program navigates to the corresponding directory and loops through all available measurement repeats (`Index [j]`). Within this inner loop, it constructs the filename for each repeat based on a predefined naming convention, then reads the associated `.wdf` file using a data-reading function. The X-axis values (Raman shifts) and intensity data (spectra) are extracted from the file and stored in a structured array (`RamanSpectra`). This systematic looping ensures all datasets and their repeats are

processed efficiently, enabling organised extraction and storage of spectral data for subsequent analysis.

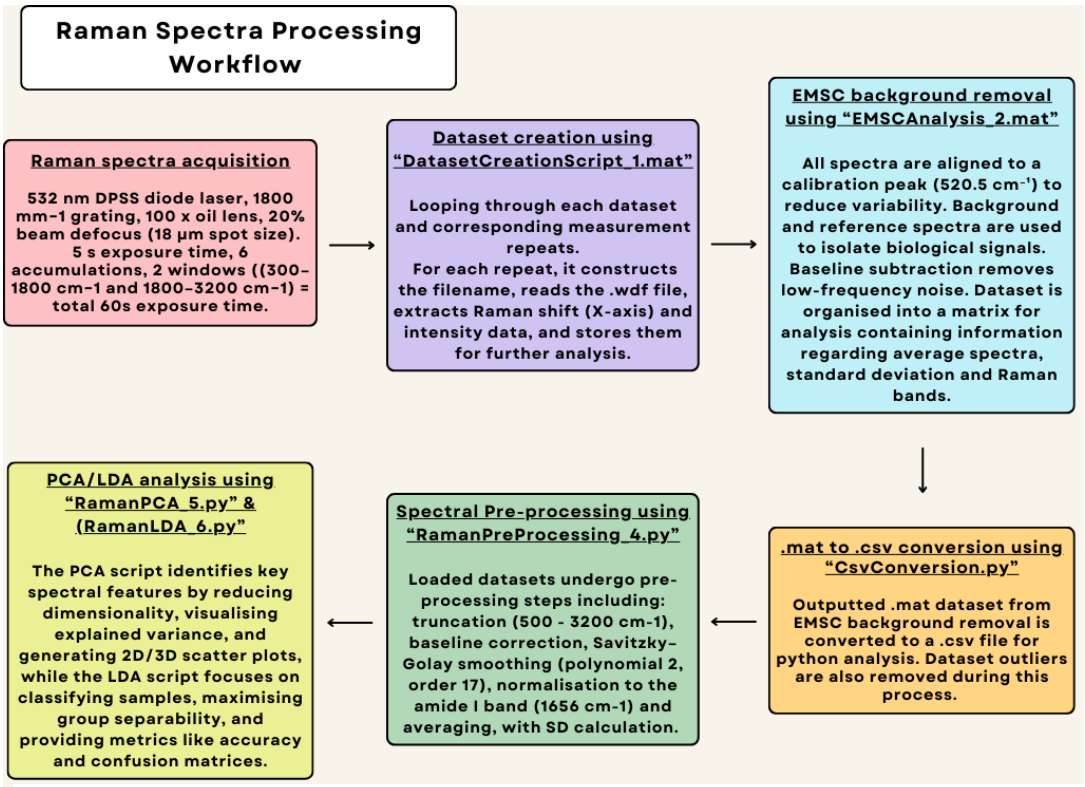

Figure 2.7: Comprehensive Workflow for Raman Spectra Acquisition, Pre-processing, and Analysis.

A second Matlab script was used for background removal (EMSCAnalysis_2.mat). The code performs EMSC to remove background signals, such as PDMS, from Raman spectra. It aligns all spectra to a calibration peak (520.5 cm^{-1}) to reduce variability and uses background and reference spectra to isolate biological signals. An example of the inputted raw, biological, background, and post-EMSC spectra is shown in below Figure 2.8. After applying EMSC, the baseline is subtracted to remove low-frequency noise, ensuring the spectra focus only on meaningful peaks. The processed data are organised into a matrix for analysis, with key features like average spectra, standard deviation, and peaks identified for further interpretation. Examples of these steps are shown in Figure 2.9. And 2.10

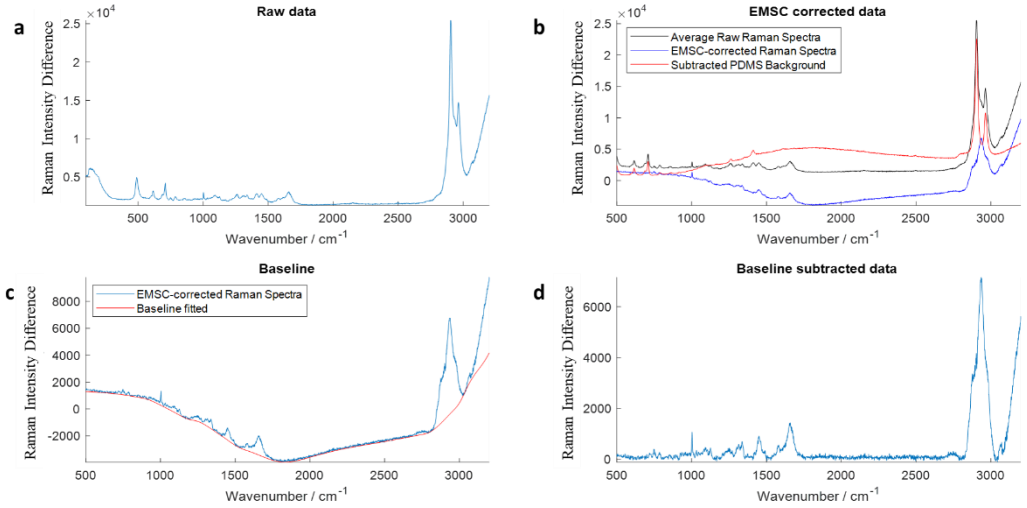

Figure 2. 8: Examples of raw CP-A single cell spectra (blue), EMSC-corrected CP-A single cell spectra (green), background PDMS spectra (black) and raw biological reference spectra (red). All spectra have been interpolated, averaged, and smoothed using a Savitzky-Golay filter (sgolayfilt function) with a polynomial order of 2. Examples of raw CP-A single cell spectra (blue), EMSC-corrected CP-A single cell spectra (green), background PDMS spectra (black) and raw biological reference spectra (red). All spectra have been interpolated, averaged, and smoothed using a Savitzky-Golay filter (sgolayfilt function) with a polynomial order of 2. The spectrum of a single CP-A cell showing (a) post-spline interpolation; (b) EMSC-corrected Raman spectra compared to the raw data and the background contribution; (c) baseline alterations post-EMSC correction; (d) the final outputted spectrum.

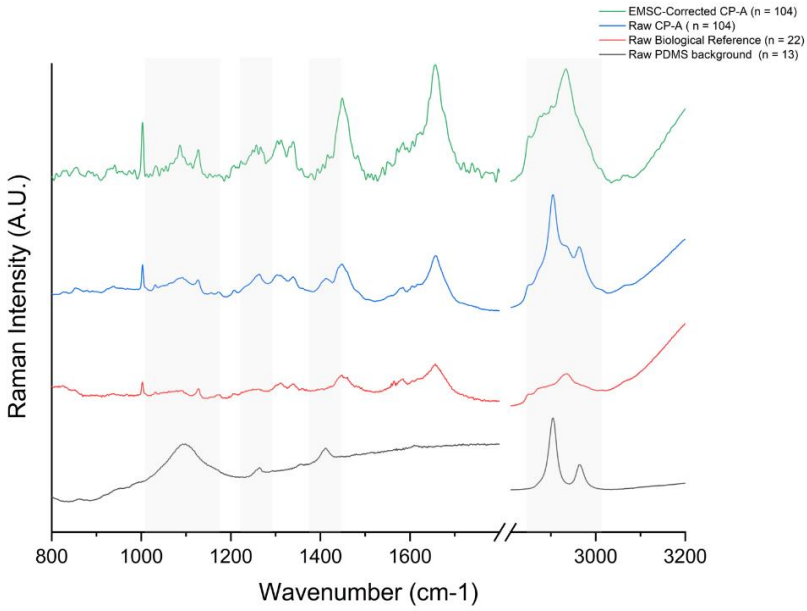

Figure 2.9 Examples of raw CP-A single cell spectra (blue), EMSC-corrected CP-A single cell spectra (green), background PDMS spectra (black) and raw biological reference spectra (red). All spectra have been interpolated, averaged, and smoothed using a Savitzky-Golay filter (sgolayfilt function) with a polynomial order of 2.

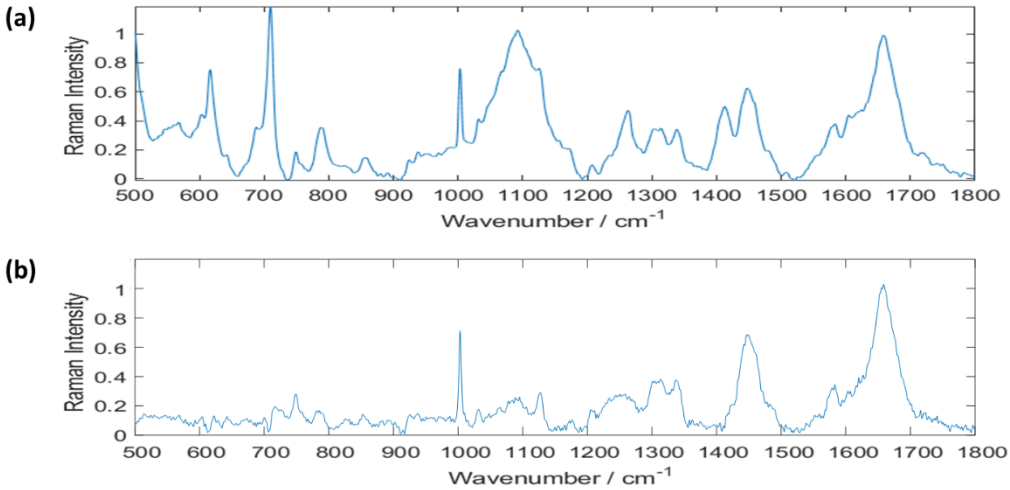

Figure 2. 10: A (a) raw and (b) analysed Raman spectrum of a single CP-A cell, post EMSC background removal and baseline correction.

EMSC-corrected datasets (.mat) were exported to Python and converted to .csv files using the CsvConversion.py script. These .csv files were then imported into RamanPreProcessing_4.py for the main pre-processing steps. The EMSC-corrected data (Figure 2.11a) was first truncated between 500–3200 cm^{-1} (Figure 2.11b), and baseline correction was applied using the Koch et al. (328) algorithm (Figure 2.11c). Additional pre-processing included smoothing (polynomial 2, window 17) (Figure 2.11d), averaging, and standard deviation calculations (Figure 2.11e).

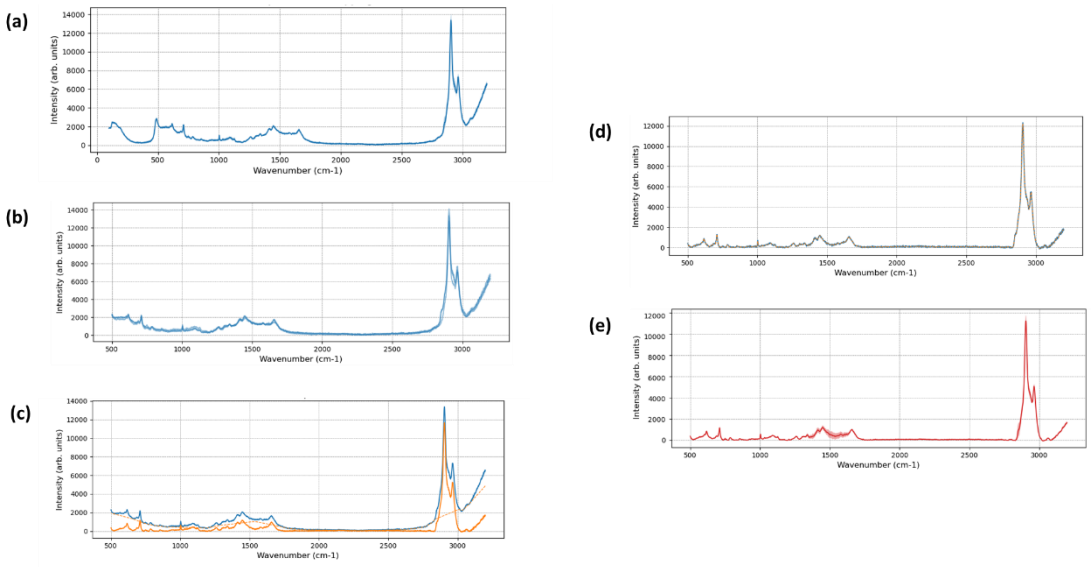

Figure 2.11: Raman spectrum pre-processing steps for 10 CP-A single cell spectra. (a) Post-EMSC correction; (b) Truncated between 500 – 3200 cm^{-1}; (c) Baseline corrected; (d) Smoothed [2, 17]; (e) Averaged, with the standard deviation calculated.

2.5.2 Multivariate Analysis

Principal component analysis and linear discriminant analysis (PCA-LDA) was conducted using the EMSC-corrected datasets, truncated within the fingerprint region (800–1800 cm^{-1}). This was carried out using Python scripts RamanPCA_5.py and RamanLDA_6.py, written by Professor Stephen Evans. All confusion matrices presented from this data are normalised using the fourfold cross-validation method. 15 principal components were chosen for the PCA as this was representative of > 90 % of the overall variance within the data. The variance explained reaches a plateau after 15 PCs, indicating that the remaining components are representative of little to no change. The fourfold cross-validation was conducted using a StratifiedKFold, whereby the datasets for each set of analysis were split into k (4) groups of equal size, with the first group held as a validation set, and the remaining groups (k − 1) used to fit the method.

2.6 Deformation Cytometry Procedure

The deformation cytometry setup is shown in Figure 2.12a, utilising the microfluidic device described in Section 2.3.2. A 1 ml glass syringe, cleaned with ethanol and deionized water, was filled with a cell sample. Fine bore polythene tubing (Smiths Medical™, Kent, UK) with 0.36 mm ID and 1.09 mm OD connected the syringe to the microdevice, fitting directly into 1 mm punctured PDMS inlets and outlets. The syringe was placed in a syringe driver (WPI, Hertfordshire, UK) to control the volumetric flow rate (µl/min) through the device. Tubing from the outlets directed the flow into a glass vial or microcentrifuge tube for sample collection. Microdevices were positioned on an inverted bright-field microscope (Figure 2.11b), using a 10x objective to capture cell deformation events.

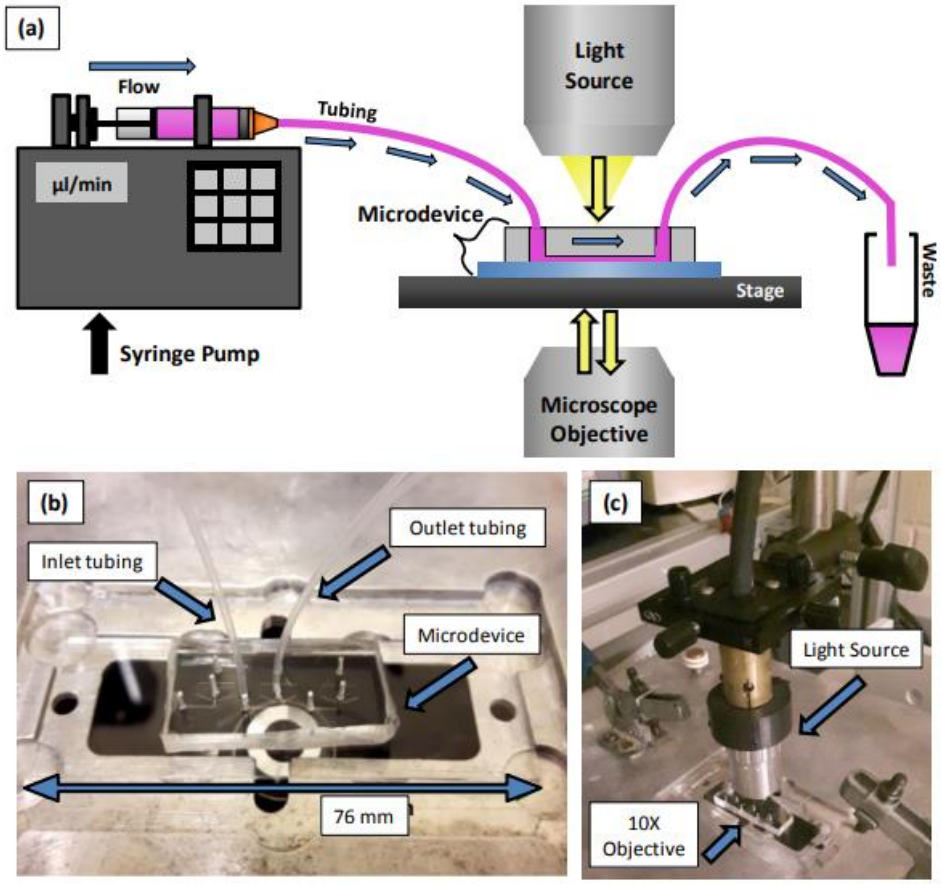

Figure 2.12: (a) Schematic of the microfluidic setup for deformation cytometry. (b) PDMS device mounted in a holder with inlet and outlet tubing for high-speed microscopy of cell deformation. (c) The device positioned between a standard inverted bright-field microscope and an additional light source to enhance image intensity. (6)

High-speed microscopy recorded these events, with the camera (Photron Fastcam SA-X) operating at frame rates of 7000 to 260,000 fps and exposure times of 0.37 to 6.67 µs. An external light source mounted over the setup (Figure 2.11c) enhanced image capture at higher frame rates and reduced exposure times to minimize motion blur. Flow rate and camera settings are detailed below in Table 3: frame rate, exposure time and camera resolution for each flow rate.

Flow rate (µL/min)	Frame rate (frames/s, fps)	Exposure time (shutter, nS)	Resolution (size of image)
5	10,000	50,000	320×264
10	15,000	70,000	320×264
20	20,000	91,000	320×264
40	42,000	150,000	320×264
60	60,000	263,000	320×192

Table 3: frame rate, exposure time and camera resolution for each flow rate.

The deformation cytometry experiments utilised an inverted microscope (Eclipse Ti-U, Nikon, Japan) with reflection bright-field microscopy, which illuminates a sample with white light and collects the reflected light. In this technique, contrast is generated through the sample's absorbance of light, though biological samples, like mammalian cells, tend to exhibit low contrast due to their optical transparency and lack of colour. Bright-field images typically lack detail unless the sample is stained, a process that often requires fixation and is not suitable for live cell imaging in suspension.

A high-speed camera is defined as having a frame rate exceeding 250 frames per second (fps) and an exposure time shorter than 1 ms. In this study, cells were deformed and recovered in less than a millisecond through microfluidics, necessitating high-speed microscopy to capture these rapid events. Key factors in high-speed cameras include image resolution, frame rate, exposure time, and data rate. The resolution must be sufficient to capture the cell before, during, and after deformation, while the frame rate should be high enough to avoid missing any critical deformation. Short exposure times are required to minimize motion blur. For this work, the FASTCAM SA-X camera (Photron), featuring a CMOS (complementary metal–oxide–semiconductor) sensor, was employed. It offers a maximum resolution of 1024 x 1024 pixels at frame rates up to 12,500 fps. Beyond this frame rate, the resolution is constrained by the data rate, reducing the field of view as frame rate increases. The camera has an ISO value of 25,000, which enables exposure times as low as 293 nanoseconds (ns).

2.7 Single Cell Deformation

2.7.1 Python Analysis

The analysis conducted in this section was conducted using Python scripts CellMovieTrimmer_ and CellMovieTracker_, both written by Professor Stephen Evans.

Videos up to 40,000 frames were uploaded into CellMovieTrimmer_, and were subsequently trimmed down to remove empty frames and reduce file size, with the outputted file uploaded for analysis in CellMovieTracker_. Through changing parameters such as flow rate and frame rate, CellMovieTracker was used to determine the average number of frames, velocity, cell size and maximum deformation for each single cell, allowing this information to be plotted (Chapter 6).

2.7.2 CNN Analysis

The analysis detailed in this portion of the thesis was conducted by MPhys student Henry Aldridge.

To prepare images for CNN training and extract parameters for LDA, cells were tracked and analysed from greyscale AVI videos recorded at 10,000 fps. These videos captured cells traveling through a DC device at a flow rate of 20 µl/min in a shear-dominant regime. Each video was processed using a Python cell-tracking script (illustrated in Figure 2.12) which included rotating frames to align channels horizontally, scaling to the lowest magnification, and cropping to show a consistent area of the device. Cells were tracked using the OpenCV module, ensuring debris and large cell clumps were excluded. Cell metrics such as area, perimeter, circularity, D.I., and strain were recorded. Duplicates were removed based on frame occurrences per video, and frames where cells exhibited maximum deformation (1 − circularity) were identified. Image processing steps included smoothing, background subtraction, colour normalisation, and shifting pixel value medians to 150 for uniform backgrounds. Additionally, corner segment datasets (~25% the size of main dataset images) were filtered to exclude partial cell segments based on mean pixel values ± 2 % of the median

Machine learning (ML) is a branch of artificial intelligence (A.I.) that uses statistical and computational methods for recognising patterns within data, enabling predictions for new unseen data (329). This is invaluable for large and complex biological datasets, which often involve high-dimensional, non-linear relationships, an example being the interactions between genes, proteins, or metabolic pathways.

Linear Discriminant Analysis (LDA) was initially used for analysing the deformation cytometry data, as it can help classify and interpret differences in cell mechanics, such as deformation, elasticity, and stiffness, which are often linked to cell state or disease. Deformation

cytometry generates high-dimensional data, and this complexity is reduced during LDA by projecting the data into a lower-dimensional space while maximising the separation between predefined groups, such as healthy and diseased cells. This makes it easier to identify key mechanical features, such as D.I., that are most relevant for classification. LDA is particularly useful for distinguishing subtle mechanical differences between cell types or states, such as normal, dysplastic, and cancerous cells, and can help identify potential biomarkers or mechanical signatures. Additionally, it is computationally efficient, making it well-suited for high-throughput applications like disease screening or hypothesis testing in clinical and research settings (330) .

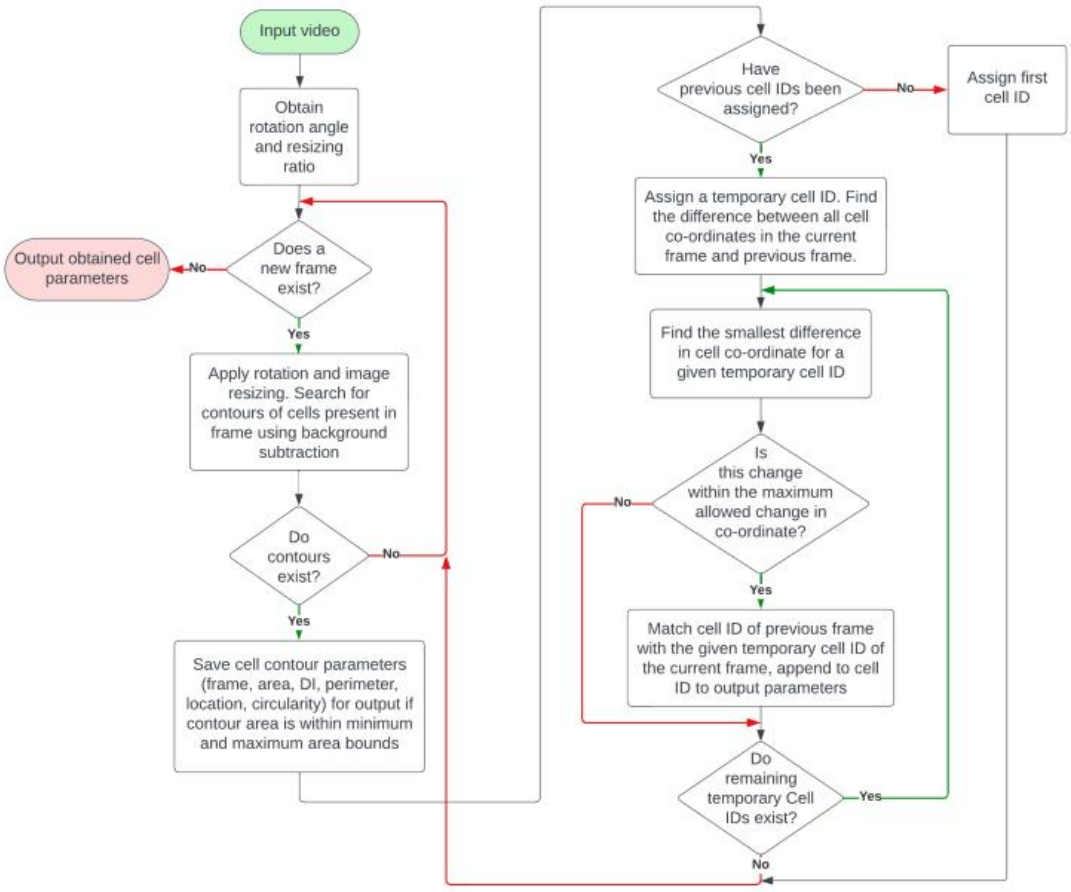

Figure 2.13: Flowchart illustrating the cell tracking algorithm logic for obtaining cell parameters. Green arrows indicate True statements, red arrows indicate False. The parameters help identify frames where cells are at maximum or no deformation, providing images for CNN training and data for LDA. (5)

CNN training was primarily based on Géron's ML textbook *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (331), utilizing two architectures: Google's Xception (332) and Microsoft's ResNet50 (333). ResNet50 was chosen due to its success in similar studies (334, 335). TensorFlow and Keras were employed for model development. Data was divided using an 80:10:10 stratified split into training, validation, and test sets. Video still images were used as the inputs, resized to 72×72 pixels and processed using the Keras preprocessing function. Xception and ResNet50 models were initialized with random weights, and their layers remained unfrozen. Each model was completed by adding a global pooling layer followed by five fully connected layers, as per Géron's approach. The global pooling layer reduced dimensionality. Both models were trained for 10 epochs using the Adam optimizer with a learning rate of 0.1, chosen for its efficiency in similar studies (334). The models were then evaluated on test datasets, and confusion matrices were generated. Metrics such as accuracy (2.7.2.1), recall (2.7.2.2), and precision (2.7.2.3) were calculated from the confusion matrices. Where *TP* is the number of true positives, *TN* is true negatives, *FP* is false positives and *FN* is false negatives (331) . These metrics allow different model qualities to be identified, such how many true positives it identifies or how many false positives it provides, which are crucial in clinical settings.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad \textbf{2.7.2.1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad \textbf{2.7.2.2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \textbf{2.7.2.3}$$

The training datasets were augmented using Keras to increase size and variability, aiming to reduce overfitting and improve model generalization. Augmentations included random vertical/horizontal flips, brightness/contrast adjustments, and small rotations, expanding the dataset fourfold. A balanced dataset was also created by duplicating images from undersampled cell lines before augmentation. For 5-fold stratified cross-validation, LDA was

applied using the scikit-learn module (6). Parameters included cell area, perimeter, circularity, DI, and strain. As with the CNN analysis, confusion matrices, accuracy, recall, and precision were calculated. Tests were repeated with a balanced dataset, achieved by randomly sampling from oversampled cell lines.

3 Raman Spectroscopy for the Staging of Oesophageal Adenocarcinoma Progression

The previous chapters of this thesis have provided an overview of the current literature in this field, detailed the technical theory behind the experimental methods, and outlined the aims of this project. This chapter presents the first key findings: novel on-chip live single-cell analysis of oesophageal adenocarcinoma development, characterised using Raman spectroscopy. Representative cell lines were employed to model healthy oesophageal epithelium (HET-1A), oesophageal adenocarcinoma (OE19), and the progressive stages of Barrett's oesophagus (CP cell lines: -A, -B, -C, and -D). Data was consolidated across multiple experiments and cell passages to minimise variability and ensure high reproducibility. Principal component analysis (PCA) combined with linear discriminant analysis (LDA) enabled accurate classification of each cell population, achieving robust differentiation both between and within disease stages.

3.1 On-Chip Optimisation

3.1.1 Trapping Efficacy

The first parameters to be defined were the conditions for single cell trapping. This involved refining conditions such as flow rates and cell concentrations to ensure consistent and reliable isolation of individual cells in each cell trap. Optimal trapping conditions were necessary to prevent on-chip cell clumping. CP-A single cells were used for this optimisation, with cells passaged, filtered through 20 µL cell strainers, and flown though the microfluidic device at flow rates between 1- 5 µL/ min. Cell concentrations for 0.5, 1, 2 and 3 million cells / mL were used to determine ideal conditions.

Figure 3.1 provides a visual and quantitative assessment of the efficacy of single cell trapping under a range of conditions. Imaged in Figure 3.1a and Figure 3.1b, lower concentrations of 0.5 and 1 million cells / mL led to an inadequate number of trapped single cells available for data collection, whilst using a higher concentration of 3 million cells/ mL (Figure 3.1d) led to cell clumping between traps, reducing the available number of traps containing only single cells. A final cell concentration of 2 million cells / mL (Figure 31.c) was established to provide the best trapping efficacy of 74 % (± 6%), calculated by the percentage of traps filled with

completely detached single cells, with results for each cell concentration plotted in Figure 3.1e. When determining optimal flow rate, it was vital that it was high enough to push cells through the device, but low enough not to force them through the traps. Detailed in Figure 3.1f, lower flow rates of 1 µL/ min and 2 µL/ min were insufficient to generate the pressure required to propel cells through the device. In contrast, higher flow rates of 4 µL/ min and 5 µL/min resulted in excessive pressure, causing cells to deform and pass through the 5 µm gaps between the traps. All single cell trapping experiments were conducted using a cell concentration of 2 million cells / mL, at a flow rate of 3 µL/ min.

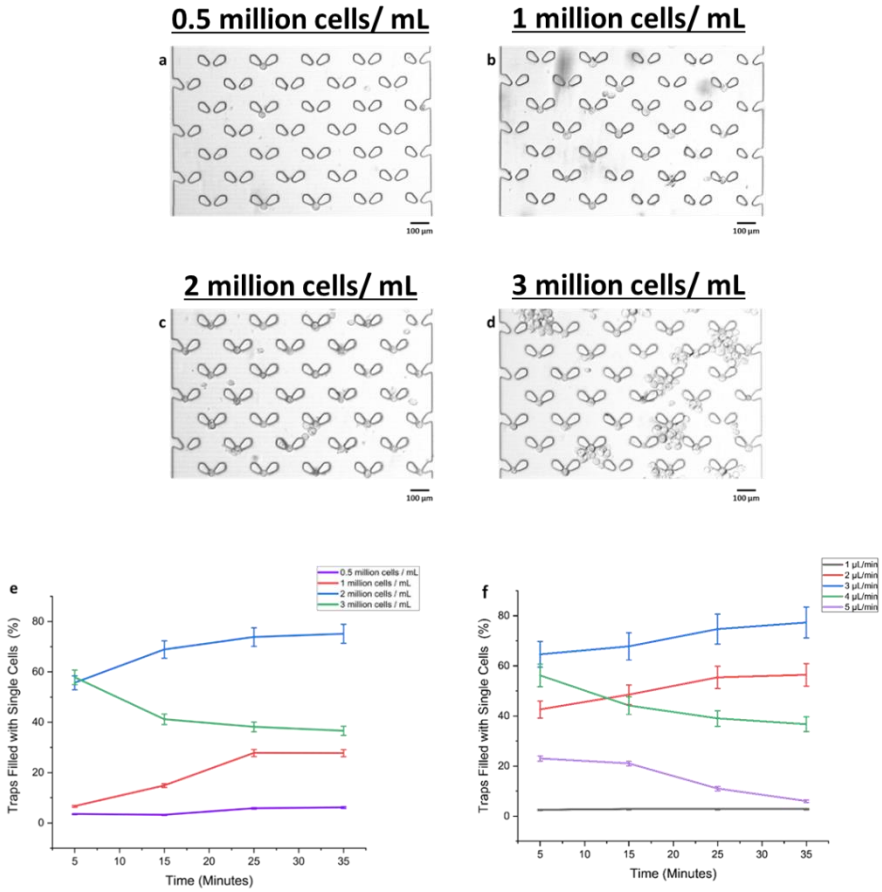

Figure 3.1: Representative brightfield images and quantitative analyses of single-cell trapping efficiency under varying cell concentrations (0.5, 1, 2, and 3 million cells/mL). Brightfield images were captured using a 10x objective lens, highlighting cell distribution within microfluidic traps. Quantitative data (bottom graphs) show the percentage of traps occupied by single cells (left) and multiple cells (right) over time. Optimal trapping conditions were achieved at a concentration of 2 million cells/mL, maximising single-cell trapping (74% ± 6%) and minimising multi-cell occupancy, while higher concentrations (3 million cells/mL) resulted in clumping and reduced trapping efficiency. Error bars represent standard deviations across three experimental replicates.

3.1.2 Cellular Viability On-Chip

Single cells were successfully trapped within the microfluidic device at a flow rate of 3 µL/min and maintained under controlled conditions at 37.5°C. To sustain cell viability, a constant flow of fresh culture media was supplied at 1 µL/min throughout the experiment. Viability was assessed using Calcein AM live stain and Sytox Red dead stain, with fluorescence images captured at hourly intervals over a 5-hour period. Each experiment was limited to a maximum duration of 5 hours, and viability data were collected for 100 single cells per independent cell line.

To evaluate the potential impact of laser exposure on cell viability, delayed apoptosis was assessed by monitoring viability over the same 5-hour period following a single laser exposure. Fluorescence viability images were acquired at hourly intervals to quantify the percentage of viable cells. The average viability for each cell line over the 5-hour experiment is summarised in Figure 3.2. These results represent the mean viability of 100 single cells per cell line across multiple experimental replicates.

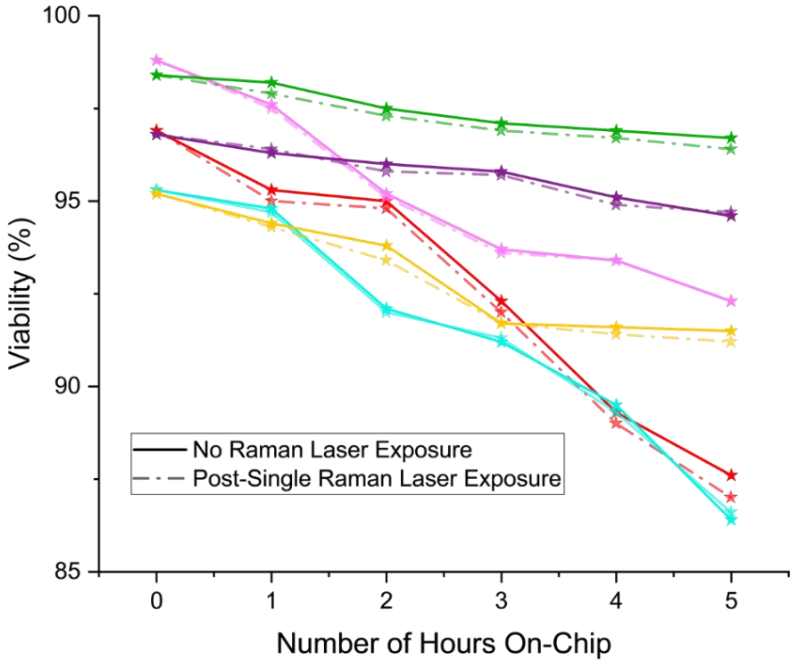

Figure 3.2: viability of single cells trapped on-chip over 5 hours for different oesophageal cell lines, including HET-1A (healthy epithelial cells, red), CP-A (non-dysplastic Barrett's oesophagus, cyan), CP-B (low-grade dysplasia, pink), CP-C (high-grade dysplasia, yellow), CP-D (high-grade dysplasia, purple), and OE19 (oesophageal adenocarcinoma, green). Cell viability was monitored hourly using Calcein AM and Sytox Red fluorescence staining, with data averaged across 100 single cells per cell line.

3.2 Raman Analysis of Live Cells

3.2.1 Spectral Profile Across All Cell Lines

Raman spectra were obtained from trapped live single cells across the cancerous cell line (OE19), those representing the healthy epithelium (HET-1A) and the intermediary Barrett's oesophagus (CP-A to CP-D), via Raman conditions detailed in the methodology (2.4 Raman Spectroscopy Protocol)

Figure 3.3 shows the averaged spectrum and standard error for each cell line, averaging over 100 single cells per sample type. All bands were pre-processed using EMSC and normalised to the Amide I peak at 1656 cm^{-1}. Each peak is labelled corresponding to the associated cellular component, with Table 4 detailing further covalent bond information.

Peak Position (cm^{-1})	Assignment	Vibrational Mode
880	Tryptophan	Ring breathing
1004	Phenylalanine	Symmetric ring breathing
1033	Phenylalanine	C–H in-plane bending
1089	PO_2^- in DNA	Symmetric stretching
1122	Glycogen	C–O stretching
1129	Skeletal of acyl backbone	C–C stretching
1171	Tyrosine in collagen	C–H bending
1206	Hydroxyproline, tyrosine	C–H bending
1312	CH_3/CH_2 twisting mode of lipids	CH_2 twisting
1336	Lipid C–H deformation	CH_2 wagging
1416	C–C stretching in quinoid ring	C=C stretching
1450	CH_2 bending	CH_2 scissoring
1570	COO^- stretching	Asymmetric COO^- stretching
1584	C–C stretch	Aromatic ring stretching
1610	Cytosine (NH_2)	N–H bending
1656	Amide I	C=O stretching
2850	Lipids	CH_2 symmetric stretching
2881	Lipids and proteins	CH_2 asymmetric stretching
2935	Chain-end CH_3 symmetric band	CH_3 symmetric stretching
2980	Asymmetric CH_3 stretch (protein)	CH_3 asymmetric stretching

Table 4: Key Raman spectral peaks and their corresponding molecular assignments and vibrational modes. Peak positions (in cm^{-1}) are associated with specific biomolecules such as proteins, lipids, and nucleic acids, providing insights into the biochemical composition and structural properties of single cells.

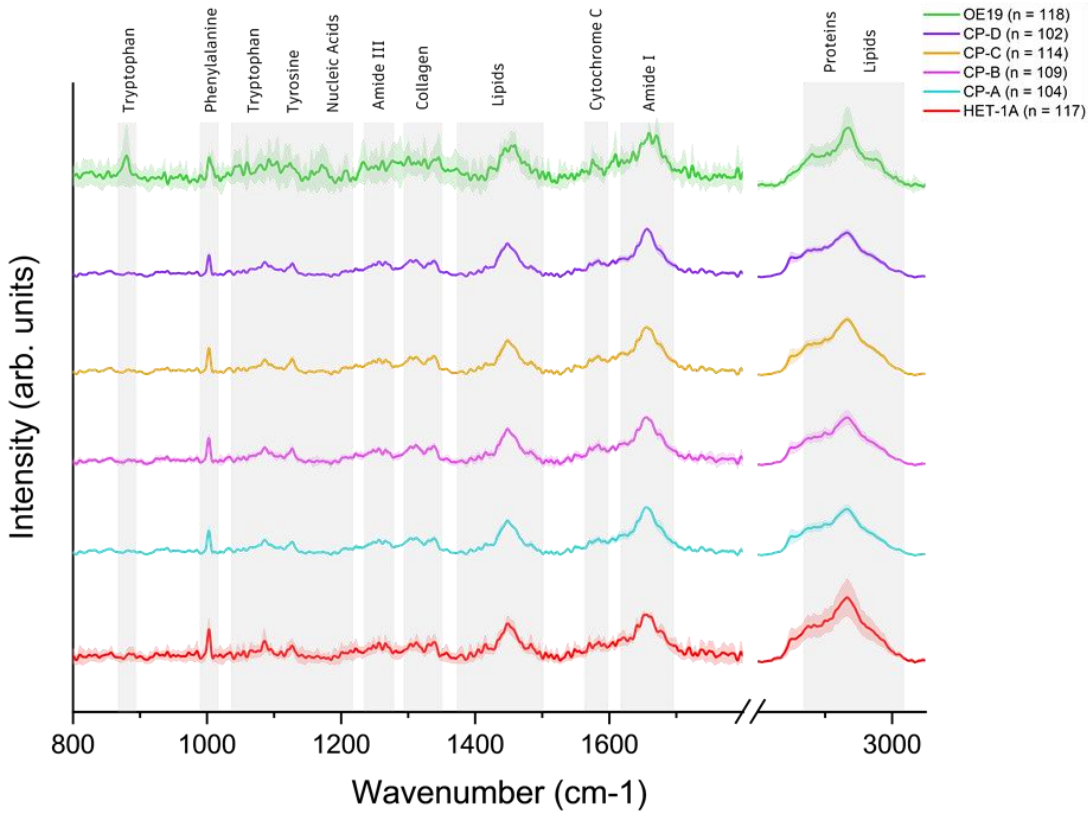

Figure 3.3: Single-cell Raman spectra representing the progression from healthy squamous oesophageal epithelia (HET-1A: red) through non-dysplastic Barrett's (CP-A: cyan), mild dysplastic Barrett's (CP-B: pink), moderate dysplastic Barrett's (CP-C: yellow) and severe dysplastic Barrett's (CP-D: purple), to oesophageal adenocarcinoma (OE19: green) cells. All spectra were normalised to the Amide I band at 1656 cm-1. The shaded area around each cell line is representative of the standard deviation. The grey shading across all cell lines highlights bands of interest (labelled). n = number of single cell Raman spectra. The high wavenumber region (CH2CH3) has been multiplied by a factor of 0.25 to fit on the same scale.

At an initial glance, all cell lines contain a prominent band corresponding to the presence of phenylalanine at 1004 cm^{-1}. The spectra for the oesophageal adenocarcinoma cell line (OE19), displays an increase in irregularity within peaks when compared against other stages of disease progression, alongside an increase aromatic amino acid presence (tryptophan (336, 337), tyrosine (246, 338-340) and phenylalainine (246, 339)). These components are associated with various cancers such as breast (341) and colorectal (342)**,** as well as with the acquisition of mutated transforming functions in cancer development (343)**.** An additional band also related to this function is the tryptophan shoulder shown at 1009 cm^{-1} (341, 343).

Nucleic acid and protein-related bands, such as Amide III and Amide I, are displayed across all six cell lines. DNA-associated bands, as seen at 1089 cm^{-1}, are common within transformed cancerous cells (246) with an increased DNA concentration associated with disease progression. An increase in protein signature is observed via the heightened CH_3 stretching (2881 cm^{-1} and 2980 cm^{-1}) band for the OE19 cancer cells. Such an increase in protein content is consistent with previously reported data (266),(344, 345).

CH_3 and CH_2 twisting modes previously attributed to collagen (346) can be observed between 1310 - 1340 cm^{-1} (238, 343). Lipid-associated bands are prominent in Figure 4.5, reported as being attributed to the acyl (C – C) backbone in lipids (238, 347), CH_3 rocking / C – O vibrations (266), bending modes of C – H in CH_2 moieties (348, 349) as well as methylene deformation (246). These changes in cellular lipid quantity are significant biomarkers of abnormal membrane composition, characteristic of neoplastic cells (350).

3.2.2 Healthy vs Adenocarcinoma

Analysis was carried out to determine if the OE19 oesophageal adenocarcinoma cell line could be distinguished from HET-1A healthy cells. PCA was conducted as an unsupervised technique with the goal of maximising and identifying the variance of the data along the principal components (PCs). These PCs can be visualised and plotted to identify features within the spectra that help distinguish the different cell lines.

PCA was undertaken using the first 15 PCs of a truncated Raman spectra (Figure 3.4a). Differences across the first 5 PCs, accounting for > 80 % variance, are shown in Figure 3.4b, with bands showing significant differences highlighted in blue. The cumulative variance and variance explained for all 15 PCs is shown in Figure 3.4c. Variation within the amino acid bands (tryptophan, tyrosine and cytosine) are seen in the first 5 PCs at 880 cm^{-1}, 1171 cm^{-1} and 1610 cm^{-1} respectively. Additional bands showing variation across all 5 PCs include DNA (1089 cm^{-1}), collagen (1310 – 1340 cm^{-1}), methylene deformation (1450 cm^{-1}) and Amide I (1656 cm^{-1}).

Further analysis was undertaken using LDA, a supervised technique used for classification, by maximising the separation between different classes within a dataset via their discriminant functions. The LDA was trained on 75% of the data and tested on the remaining 25% using a 4-fold cross-validation approach. The combined PCA-LDA was able to separate the datasets

into distinct classes with high accuracy. The histogram (Figure 3.4d) indicates a clear separation of both populations. It is also evident that the representative healthy cell line (HET-1A, red) provides a narrower distribution compared to the cancerous equivalent (OE19, green).

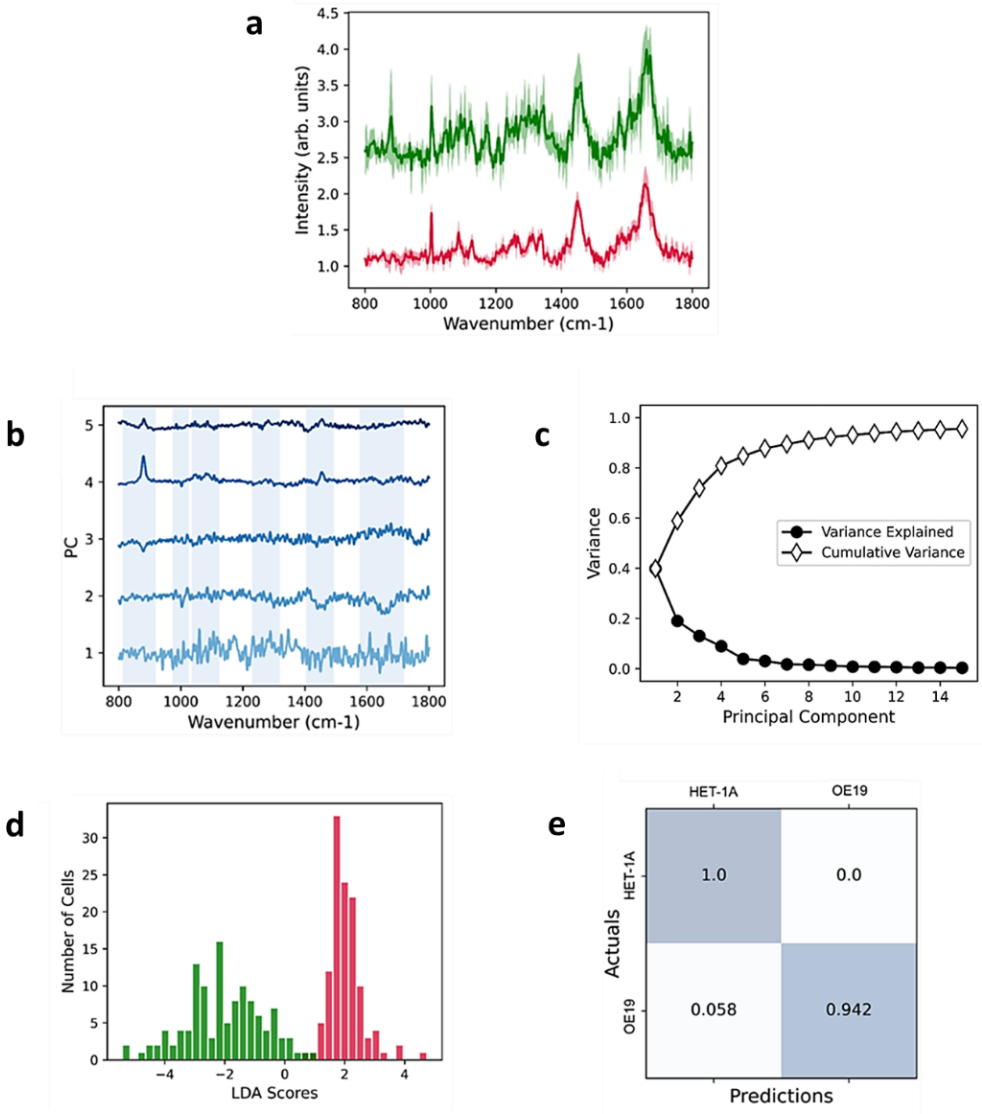

Figure 3.4: PCA-LDA of HET-1A (red) and OE19 (green) EMSC-corrected spectral data. (a) Truncated Raman datasets between 800 and 1800 cm−1 used for analysis. (b) Principal component loadings taken from PCA. Scores 1–5 are shown, accounting for > 85% variability. Bands of interest are highlighted in blue. (c) Cumulative variance and variance explained for the first 15 PCs. (d) LDA histogram displaying the spread of the scores for each dataset. (e) Confusion matrix for statistical analysis of the PCA-LDA. The outputted values are representative of the False Positive (0.058), True Positive (1), False Negative (0.942) and True Negative (0).

The confusion matrix obtained from this analysis is diplayed in Figure 3.4e. Whilst all HET-1A cells were correctly categorised, approximately 6 % of OE19 cells were mistaken for HET-1A, this is presumably a consequence of their heterogeneity. Precision [(true positive value / (true positive + false positive)], Recall [true positive value / (true positive + false negative)], and F1-Score [2 * ((precision * recall) / (precision + recall)), were determined to be 1, 0.972 and 0.942 respectively. The average accuracy for the classification of healthy vs. cancer single cells was 97 %.

3.2.3 Disease Progression

Intermediate between the healthy (HET-1A) and cancerous (OE19) stages are the progressive stages of Barrett's oesophagus, represented in these results by cell lines CP-A (non-dysplastic) and CP-D (severe dysplasia). Shown in Figure 3.5a**,** the spectra for each cell line was truncated to 800 – 1800 cm^{-1} prior to PCA-LDA. Analysis of the first 5 PC loadings (Figure 3.5b), covering 85% variance (Figure 3.5c), confirms the influence of amino acid bands, with phenylalanine and tryptophan bands present across all 5 PCs. Other cellular components such as Amide III, Lipids and Amide are present in PCs 1 and 5, PCs 2, 4 and 5, and PCs 2 and 5, respectively. Additionally, between disease stages, the Cytochrome C band (1583 cm^{-1}) is present within PC1 and PC2. This peak can be attributed to the presence of aromatic side chains within the Cytochrome C compound(244, 351, 352)**.**

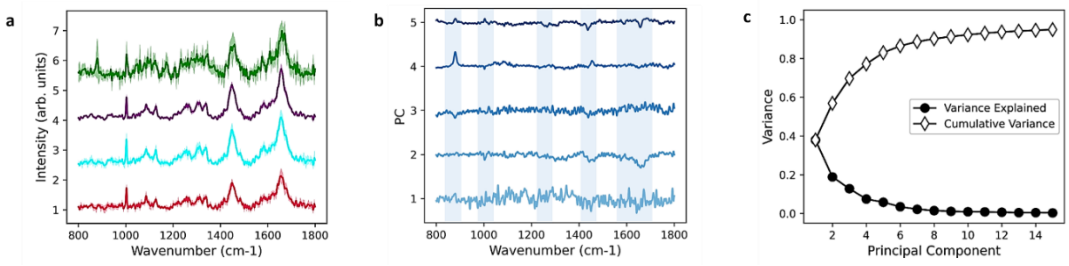

Figure 3.5: (a) Truncated Raman spectra (800 -1800 cm-1) for each stage of cancer development. The healthy (HET-1A), non-dysplastic (CP-A), severely dysplastic (CP-D) and cancerous (OE19) cell lines are represented by red, cyan, purple and green spectra, respectively. (b) The PC loadings from the data shown in Figure 4.7a. Bands of interest have been highlighted in blue. (c) 2D plot representing the variance explained and the cumulative variance for HET-1A (healthy), CP-A (non-dysplastic), CP-D (severely dysplastic) and OE19 (cancerous) single cells.

The 2D scatter plot for the LDA is displayed in Figure 3.6a, with HET-1A (red), CP-A (cyan), CP-D (purple), and OE19 (green), along with their respective histograms. From this scatter plot we see that the OE19 dataset is separated from the others and also shows greater scatter as might expected from a cancerous dataset with uncontrolled mutation. The HET-1A appears to be separated from CP-A and CP-D datasets in LDA2, whilst the CP-A and CP-D cells appear to exhibit strong overlap. This is further supported by looking at the confusion matrix in Figure 3.6b, which suggests that neither of the Barrett's Oesophagus cell lines (CP- A and CP-D) nor the healthy cell line (HET-1A) were mis-categorised as the cancer cell line (OE19). The HET-1A cells were correctly identified with ~91 % accuracy whilst the non-dysplastic (CP-A) and dysplastic (CP-D) cell lines were only correctly classified correctly 63 % and 64 % of the time. The largest errors occur between CP-A and CP-D cells, with ~36 % inaccuracy. The OE19 cells could be inaccurately categorised as any of the other three cell lines but with low probability – HET-1A (4 %), CP-A (6 %) and CP-D (2 %). The precision, recall and F1-score for this data is shown in Table 5.

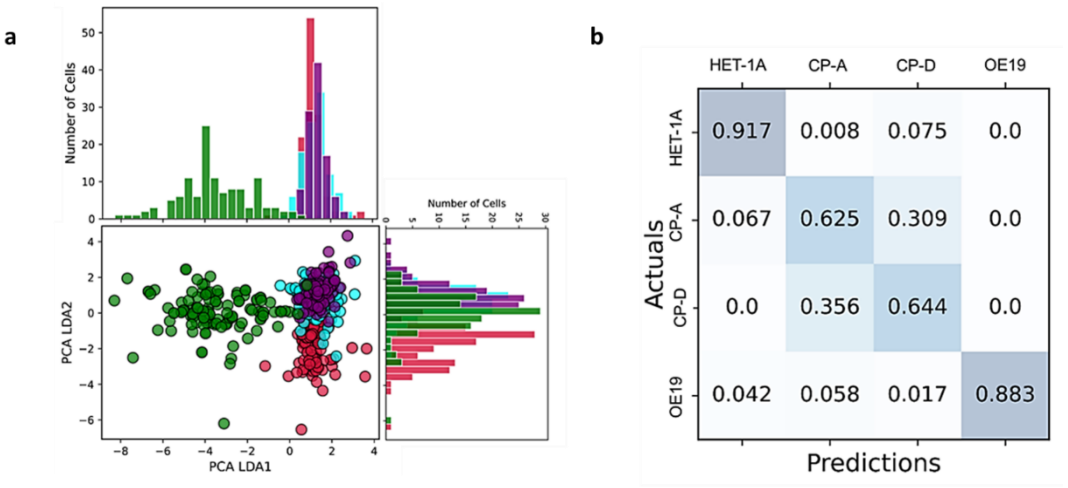

Figure 3.6: (a) 2D plotted PCA LDA1 vs PCA LDA2 scores for HET-1A (red), CP-A (cyan), CP-D (purple) and OE19 (green) and their relative LDA histograms. (b) Confusion matrix for statistical analysis of the PCA-LDA.

Cell line	Precision	Recall	F1-Score
HET-1A	0.917	0.917	0.905
CP-A	0.624	0.624	0.61
CP-D	0.644	0.644	0.63
OE19	0.833	0.833	0.938

Table 5: Precision, recall and F1-scores for 4-fold cross-validation analysis of the LDA scores for HET-1A (healthy), CP-A (non-dysplastic), CP-D (severely dysplastic), and OE19 (cancerous) single cells. The average model accuracy for this analysis was 88 %.

3.2.4 Dysplastic Progression

Next, the focus was on the analysis of cell lines representing high-grade Barrett's dysplasia. Identification of measurable differences in the spectral data could provide a promising route to the diagnosis of Barrett's Oesophagus severity in patients.

The PCA-LDA was carried out on the fingerprint region of the dysplastic cell lines, with the first 5 PC loadings in Figure 3.7b responsible for 80% of the variance, with the cumulative variance and variance explained detailed in Figure 3.7c. Negative bands in PC1 correlated with the aromatic amino acid content and Cytochrome C. These bands become positive between PC2 – PC4. Other notable bands present in PC1 are associated with Amide I, Amide III, lipid and collagen content. The PO^{2-} DNA-associated band (1089 cm^{-1}) also displays an increase in PC1.

The scatterplot, Figure 3.7a, displays how CP-B (pink), CP-C (orange) and CP-D (purple), are well separated by LDA1 and 2. The confusion matrix for this data is shown in Figure 3.7d, with precision, recall, and F1-scores detailed in Table 6. The average accuracy was calculated at 92 %, with PCA-LDA classification accuracy at 77 %. The PCA-LDA is able to correctly identify each stage of Barrett's dysplasia with accurate detection of CP-B at 80 %, CP-C at 93 % and CP-D at 91 %. With CP-B cells have a 13 % chance of being identified as CP-C cells.

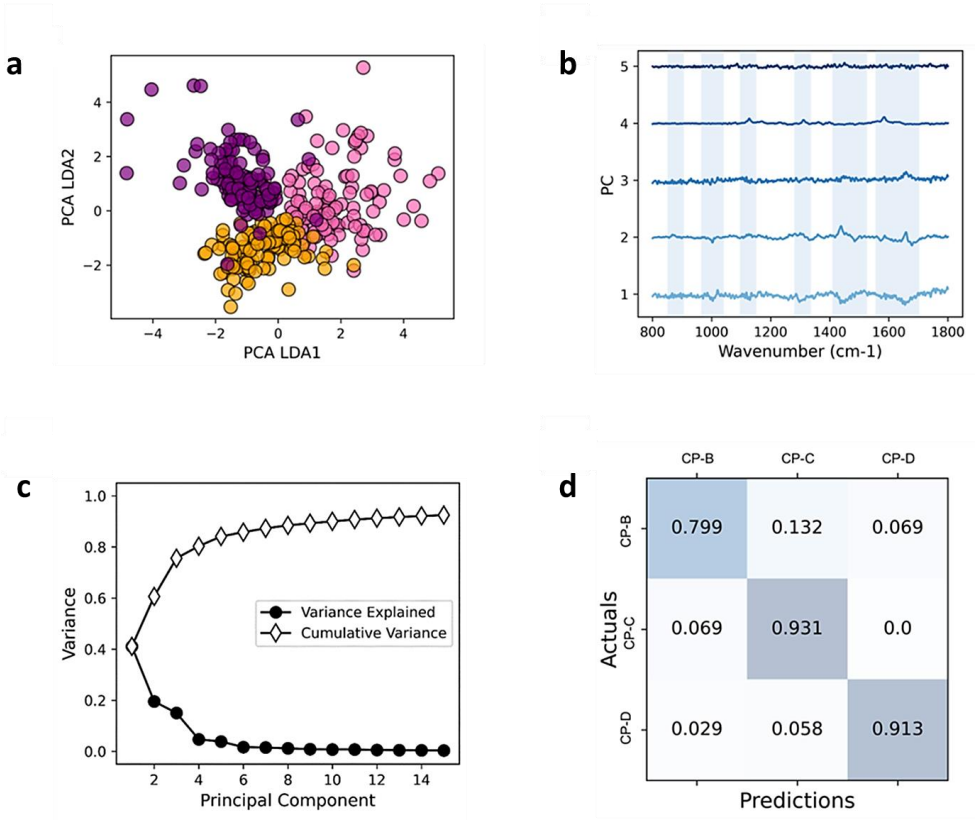

Figure 3.7: (a) 2D plotted PCA LDA1 vs PCA LDA2 scores for CP-B (pink), CP-C (orange) and CP-D (purple). (b) Principal component loadings taken from PCA. Scores 1-5 are shown, accounting for > 80 % variability. Bands of interest are highlighted in blue. (c) Cumulative variance and variance explained for the first 15 PCs. (d) Confusion matrix for statistical analysis of the PCA-LDA.

Cell line	Precision	Recall	F1-Score
CP-B	0.799	0.799	0.8424
CP-C	0.931	0.931	0.878
CP-D	0.913	0.913	0.921

Table 6: Precision, recall and F1-scores for 4-fold cross-validation analysis of the LDA scores for CP-B (mild dysplasia), CP-C (moderate dysplasia) and CP-D (severe dysplasia) single cells. The average accuracy for this classification model was 92 %.

3.3 Summary of Single Cell Analysis

In this chapter, it was shown that Raman spectroscopy can be utilised as a classification tool for determining disease stages within cancer progression. Distinct spectral signatures were exhibited as healthy cells progressed towards oesophageal adenocarcinoma, allowing us to observe changes within molecular components at each stage of disease development.

In the case of the OE19 cell line, peaks associated with changes in the DNA backbone geometry (Figure 3.4b) are indicative of phosphate ion interactions found in neoplastic cells(26, 344, 353), arising due to the erratic accumulation of DNA and RNA in the nuclei of cancer cells (354).

The increase of the aromatic amino acids associated with advanced disease progression is of note, with previous papers also highlighting the influence of these compounds in various cancers and diseases (341, 355). These compounds have been shown to play a significant role in orchestrating morphological changes within malignant tissues (356). Moreover, the increase in transcription and replication of cancerous cells requires a concomitantly increased volume of proteins for sustaining the excessive RNA synthesis. This increase in protein expression during cancer progression is expected due to an overall increase in protein overexpression with multiple oncogenic pathway activation and upregulation (266, 344, 345), as shown in Figure 3.3.

Increase in cellular lipid quantity across cell lines (Figure 3.3) are significant biomarkers of abnormal membrane composition, characteristic of neoplastic cells. These increases occur due to the modulation of specific enzymes and pathways within the cells, allowing for a higher proliferation rate, increased chemotherapy resistance and attenuation of the immune response (350).

This low risk of misclassification identifies a new method of detecting progression within Barrett's dysplasia at a single-cell level, which could allow earlier detection of progression towards cancer within a clinical setting. Disease progression classification has been conducted using PCA-LDA methods within other cancer sub-types such as oral cavity cancer (357) using statistical image analysis, and breast cancer (358), through sensor response curves, each derived from patient samples. They produced an accuracy classification of 98.2(±3.9) % and 95.4(±4.9) %, and 96.4% and 95.6%, respectively. The experimental results

of this work support the use of PCA-LDA as an effective and reliable tool for the diagnosis of cancer lesions.

The results from the work presented in this chapter not only highlight the use of single-cell Raman spectroscopy alongside PCA-LDA for the identification of healthy versus cancer cells but also for distinguishing between progressive disease stages within Barrett's oesophagus dysplasia itself.

4 The Influence of a Low pH Bile Salt Environment on Barrett's Oesophagus Single Cells

The previous chapter detailed the use of single cell Raman spectroscopy for characterising novel on-chip progression of oesophageal adenocarcinoma. The results showed successful characterisation of each disease stage using principal component analysis and linear discriminant analysis, with the ability to differentiate between healthy and cancerous cells with an accuracy of 97 %. Intra-class separation between dysplastic stages was determined with an accuracy of 68 %.

In an effort to identify changes in disease progression with increasing acid-bile exposure, this chapter highlights the impact of the increased exposure of Barrett's oesophagus single cells to an acidic bile salt solution (ABSS) over a prolonged duration. Bile is a complex digestive fluid primarily composed of bile acids, cholesterol, phospholipids, bilirubin, and electrolytes, synthesised in the liver and stored in the gallbladder before being secreted into the duodenum during digestion. Its principal role in digestion is to emulsify dietary lipids, facilitating their absorption in the small intestine (359). The amphipathic nature of bile acids allows them to interact with and disrupt lipid bilayers, which is particularly relevant in the context of oesophageal epithelium repeatedly exposed to bile during episodes of reflux. In vitro and in vivo studies have shown that bile acids can integrate into and destabilise cellular membranes, increase epithelial permeability, and induce oxidative stress and DNA damage (360).

Changes in cellular biochemistry were monitored on-chip using Raman spectroscopy, supplemented with morphological analysis via bright-field images. Live cells were trapped on-chip before being subject to low pH – bile salt conditions for a duration of 60 s. This was repeated for 10 exposures, with a 30-minute "recovery window" during which the pH was raised to neutral and data was collected. This recovery window is a dynamic period where pro-survival and pro-oncogenic signalling and repair mechanism pathways are activated, which in turn contribute to disease progression. Previous studies have reported recovery windows, albeit each study having a different duration (361, 362). Representative Barrett's

cell lines CP-A, CP-B, CP-C, and CP-D were utilised, with all data averaged over 10+ experimental repeats.

4.1 Single-Cell Viability: On-Chip vs. Acidic Bile

Whilst there is significant focus on replicating true phenotypic conditions *in vitro*, model refluxate conditions in the literature show high variation in exposure time and conditions, a large majority of research has been conducted on 2D and 3D cultures as well as patient tissues using solutions at pH 4.0 (361, 363-369). Ambulatory pH monitoring reports show that pH levels can drop as low as 4.0 for approximately 20% (< 4 hours) of the day (370-372) with most reflux episodes lasting ten minutes or less (373). For most Barrett's patients, the bile salt concentration can vary between 0.03 and 0.82 mmol/L (374-376).

To ensure that any biochemical changes within the single cells was due to external stressors (ABSS) and not continuous exposure to the 532 nm Raman laser, on-chip viability was monitored for 100 single cells per cell line. Each cell underwent a 60 s exposure to the Raman laser (2.4.1) every 30 minutes, over a total duration of 5 hours. Cell death was determined using Sytox Red (2.1), with average viability plotted in Figure 4.1. No significant changes were observed in the Raman spectra, an example for CP-A shown in Figure 4.3, indicating that the cells remained viable and healthy on the chip. This suggests that the experimental conditions did not induce stress or excessive cell death, allowing the cells to maintain their normal physiological state.

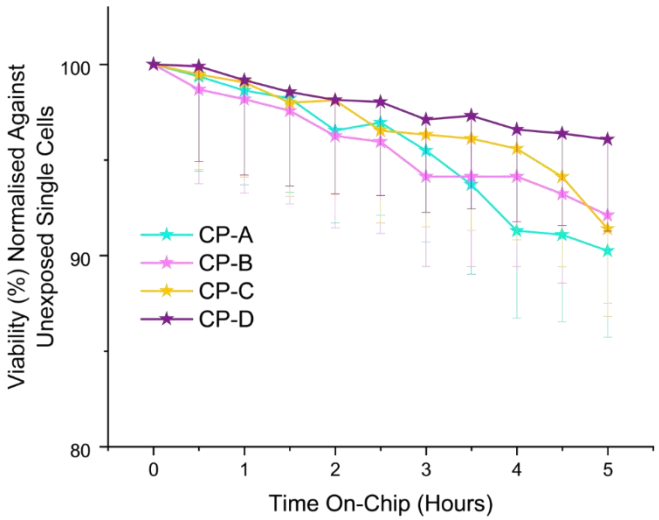

Figure 4.1: Line graph displaying the average viability of 50 single cells exposed to a Raman laser (2.4.1) every 30 minutes, to a total of 10 Raman laser exposures. Values were normalised against the average viability of trapped single cells on-chip which were not exposed to the Raman laser.

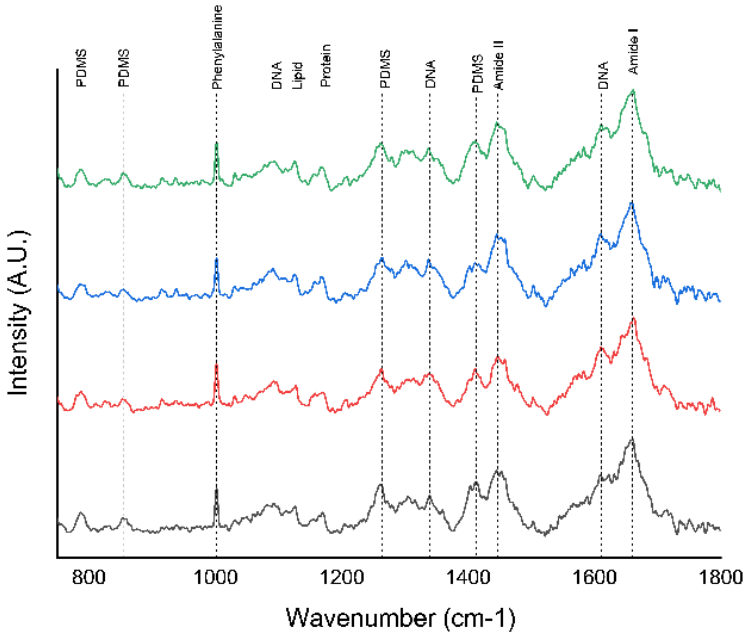

Figure 4.2: Offset Raman spectra of a CP-A cell are shown at four time points: immediately before laser exposure (black), post-2 hours (red), post-4 hours (blue), and post-5 hours (green). The study aimed to assess whether a single Raman laser exposure could induce delayed apoptosis by monitoring untreated cells over time. Each spectrum represents the average of 100 individual spectra, all of which were baseline-corrected and smoothed using the Savitzky-Golay method for noise reduction.

In order to remain as physiologically accurate as possible, initial experiments focused on monitoring changes over a singular 10 minute ABSS exposure. However, due to the fragility of isolated single cells, the prolonged exposure to a low pH solution led to cellular damage beyond recovery and repair. This was observed through both morphological changes (Figure 4.2), where signs of damage were observed through blebbing and irregularity of the lipid membrane, however this caused extreme disintegration of the cellular structure, notably between 1.5 and 2 hours (Figure 4.3e). Immediately before exposure, the CP-A cell appears healthy and structurally intact, with a smooth and uniform morphology (Figure 4.3a). There seem to be no visible signs of damage or stress, with this image providing a baseline reference point for assessing the effect of ABSS exposure. By 0.5 hours post-exposure (Figure 4.3b), there are early signs of morphological stress, with slight irregularities in the cell membrane. Internal structures also appear darker and more pronounced, possibly due to internal swelling or cytoplasmic reorganisation. These changes could be indicative of the cell attempting to repair damage caused by the ABSS exposure or adapting to the acidic stress. Figure 4.4c presents the cell at 1 hour post-exposure, where there are more pronounced morphological changes such as the cell membrane becoming visibly distorted, with an uneven surface. The internal structures, previously well defined, are now less prominent. This could be due to ongoing damage, or the process of necrotic or apoptotic cell death. The morphology deteriorates by 1.5 hours post-exposure, (Figure 4.3d) where the internal structures appear disorganised and the membrane has fragmented and partially collapsed. By 2 hours post exposure (Figure 4.3e), the cell is completely disintegrated.

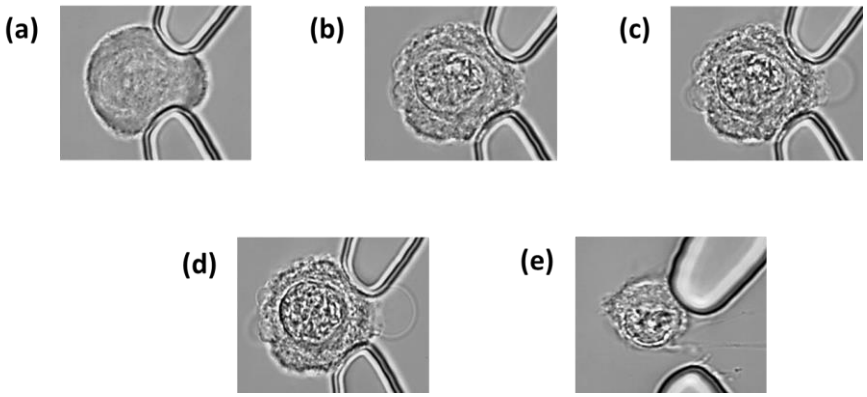

Figure 4.3: Bright-field images of a trapped CP-A cell on-chip, illustrating morphological changes at various time points: (A) immediately before exposure, (B) 0.5 hours post-exposure, (C) 1 hour post-exposure, (D) 1.5 hours post-exposure, and (E) 2 hours post-exposure.

The experiment aimed to observe biochemical changes within single cells during recovery after recurrent acid bile exposure, necessitating that the cells remain intact for long-term monitoring. For this reason, varying ABSS exposures were tested to identify conditions that would trigger biochemical recovery changes in the cells without over-compromising overall viability. To ensure an adequately sized dataset for analysis, a minimum final viability of 25 % was required.

Single Barrett's oesophagus cells were subjected to a pH 4.0 ABSS for 1, 2, 5 and 10 minutes, to give 10, 5, 2 and 1 exposure(s) respectively, totalling a final exposure duration of 10 minutes. Exposures were separated by a 30-minute recovery period at pH 7.0, with cell media ran through the device to remove any ABSS residue. Shown across Figures 4.3, 4.4 and 4.5, increasing the duration of exposure to ABSS at pH 4.0 caused a drastic reduction in cellular viability.

The bar graphs shown in the right-hand column of Figure 4.4 provide a clearer visual representation of differences in cellular viability across the CP cell lines. CP-C demonstrated the highest percentage of dead cells following each exposure to the ABSS. This pattern indicates that CP-C cells are particularly vulnerable to environmental stressors, including repeated exposures of a low pH solution, and their viability decreases sharply as the number of exposures increases.

The CP-A single cells exhibit a similar trend, with the second-highest mortality rate. Interestingly, after six exposures, the viability of CP-A single cells was comparable to that of CP-C single cells, suggesting a possible threshold beyond which these cells may lose their ability to maintain homeostasis. The steep decline in viability for CP-A cells is indicative of a high sensitivity to ABSS conditions, despite initially appearing to be more robust than CP-C.

Contrasting this, the CP-D cell line displays a gradual, more linear decline in viability over each exposure, possibly attributed to CP-D's classification as a severely dysplastic Barret's cell line, which is a more transformed cell type than CP-A or CP-C. These cells may have altered adaptations, metabolic or structural, which make them less immediately susceptible to extracellular environmental changes. CP-B cells displayed the greatest viability amongst the CP- cell lines, with the lowest level of recorded cell death. By the final exposure, CP-B cells had an overall viability of 56 %, significantly higher than the other cell lines. This could suggest that CP-B cells have intrinsic properties or mechanisms that provide them with greater resistance to the damaging effects of ABSS exposure.

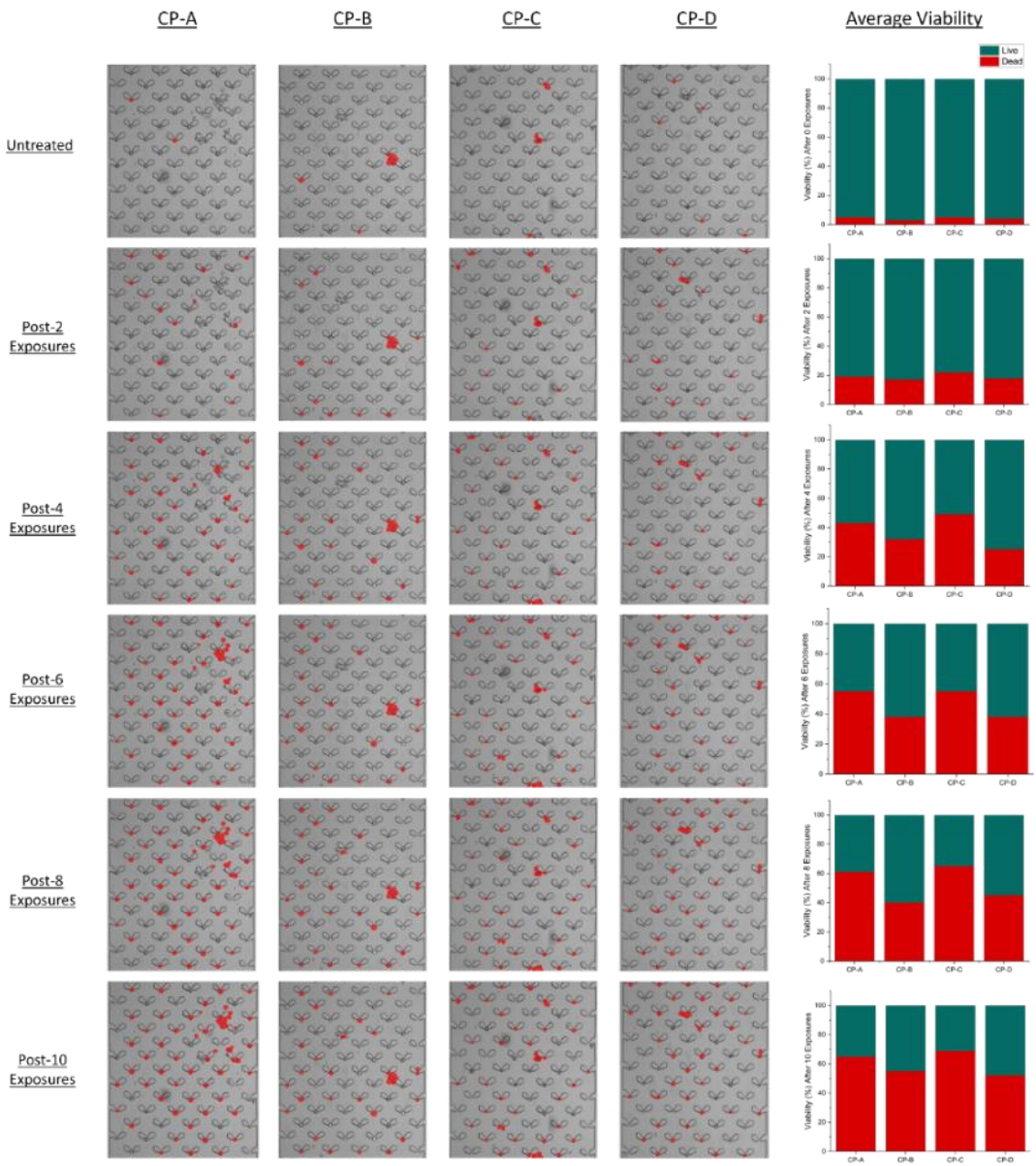

Figure 4.4: Overlapped bright field and fluorescence images of single cells trapped on-chip. All images were taken on a 10x dry lens. Each column (left to right) is representative of the average viability of cell lines CP-A, CP-B, CP-C and CP-D. The top row displays results from cells that are untreated, with the remaining rows (top to bottom) displaying cells that have been treated with 2, 4, 6, 8, and 10 acidic-bile salt solution exposures. Viability was determined via Sytox Red dead stain fluorescence imaging. The right-hand panel displays the percentage viability for each cell line, calculated over 10+ independent experiments, for >100 single cells. The red bars are indicative of increasing cellular death, whilst the blue bars are indicative of respective cell viability %.

The comparison between cell lines reveals that the CP-C cell line shows the highest % of dead cells after each exposure. CP-A single cells had the second highest rate of mortality, with the cellular viability equal to CP-C single cells after exposure 6. The severely dysplastic Barrett's cell line CP-D shows a linear decrease in viability, possibly attributed to being the most transformed cell type, and therefore less sensitive to extra-cellular environmental changes. CP-B single cells had the least cell death compared to all cell lines, with a final overall viability of 56 %. When comparing the viability shown in Figure 4.6 against the line graphs in Figure 4.4, it is immediately visible that both CP-A and CP-C Barrett's single cells show a drastic reduction in cellular viability after ABSS exposure. When compared against CP-B and CP-D Barrett's, the death rate was 1.5 x higher after the 4th exposure.

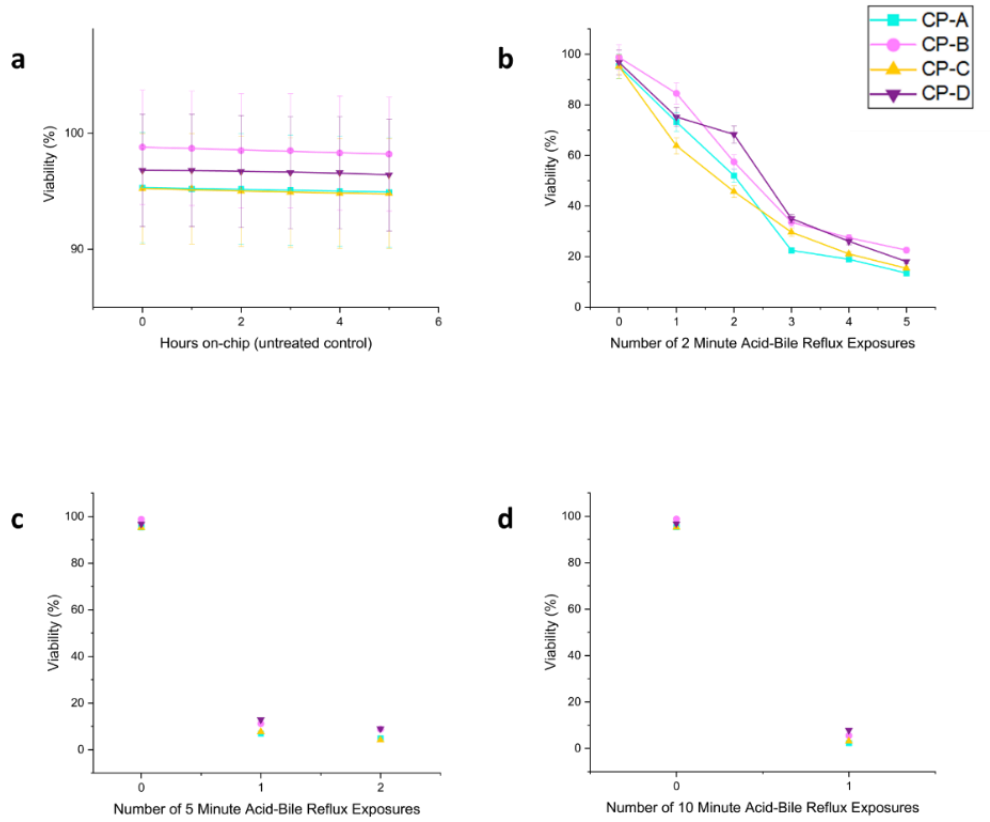

Figure 4.5: On-chip viability data for CP-A (cyan), CP-B (pink), CP-C (mustard) and CP-D (purple) Barrett's single cells. Cells were subject to 2 (b), 5 (c) and 10 (d) minute durations of pH 4.0 acidic-bile salt solution exposure, to give an overall exposure time of 10 minutes. Untreated cells were monitored on-chip for a total duration of 5 hours (a). The pH was raised to 7.0 post-exposure with the addition of serum-free media. The duration between each exposure was 30 minutes. Viability was determined via Sytox Red dead stain fluorescence imaging. Each value is the average of >100 single cells, taken over 10+ independent experiments.

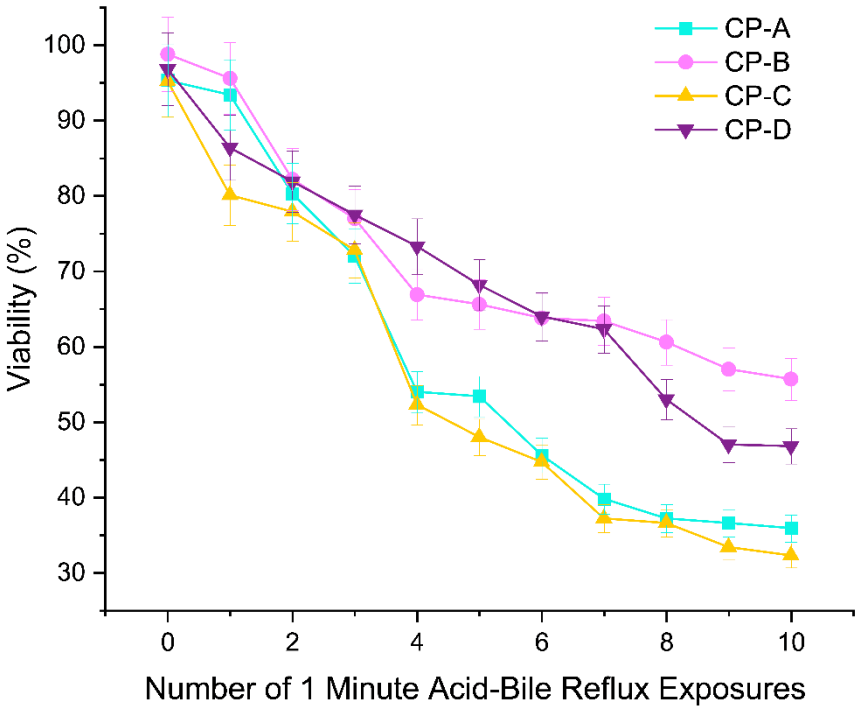

Figure 4.6: CP-A (cyan), CP-B (pink), CP-C (mustard) and CP-D (purple) Barrett's single cells were subject to 10 x 1-minute exposures of pH 4.0 acidic-bile salt solution, giving a total exposure duration of 10 minutes. The pH was raised to 7.0 post-exposure with the addition of serum-free media. The time between each exposure was 30 minutes. Viability was determined via Sytox Red dead stain fluorescence imaging. Each value shown is the average of >100 single cells, taken over 10+ independent experiments

All cell lines, CP-A (cyan), CP-B (pink), CP-C (mustard) and CP-D (purple), show a steady decrease in viability over multiple 1 minute (Figure 4.6) and 2 minute (Figure 4.5) exposures with a visible decline seen after 4 and 2 exposures, respectively. All cells exposed to ABSS for 5 minutes (Figure 4.5c) and 10 minutes (Figure 4.5d) had < 10 % viability after a single exposure. Final viability for all cell lines after 5 x 2 minute exposures was > 25 %, with final viability for 10 x 1 minute (Figure 4.6) exposures exceeding 30 % for all cell lines. For this reason, 1-minute exposures were chosen as the optimal experimental conditions.

4.2 Acid-bile salt exposure of Barrett's oesophagus cells

Single- cell Raman spectra were acquired from individual CP-A, -B, -C, and -D, Barrett's cells which were trapped on-chip. Data was collected following a 1-minute ABSS exposure, during the 30 minute recovery phase after the pH was restored back to 7. The aim of this research was to investigate what biochemical changes occur as part of BO cells' immediate response to fluctuations in pH, namely acidic pH. The average spectra for 50 single cells, per cell line, and per exposure, is presented in Figure 4.7. All spectral data were pre-processed as detailed in 2.5.1, including EMSC background removal and normalisation to the Amide I peak (1656 cm^{-1}).

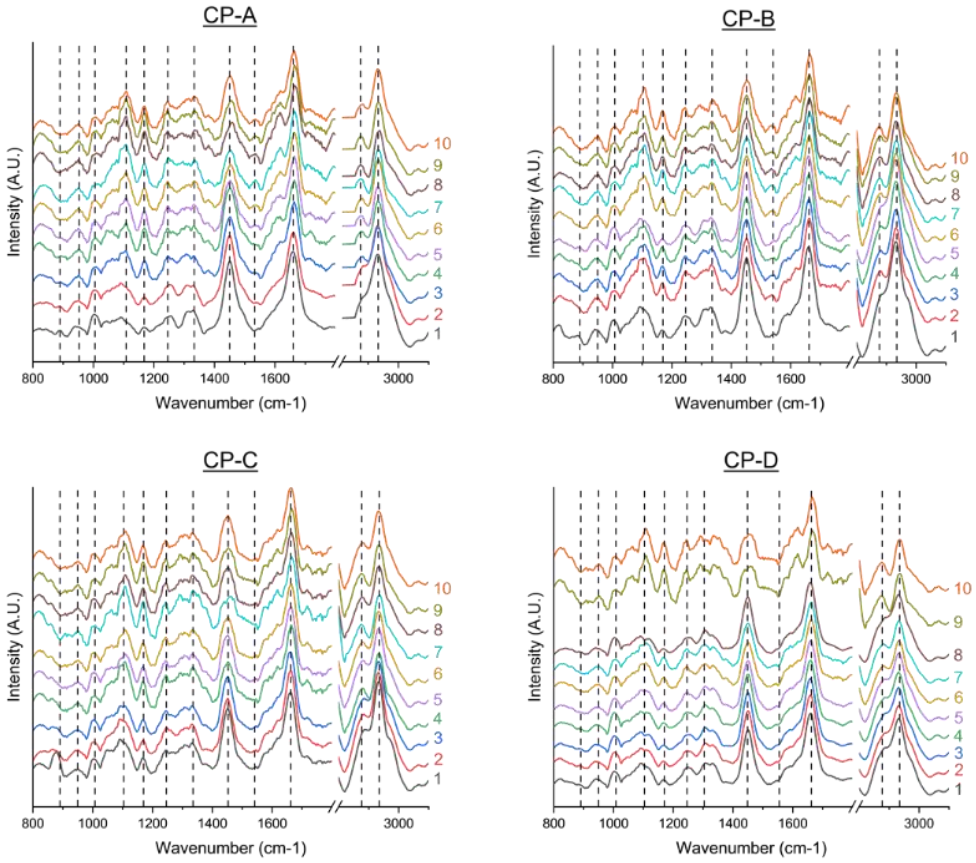

Figure 4.7: Raman spectra for Barrett's oesophageal cell lines CP-A (A), CP-B (B), CP-C (C), and CP-D (D). Single cells were subjected to 10 individual 1-minute exposures to acidic-bile salt solution at pH 4.0, resulting in a total exposure duration of 10 minutes. Following each exposure, the pH was neutralised to 7.0 with serum-free media, with a 30-minute recovery period between exposures. Spectra were collected during the neutral pH recovery phase. The average of 50 spectra per condition is displayed, offset by 0.1 units for clarity, with colours representing ascending exposures: exposure 1 (black), exposure 2 (red), exposure 3 (blue), exposure 4 (green), exposure 5 (purple), exposure 6 (mustard), exposure 7 (cyan), exposure 8 (brown), exposure 9 (olive), and exposure 10 (orange).

When looking at the Raman bands displayed for CP-A single cells, there are prominent biochemical changes occurring over recurring ABSS exposures. The band at 880 cm^{-1}, assigned to tryptophan vibrations, reflects symmetric stretching associated with protein stability and cellular stress response. As the number of exposures increase, the phenylalanine symmetric ring breathing band at 1004 cm^{-1}, suggests that there is retained protein conformation, despite the environmental stress. Lipid-associated peaks at 2850 cm^{-1} (CH_2 symmetric stretch of lipids) and 2881 cm^{-1} (CH_2 asymmetric stretch), show a gradual decrease in intensity, possibly indicative of lipid peroxidation or lipid membrane destabilisation (246, 377). This would suggest that CP-A cells experience membrane integrity issues over a prolonged exposure of ABSS.

In contrast, CP-B single cells have fewer intensity changes than CP-A. The glycogen band at 1122 cm^{-1} remains constant, similarly the 1312 cm^{-1} CH_3/CH_2 twisting mode in lipids demonstrates greater lipid configuration stability. The amide I band (1656 cm^{-1}) has minimal variation, demonstrating protein structural integrity. The 1171 cm^{-1} tyrosine and collagen band present in CP-C single cells has altered intensity, potentially reflective of stress-induced protein degradation/modifications (244, 363, 364). Lipid peaks at 1336 cm^{-1} (C–H deformation) and 1450 cm^{-1} (CH_2 bending) have lower intensities, possibly due to lipid bilayer disruption or loss of integrity. The 1570 cm^{-1} (COO^- stretch) band has a slight shift, indicating changes in the carboxylate group environments linked to cellular metabolic stress.

The CP-D single cells appear to show the most stable biochemical response out of all four Barrett's cell lines. The 1089 cm^{-1} DNA-associated band (PO_2^- vibrations) is largely unchanged, which indicates good genomic stability despite ABSS exposure. Additionally, the 1206 cm^{-1} hydroxyproline and tyrosine band remains at the same intensity, suggestive of preserved protein and collagen structures. There are minor reductions in the lipid-associated band at 2935 cm^{-1} (CH3 symmetric stretch) band, however the overall lipid profile for CP-D remains less affected when compared against cell lines CP-A and CP-C. The presence of the amide I band (1656 cm^{-1}), indicates that there is maintained protein structural integrity which could explain the linear decrease in viability, shown in Figure 4.6, as opposed to the sharp decline observed in the other cell lines.

This spectral data is further detailed in a heatmap, as shown in Figure 4.8. This figure provides a visual representation of trends in intensity changes over time, making it easier to observe progressive shifts in specific spectral regions.

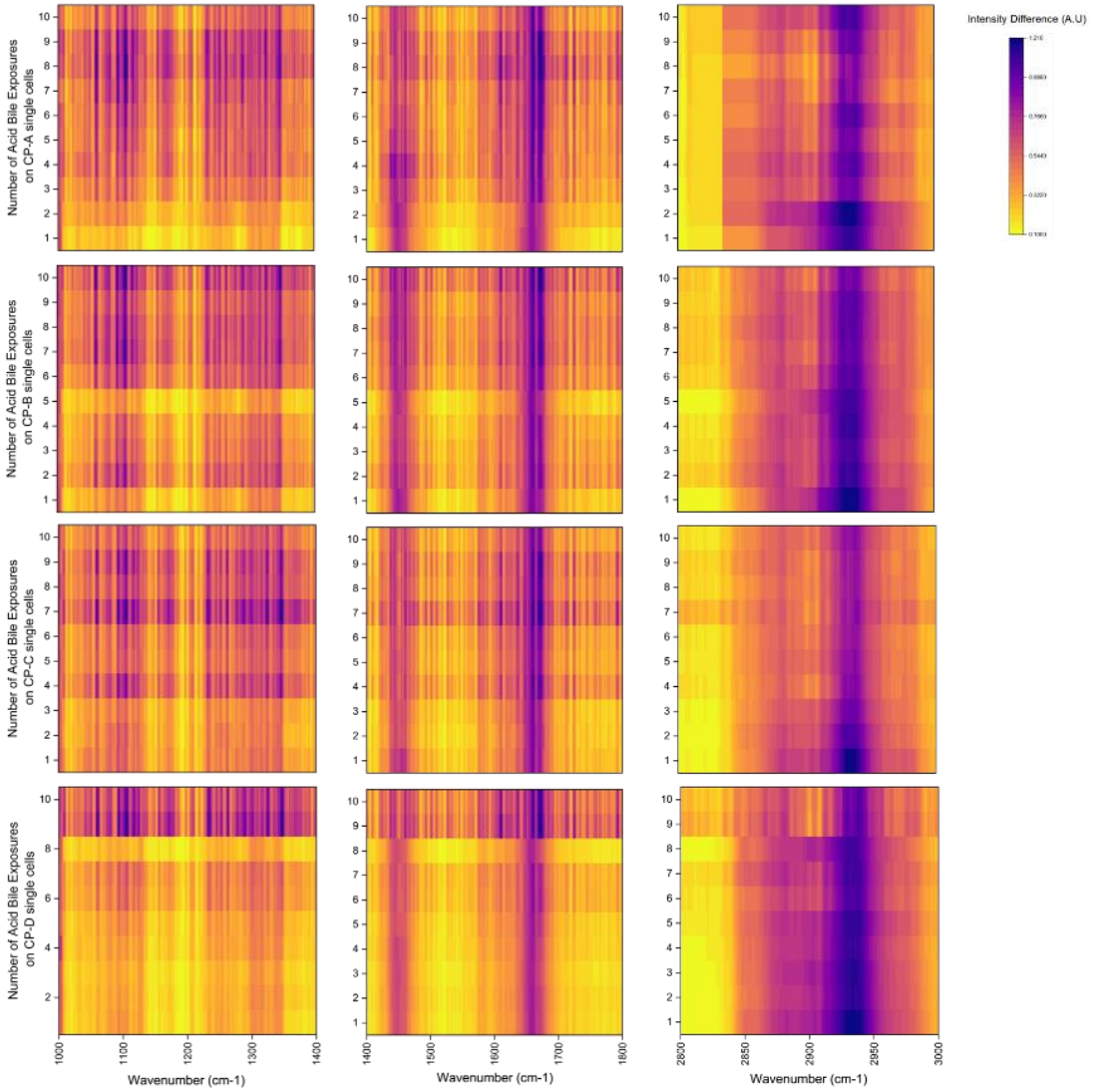

Figure 4.8: Heatmaps showing changes in spectral intensity for CP-A (top row), CP-B (second row), CP-C (third row), and CP-D (bottom row) cell lines. Spectral regions are displayed as follows: 1000–1400 cm^{-1} (left column), 1400–1800 cm^{-1} (middle column), and 2800–3000 cm^{-1} (right column). Each intensity value represents the average of 50 individual spectra per cell line under each condition. Cells were exposed to 60-second treatments with pH 4.0 acidic-bile salt solution, with the total number of exposures indicated on the left side of each heatmap. Data was collected during a 30-minute recovery period, following neutralisation to pH 7.0 with serum-free media.

Whilst heatmap analysis can provide a broad overview of intensity changes across spectral regions, the Raman spectra shown in Figure 4.9 offers a more detailed explanation into biochemical alterations occurring in the single cells after repeated ABSS exposure, compared to their untreated counterpart.

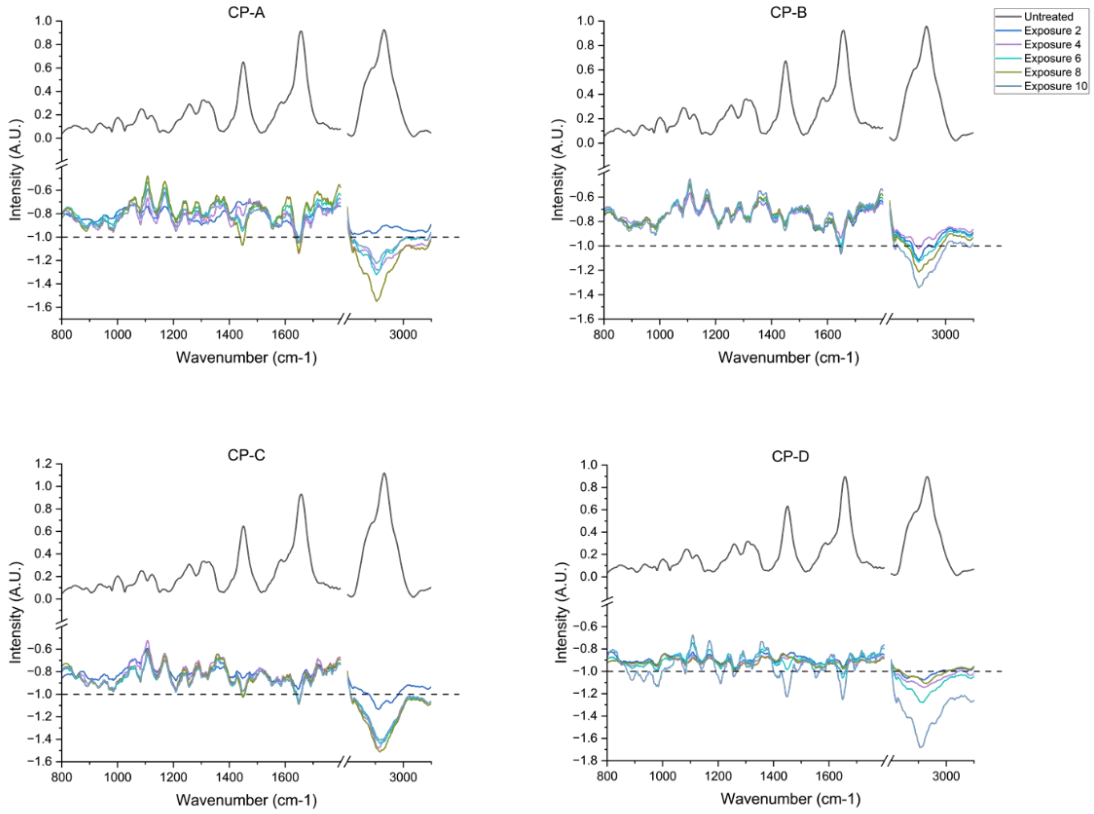

Figure 4.9: Raman spectra for CP-A (a), CP-B (b), CP-C (c) and CP-D (d). The upper trace (black) displays the average spectrum for untreated single cells. The lower trace displays the average spectra for single cells after 2 (blue), 4 (green), 6 (light purple), 8 (teal) and 10 (dark purple) exposures of ABSS, plotted as a function of peak differences against the untreated spectra. Each spectrum shown is the average of 50 measurements taken at pH 7.0. All spectra were normalised to the Amide I band (1656 cm^{-1}) prior to subtraction from untreated single-cell data to show the peak difference over multiple exposures. A dotted line on the upper trace marks Y(0). The high wavenumber region (CH_2CH_3) was multiplied by a factor of 0.25 to fit on the same scale.

The upper black trace in each panel is representative of the average spectrum of untreated cells for respective cell line, serving as a baseline for comparison against the treated spectra after 2 (blue), 4 (green), 6 (light purple), 8 (teal) and 10 (dark purple) exposures of ABSS. Each spectrum shown is the average of 50 spectra, per cell line, per condition. All bands were pre-processed using EMSC and normalised to the Amide I peak at 1656 cm^{-1}. The lower trace (untreated cell spectra) has been offset, with a dotted line plotted across Y = -1 to visualise negative and positive peaks in the upper traces.

When analysing CP-A, the spectra show noticeable differences in the lipid-related CH_2/CH_3 stretching region (2800 – 3200 cm^{-1}) (238, 349, 360-362). After multiple exposures, the bands present in this region become less intense relative to the untreated cells, indicating a loss or reorganisation of membrane lipids (346). The band around 1089 cm^{-1}, associated with DNA phosphate vibrations (246), has a gradual decrease similar to the lipids, also suggesting a reduction in nucleic acid integrity or changes in chromatin structure.

The differences in CP-B are less intense than CP-A, however there is a consistent reduction in the amide I band (1656 cm^{-1}) which may be associated with protein denaturation / structural changes (267, 344, 345). The DNA-associated band at 1089 cm^{-1} remains stable across all exposures, indicating that CP-B single cells were able to maintain genomic stability under stress (343). These changes align with the previous higher viability data, indicating a more robust response to ABSS compared to other cell lines.

CP-C cells display significant changes associated with the amide I and lipid-stretching region. CH_2 and CH_3 vibrational bands increase in intensity over multiple exposures, suggesting lipid damage is a large factor in the stress response of CP-C. The DNA-phosphate band at 1089 cm^{-1} shows a decrease, pointing to less genomic stability than that of CP-B. The CP-D dataset displays a decrease in intensity for lipid-associated peaks at 2800 – 3200 cm^{-1}, despite the amide I and DNA-related bands remaining stable over multiple exposures.

Peak Position (cm^{-1})	Assignment	Vibrational Mode
880	Tryptophan	Ring breathing
1004	Phenylalanine	Symmetric ring breathing
1033	Phenylalanine	C–H in-plane bending
1089	PO_2^- in DNA	Symmetric stretching
1122	Glycogen	C–O stretching
1129	Skeletal of acyl backbone	C–C stretching
1171	Tyrosine in collagen	C–H bending
1206	Hydroxyproline, tyrosine	C–H bending
1312	CH_3/CH_2 twisting mode of lipids	CH_2 twisting
1336	Lipid C–H deformation	CH_2 wagging
1416	C–C stretching in quinoid ring	C=C stretching
1450	CH_2 bending	CH_2 scissoring
1570	COO^- stretching	Asymmetric COO^- stretching
1584	C–C stretch	Aromatic ring stretching
1610	Cytosine (NH_2)	N–H bending
1656	Amide I	C=O stretching
2850	Lipids	CH_2 symmetric stretching
2881	Lipids and proteins	CH_2 asymmetric stretching
2935	Chain-end CH_3 symmetric band	CH_3 symmetric stretching
2980	Asymmetric CH_3 stretch (protein)	CH_3 asymmetric stretching

Table 7: Raman peak assignments for biochemical components observed in Barrett's oesophagus cell lines undergoing ABSS exposure. Peaks are listed with their corresponding assignments and vibrational modes.

4.3 Real time changes in Barrett's oesophagus morphology

Bright field and fluorescent images were taken after each ABSS exposure for monitoring morphological changes within Barrett's oesophagus cells due to pH fluctuations. Cellular viability was determined after each exposure via Sytox Red dead stain to ensure that all observed cells were viable (non-apoptotic) and therefore capable of recovery and repair after exposure to low pH. For conciseness, 10 single cells are shown both prior to exposure, and after 2, 4, 6, 8 and 10 exposures for CP-A in Figure 4.10. Images for CP-B, CP-C and CP-D single cells are highlighted in Figures 4.11, 4.12, and 4.13, respectively. The membrane of each cell is outlined to highlight changes occurring over time.

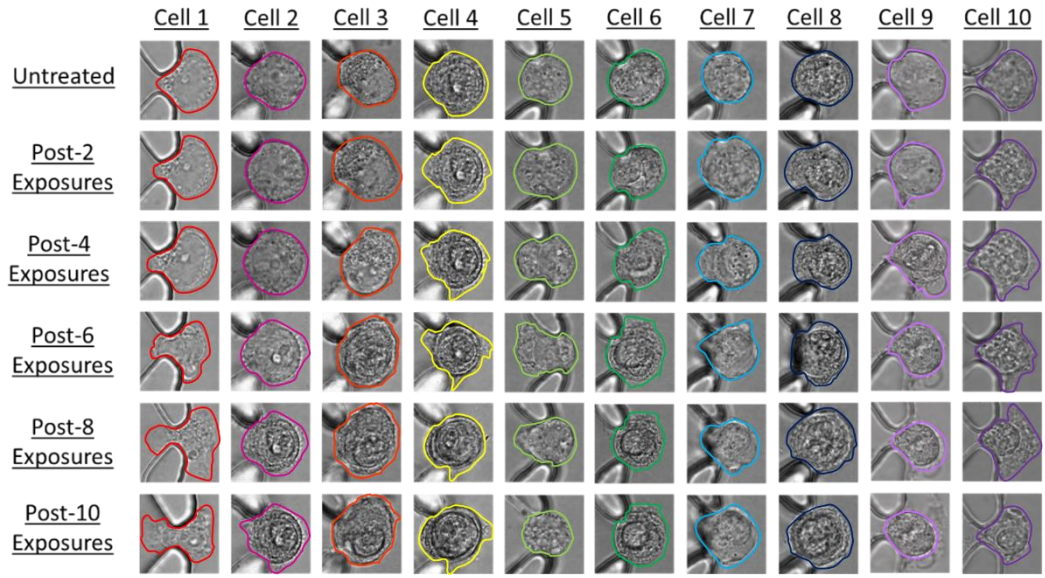

Figure 4.10: Bright-field images illustrating the morphological changes in CP-A cells subjected to pH 4 acidic-bile salt solution. Cells were exposed for 60 seconds to low pH conditions, followed by restoration to pH 7.0 for 30-minute recovery intervals. Images were captured after 0, 2, 4, 6, 8, and 10 exposures. The outlines of cell membranes are highlighted to emphasise changes in shape, size, and integrity. Significant structural degradation is observed with increasing exposures, correlating with reduced viability as determined by Sytox Red dead stain fluorescence imaging.

At 0 exposures, CP-A single cells (FFigure 4.10) display a round and smooth morphology, typical of healthy and viable cells. After 2 and 4 exposures, there are subtle changes in morphology. Whilst the cells remain rounded, there are slight irregularities within the membrane that become visible. These changes may represent the onset of the stress

response, potentially cytoskeletal or membrane reorganisation as the cells adapt to the low pH environment. Despite these changes, the cells seem to remain structurally stable. As they experience 6 exposures, there are more pronounced changes occurring – membrane irregularities are prominent with signs of blebbing and elongation appearing. These alterations in the cellular structure suggest that prolonged exposure to ABSS will lead to compromising on cellular integrity. The changes at this stage could be linked to increased strain on the membrane and cytoskeleton as the cells struggle to maintain homeostasis under acidic conditions. Reaching 8 and 10 exposures, the CP-A cells exhibit significant morphological degradation. The membranes begin fragmenting and distorting, with a large loss of viability. Shrinkage and membrane disruption is evident, with the extent of damage suggesting that the cells are no longer able to recover during the neutral pH recovery intervals.

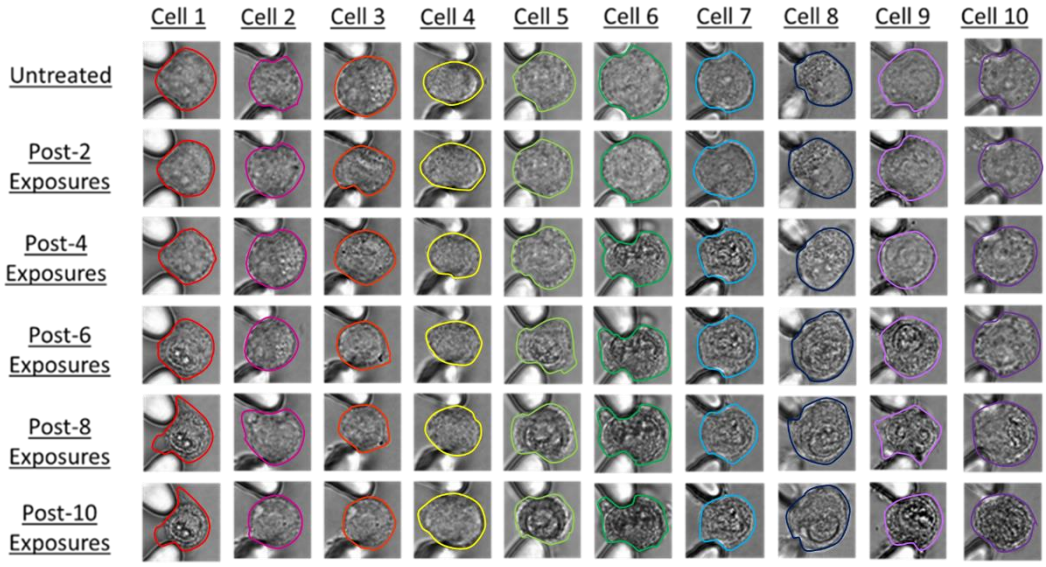

Figure 4.11: Bright-field images illustrating the morphological changes in CP-B cells subjected to pH 4 acidic-bile salt solution. Cells were exposed for 60 seconds to low pH conditions, followed by restoration to pH 7.0 for 30-minute recovery intervals. Images were captured after 0, 2, 4, 6, 8, and 10 exposures. The outlines of cell membranes are highlighted to emphasise changes in shape, size, and integrity. Significant structural degradation is observed with increasing exposures, correlating with reduced viability as determined by Sytox Red dead stain fluorescence imaging.

CP-B cells exhibit subtle but progressive morphological changes across increasing exposures. They are able to maintain their round uniform structure across exposures 2 and 4, however,

by the 6th exposure, the presence of membrane irregularities becomes apparent. These increase with further exposures, with noticeable blebbing observed after exposures 8 and 10. The ability of CP-B cells to retain consistent morphology up to the 6th exposure may indicate a moderate resistance to pH-induced stress compared to the other CP- Barrett's-derived lines. Their eventual breakdown at higher exposures however, suggests a limited capacity to withstand cumulative stress, albeit slower than CP-A.

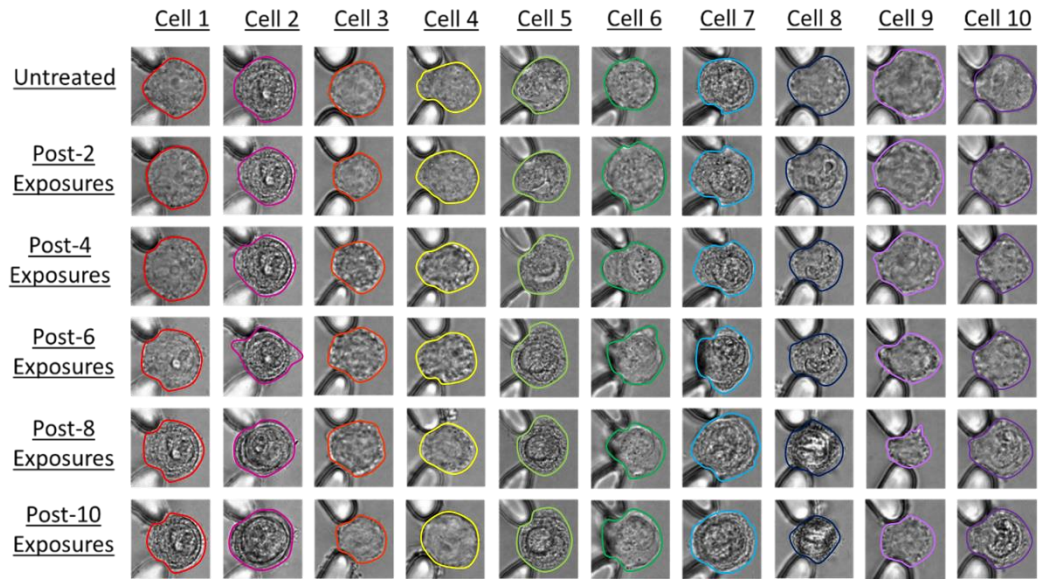

Figure 4.12: Bright-field images illustrating the morphological changes in CP-C cells subjected to pH 4 acidic-bile salt solution. Cells were exposed for 60 seconds to low pH conditions, followed by restoration to pH 7.0 for 30-minute recovery intervals. Images were captured after 0, 2, 4, 6, 8, and 10 exposures. The outlines of cell membranes are highlighted to emphasise changes in shape, size, and integrity. Significant structural degradation is observed with increasing exposures, correlating with reduced viability as determined by Sytox Red dead stain fluorescence imaging.

CP-C cells differ from CP-B and CP-A in that from the initial baseline images pre-exposure they display morphological irregularities, possibly suggestive of a predisposition to environmental stress. By the 4th exposure, there are significant membrane protrusions and irregular edges. As the exposures increase, so does the morphological degradation, characterised by cell flattening and fragmentation by the 8th and 10th exposures. The rapid changes in morphology imply a heightened sensitivity to fluctuations in pH, making CP-C single cells especially vulnerable to the ABSS, and correlating with CP-C's high mortality rates as previously analysed (Figure 4.6).

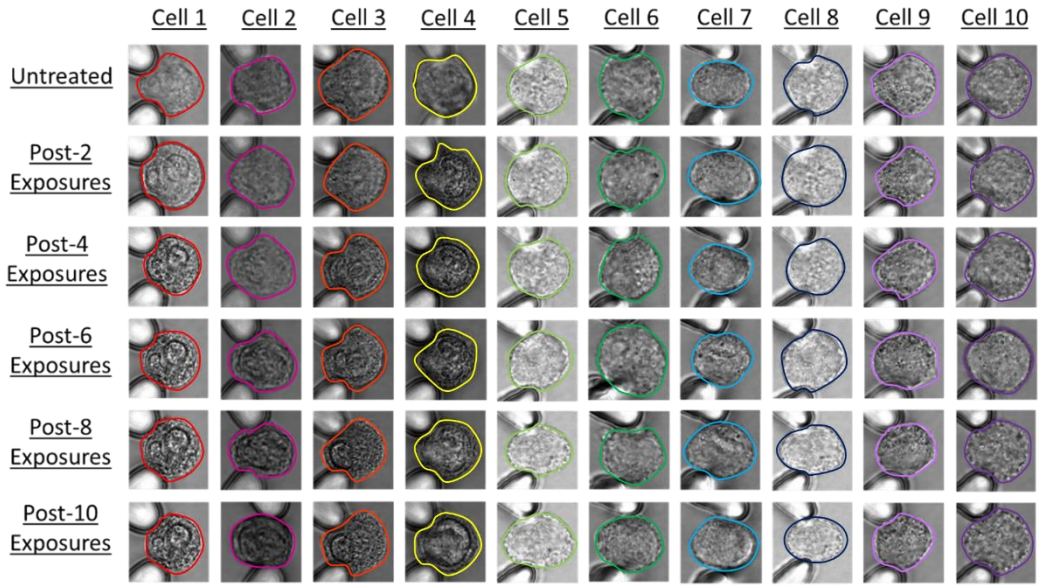

Figure 4. 13: Bright-field images illustrating the morphological changes in CP-D cells subjected to pH 4 acidic-bile salt solution. Cells were exposed for 60 seconds to low pH conditions, followed by restoration to pH 7.0 for 30-minute recovery intervals. Images were captured after 0, 2, 4, 6, 8, and 10 exposures. The outlines of cell membranes are highlighted to emphasise changes in shape, size, and integrity. Significant structural degradation is observed with increasing exposures, correlating with reduced viability as determined by Sytox Red dead stain fluorescence imaging.

CP-D appears to display the greatest resistance amongst all the cell lines. Earlier exposures (2^{nd} and 4^{th}) show minimal morphological deviations, with the single cells retaining their rounded structure. By the 6^{th} exposure, there are slight membrane irregularities, however these are less pronounced than CP-A and CP-B at equivalent stages. Even after 8 and 10 exposures, CP-D cells retain intact membranes despite some evidence of blebbing and fragmentation.

When comparing CP-B, CP-C, and CP-D, it is clear that CP-C experiences the most severe and rapid morphological changes in response to ABSS exposure. This aligns with its high death rates and suggests significant susceptibility to acidic environments. CP-B cells exhibit an intermediate response, maintaining structural integrity for longer durations but ultimately succumbing to stress at higher exposures. CP-D cells show the least morphological degradation, maintaining cellular structure even after repeated exposures, which supports the hypothesis that their transformed phenotype provides them with a degree of resilience.

4.5 Final Conclusions

Previously shown, disease development of oesophageal adenocarcinoma from healthy and Barrett's oesophagus can be determined via spectro-chemical changes within single cells, with an accuracy of 97 %.

Whilst there is alternate literature looking at the effects of ABSS on patient tissues and 2D cell monolayers (362, 378-383), little is known about the effect of a low pH microenvironment on individual Barrett's oesophagus cells in real time, including how the absence of the extracellular matrix and cell-cell signalling may influence their intracellular defence and repair mechanisms. Huo *et al.* (2010) demonstrated that acid and bile salts induce CDX2 expression via NF-κB activation specifically in esophageal squamous cells from BE patients, suggesting a predisposition at the epithelial level. Li *et al.* (2017) found that acid exposure in esophageal adenocarcinoma and dysplastic cells upregulated Silencer-of-Death Domain (SODD) protein and reduced apoptosis via NOX5-S and NF-κB pathways, indicating that acid may promote survival of genetically unstable cells in later disease stages. Zaika *et al.* (2011, 2013) focused on the DNA damage response, showing that p73 is crucial for repairing bile-induced DNA damage, while its dominant-negative isoform ΔNp73, induced by bile and cytokines, may inhibit repair and promote carcinogenesis. Das *et al.* (2011) developed a long-term BE monolayer model chronically exposed to acid and bile over 65 weeks, which showed progressive transformation to a tumorigenic phenotype, mimicking in vivo disease evolution.

To address the lack of 3D architecture in monolayer models and diversity in cellular tissues, Zhang *et al.* (2022) established human BE organoids from patient biopsies, which recapitulate key histological and molecular features of BE. Upon exposure to acidic bile salts, early-stage organoids exhibited epithelial-mesenchymal plasticity (EMP), characterized by loss of membranous E-cadherin and increased cell migration, especially when embedded in a collagen I matrix to simulate a wound-healing environment.

Data collection occurred immediately after pH neutralisation post-ABSS exposure, with the intention of determining the impact of a changing pH microenvironment on known stages of Barrett's oesophagus single cells. This may reduce the comparability against previous studies using 2D monolayers or 3D organoids, formalin-fixed tissues or patient samples due to the influence of the duration between *in vivo/vitro* exposure, sample retrieval and

preparation, prior to laboratory analysis. Rapid acquisition of this data was aided by the use of a microfluidic platform.

4.5.1 Cellular Response to ABSS

The data presented detailing exposure of Barrett's oesophageal cell lines to acidic bile salt solution revealed distinct differences in cellular responses towards extracellular pH fluctuations. The progression of ABSS treatments caused an increase in the spectral intensity for biological peaks associated with nucleic acids, proteins and lipids, indicating intrinsic biochemical changes in response to the low pH environment. As an example, the CH_2 symmetric stretch at 2850 cm^{-1} and the amide I band at 1656 cm^{-1} showed prominent increases, suggestive of increased lipid membrane reorganisation and protein structural modifications. These changes may point to cellular adaptation mechanisms, or possibly stressed-induced alterations within the cell.

CP-C demonstrated the most significant spectral changes, indicating that they have a higher degree of biochemical sensitivity to acidic conditions compared to the other cell lines. Increases were seen in peaks associated with lipid deformation and DNA-stretching, highlighting the role of oxidative stress or structural damage. On the other hand, CP-B cells showed the most minimal changes in their spectral profile, correlating with their higher viability under ABSS exposure. CP-A and CP-D both displayed intermediate responses, with increases in aromatic acids and protein-associated peaks, suggestive of protein oxidation.

The heatmap provided a visual representation which further emphasised the intensity trends over recurrent exposures. Whilst individual spectra highlight specific biochemical changes, the heatmaps allowed for identification of broader patterns such as cumulative increases in lipid-associated peaks over time, which in turn could be indicative of membrane restructuring or damage.

Exposure to a low pH solution (ABSS) would cause an influx of H^+ ions into the cells. This could potentially be redistributed within the cell through mechanisms aimed at maintaining intracellular homeostasis. Previous research suggests that organelles like the Golgi apparatus (384) or lysosomes (385) may play a role in managing intracellular pH changes. Lysosomes, in particular, are known to regulate pH and protein degradation, and while no specific

staining was performed to confirm their involvement in this study due to time restrictions, their role in acidic stress responses has been well-documented. However, lysosomal capacity to buffer excess H^+ ions is limited by their finite volume. When this capacity is exceeded, mechanisms such as exocytosis may occur, potentially leading to morphological changes such as bleb formation. These blebs, or ruptures in the actin cortex (386-388), may be observed as large protrusions after repeated ABSS exposures. The formation of blebs involves membrane compromise, cytosolic expansion, and lipid membrane tearing, which could result in a reduction of cellular lipid content (389). The presence of bile salts may have impacted this signalling between the lipid membrane and actin cytoskeleton, leading to a loss of cellular composure. This is supported by previous research indicative of bile salts influencing cell surface signalling via partitioning directly into the membrane and altering its microenvironment (390)

Changes in intracellular ion concentrations may also play a role in this mechanism. An increase in H^+ ions could lead to a reduction in intracellular K^+ ions via H^+/K^+ pump activity, potentially triggering cascades that activate caspases and nucleases (391). These processes are linked to DNA fragmentation and apoptosis, as reported in previous studies (392). The Raman spectra highlight DNA-associated peaks, may reflect such DNA damage, potentially indicating fragmentation (393-396).

4.5.2 Morphological Responses to Extracellular Stress

Bright-field imaging revealed increasing structural alterations between the cell lines, especially between CP-A and CP-C cell lines, following repeated ABSS treatments. These changes were characterised by increased irregularities in cell boundaries and decreased cellular integrity, especially after exposures 8 and 10. CP-A single cells displayed significant membrane protrusions and cytoplasmic shrinkage, indicative of cellular stress or the early stages of apoptosis. CP-C exhibited similar trends, with more severe blebbing and fragmentation, which correlates with their lower viability.

Contrasting this, CP-B single cells maintained stable morphology over the experiment, with intact membranes with minimal signs of blebbing or cytoplasmic loss, consistent with their higher viability post-exposure. CP-D cells present some morphological changes, potentially reflecting their transformed nature and grater resilience as a severely dysplastic BO cell line.

These observations suggest that CP-B and CP-D cells have robust mechanisms for maintaining cellular morphology under acidic stress.

Morphological alterations observed in the bright-field images in Figures 4.10 - 4.14 may be related to osmotic responses of the cells to their extracellular environment. Cellular swelling or shrinking is a known response to hyper- or hypotonic conditions, driven by osmotic pressure and concentration gradients causing water flux across the membrane (397). This could explain the observed changes in cell size and shape. The plasma membrane has a limited ability to expand, with volumes exceeding a 10% increase leading to membrane rupture (398, 399). Lysosomes may again play a role in regulating these changes, helping cells manage osmotic imbalances and maintain volume homeostasis to prevent cell death (400).

Building on the work conducted in the previous chapter, this ABSS-exposure study expanded the application of Raman spectroscopy by exploring how BO cells respond to a low pH bile salt solution in real time. This introduced a crucial factor – extracellular pH fluctuations – that mimic the physiological conditions of a GORD / BO patient, and play a significant role in disease development. The study's observations, including membrane blebbing and cell shrinkage, tie closely to the spectral data, providing a clearer picture of how biochemical and physical stress responses are connected.

5 Mechano-Phenotyping for Disease Progression

The previous results chapters detail the use of biochemical characterisation technique Raman spectroscopy for determining the disease stages towards the development of oesophageal adenocarcinoma. Whilst this approach provides valuable insights, it does not infer the mechanical properties of cells, which are markers for determining cellular health and disease development. Mechanical properties within cancer cells, such as stiffness and deformability, vary significantly within different disease stages, offering an additional diagnostic potential. Cell deformability, driven by interactions among subcellular components such as the cytoskeleton, nucleus, and cytoplasm, serves as a valuable biophysical marker for detecting disease (401, 402). This parameter has been previously validated and applied to diseased cells (12, 295); however, it has not yet been explored in the context of Barrett's oesophagus or oesophageal adenocarcinoma.

Through employing a multi-modal approach using both deformation data and biochemical insights through Raman spectroscopy, this method could enhance our understanding of cellular alterations associated with disease. This in turn may offer a potential diagnostic solution where untreated, unlabelled diseased cells are able to be staged on-chip, allowing for rapid disease stage determination. This chapter examines the mechanical profiling of cells to identify diseased populations across multiple datasets representing various stages of disease. Through utilising cell deformability as a key parameter, both manual and machine learning approaches are employed to characterise changes across mechanical properties associated with disease progression.

5.1 Deformation Optimisation

Previous literature details microfluidic devices which have been successfully used for determining the disease stage of single cells through their deformability, as illustrated in Figure 5.1a. The deformability of a cell can be defined through a deformation index (D.I.), which is a quantitative measure of how much a cell deviates from its natural spherical shape in solution when subjected to mechanical forces. The deformation index formula is detailed in Figure 5.1b, where cellular parameters relating to the height / width of the cell at its

maximum deformation is correlated with its stiffness and elasticity. A D.I. value closer to 1 is indicative of a more spherical, less deformed cell, with values above 1 suggestive of highly deformed and stretched cells.

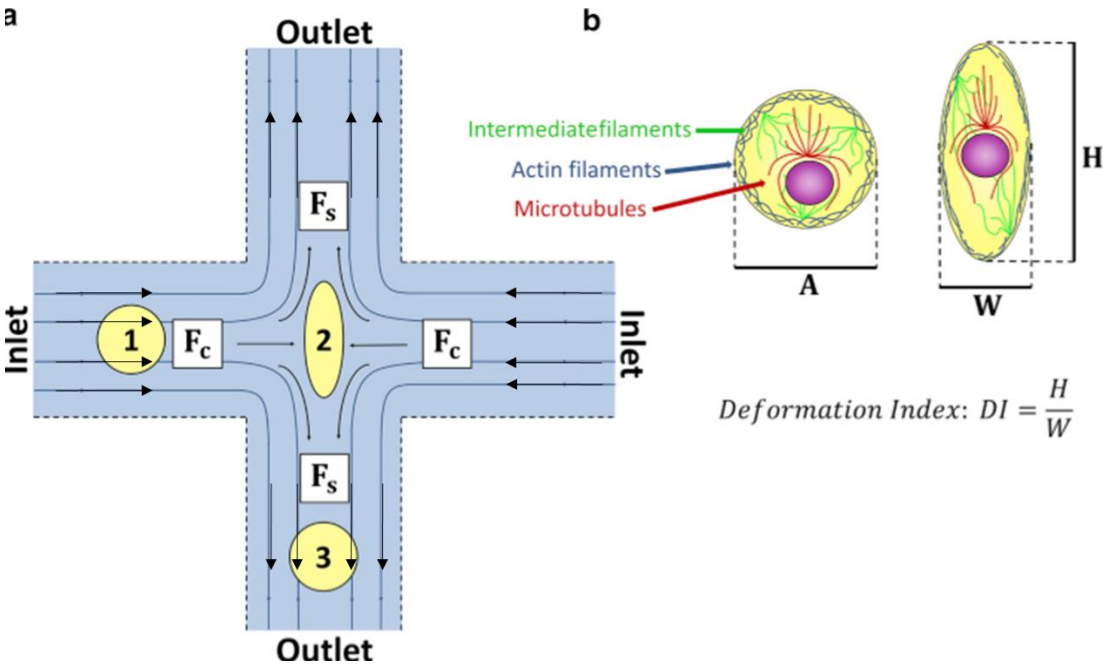

Figure 5.1: (a) Diagram illustrating the cross-flow region of the microfluidic device with shear (Fs) and inertial (Fc) forces. Changes in cell shape are labelled as 1 (undeformed), 2 (stagnation point max deformation) and 3 (undeformed) (b) Diagram of a cell, showing the nucleus and key cytoskeletal filaments (actin, microtubules, and intermediate filaments) that primarily contribute to cell stiffness. Additionally, parameters such as cell area (A), height (H), and width (W) are extracted from high-speed videos capturing cell deformation. Reprinted and amended with permission from (12).

The cross-slot microfluidic devices and experimental flow rates employed in these experiments were previously optimised by Armistead *et al.,* (12), as outlined in the Methods section (2.6 Deformation Cytometry Procedure) and depicted in Figure 5.1a. To validate cell tracking accuracy for deformation calculations, 7 µm polystyrene microspheres (Thermo Fisher) were flowed through the devices and monitored to ensure they were not flagged as deformable, consistent with their non-deformable plastic nature. They were suspended in a solution of PBS containing 0.5% methyl cellulose, creating a shear-dominant flow regime with

a viscosity of approximately $\mu \cong 33$ *cP*, consistent with the conditions determined for single cell experiments.

This optimisation is depicted in Figure 5.2a, where a superimposed bright-field image is shown of a single bead passing through the stagnation point of the cross-slot device at a flow rate of 20 µl/ min, captured using a frame rate of 20,000 fps. Further details regarding experimental setup are detailed in 2.7 Single Cell Deformation. No deformation of the beads was observed (D.I. = 1) as expected, confirming their suitability as a control for comparison against deformable objects such as single cells, with an example shown in Figure 5.2b. A superimposed image of a BO CP-A single cell is displayed, with blue arrows illustrating fluid flow, and a box outlined in red dashes highlighting the stagnation point within the device, where the cells undergo maximum deformation at the point of extensional flow.

The hydrodynamic stretching of cells undergoing deformation cytometry was first determined by Gosset *et al.,* (8). As illustrated in Figure 5.1a, the single cells experience deformation from a compressive force (F_c) generated by fluid inertia and a shear force (F_s) caused by fluid viscosity.

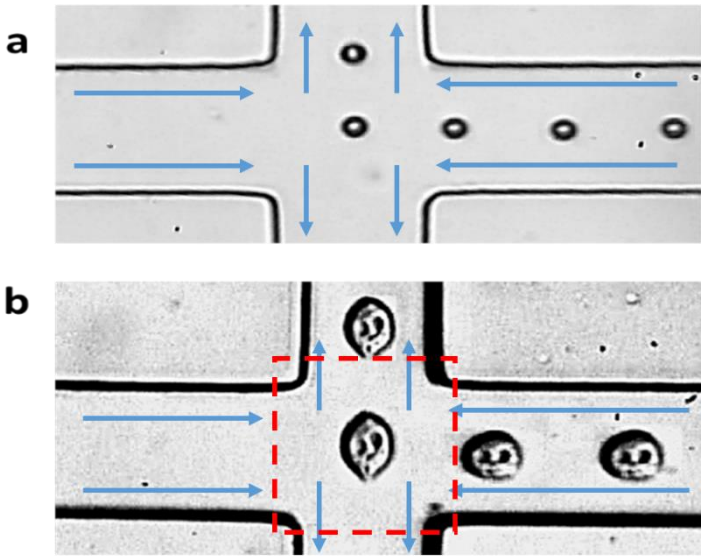

Figure 5.2: Comparison of the trajectory of a polystyrene bead in a cross-slot microfluidic device compared to a CP-A cell. Fluid flow is marked with blue arrows. (a) Bright field superimposed image of a 7 µm diameter bead passing through the extensional flow junction. The flow rate was 20 µl/min and the beads were suspended in PBS with 0.5% (w/v) methyl cellulose (µ≈33 cP). (b) A bright-field image of a CP-A cell passing through a device with identical geometry and flow conditions.

5.2 Deformation as a Marker of Cancer Development

The prospect of using deformability as a function of flow rate was evaluated across single cells from healthy, Barrett's and oesophageal adenocarcinoma cell lines (HET-1A, CP-A, CP-B, CP-C, CP-D, OE19). A detailed overview of experimental conditions is provided in the Methodology chapter, Section 2.6 Deformation Cytometry Procedure. For these experiments, all cells were suspended in a solution of PBS with 0.5 % methyl cellulose, filtered prior to experimentation (20 µm filters), and subjected to flow rates ranging from 5 – 60 µL/min.

Similar to the analysis presented in Chapter 3, this study begins by examining the potential differences between healthy and cancer cell lines, as these are expected to exhibit the largest pronounced variations. One thousand cells per cell line were tested at each of the following flow rates: 5, 10, 20, 40 and 60 µL / min. Their maximum deformation index (D.I.) was plotted against their original spherical cell size in the scatter plots presented in Figure 5.3a and Figure 5.3b. The mean deformation index (D.I.) and corresponding standard deviation for each dataset are depicted in the Y-error plot presented in Figure 5.3c. Each flow rate is distinguished through colour for clarity: yellow (5 µL / min), orange (10 µL / min), pink (20 µL / min), purple (40 µL / min) and navy (60 µL / min).

When comparing between datasets, it is evident that the mechanical properties of healthy HET-1A and cancerous OE19 single cells differ significantly. Density scatter plots for healthy (Figure 5.3a) and cancerous (Figure 5.3b) single cells highlight the distinct variation between this data. HET-1A cells present a modest increase in D.I., rising from 1.41 to 1.73 as the flow rate increases from 5 to 60 µL/ min, with an average cell size of 15.2 ± 0.046 µm. The low D.I. and lack of change across flow rates suggest that these cells are comparatively stiff and non-diseased (6). In contrast, OE19 single cells have a larger average size of 16.94 ± 0.363 µm and they display a large increase in D.I. over increasing flow rates – starting at 1.61 and rising to 2.85, a 77 % increase. The large range of D.I. values across flow rates, combined with the significant variation at intermediate flow rates (20 – 60 µL/ min) indicate that the OE19 cells are highly deformable, and therefore at an advanced stage of disease.

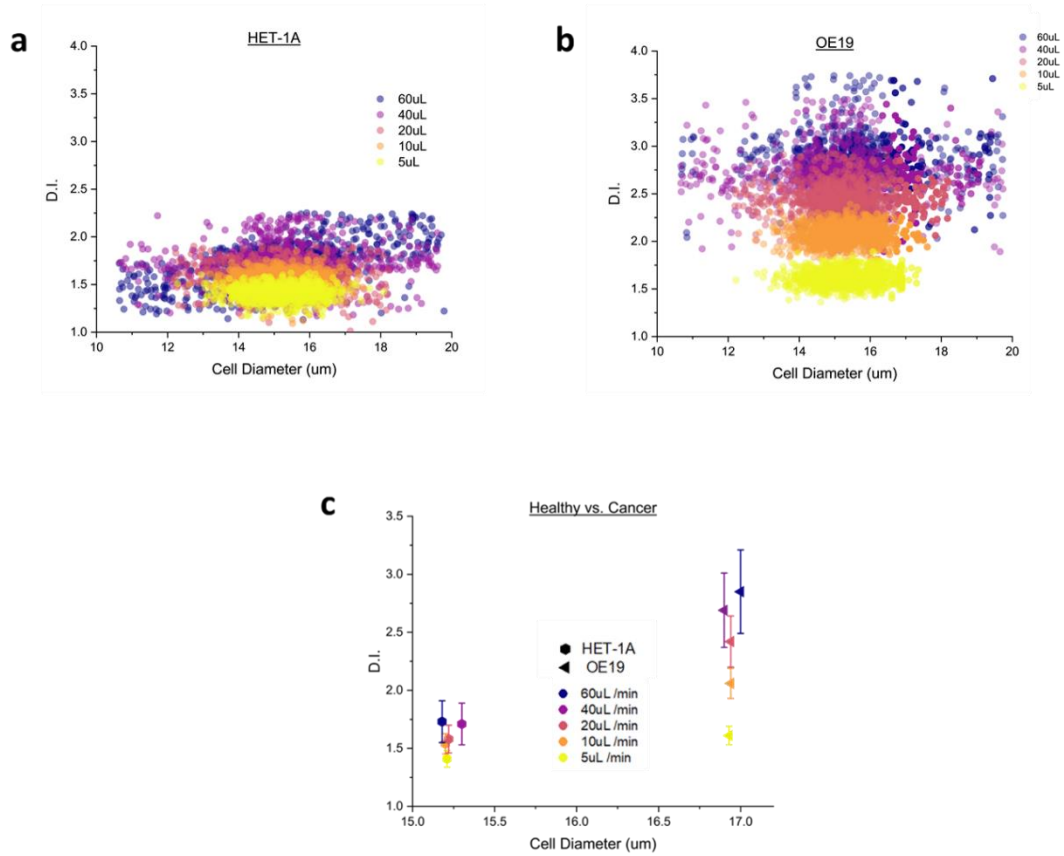

Figure 5.3: Scatter plots and corresponding mean deformation indices (D.I.) of HET-1A and OE19 cell lines across varying flow rates. The scatter plots (top panel) show the relationship between cell size (µm) and DI for each flow rate: yellow (5 µL/min), orange (10 µL/min), pink (20 µL/min), purple (40 µL/min), and navy (60 µL/min). The lower panel depicts the mean DI values and their standard deviations for HET-1A and OE19, plotted against cell size. Each point is plotted with 50 % transparency to show data concentration.

The differences in mean D.I. each flow rate for HET-1A and OE19 is displayed in Figure 5.3c, where minimal overlap is observed between the datasets within the error margins. HET-1A cells display a tightly clustered distribution, which is characterised by smaller cell sizes and lower D.I. values, which remain consistently low across all flow rates. The substantial overlap in the mean D.I. error at different flow rates indicates no significant variation in cell stiffness measurements within the HET-1A dataset, regardless of chosen flow rate. This reinforces the conclusion drawn from Figure 5.3a, that HET-1A cells represent a stiffer and minimally-deformable healthy cell population.

To contrast this, the OE19 dataset illustrates significantly larger error margins for each flow rate, which progressively increase as the flow rate rises. Interestingly, there is no overlap between the lowest (5 µL/ min) and highest (60 µL/ min) flow rates within the OE19 dataset, suggesting a large spread of data and high cell deformability / low cell stiffness, characteristic of cancer (403, 404). For experimental comparison against both single-cell datasets, there is no overlap between the mean D.I. values or errors at flow rate above 20 µL/ min, suggesting that this would be the optimal flow rate for classification of these cells as either healthy or cancer.

Changes in this cellular morphology for HET-1A and OE19 single cells during deformation can be determined during analysis, as displayed in Figure 5.4.

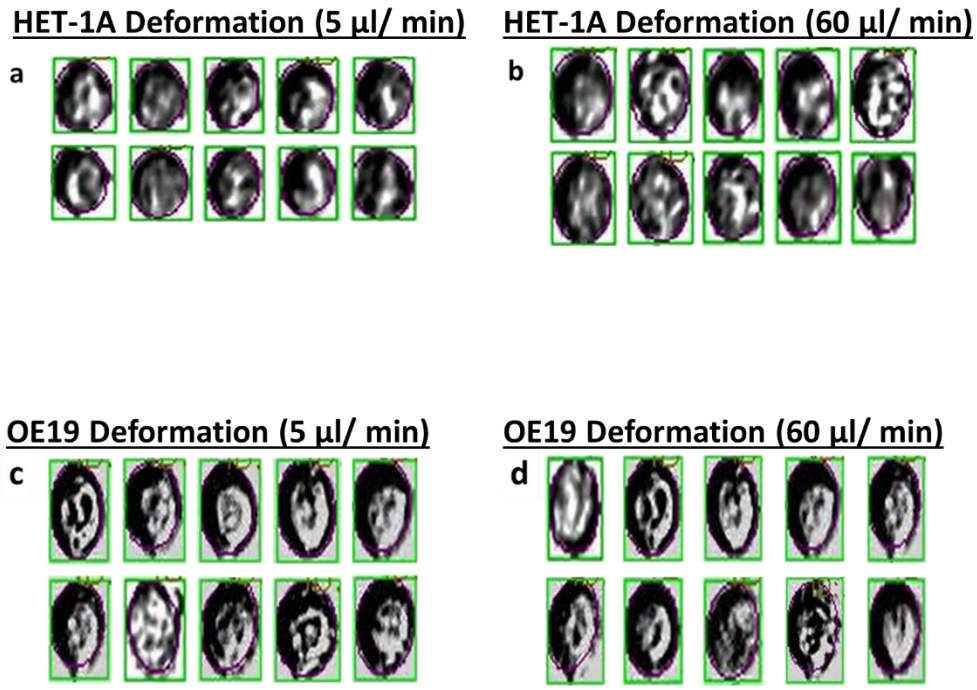

Figure 5.4: Images of single cell deformation at 5 µl/min for (a) HET-1A and (b) OE19, compared to single cell deformation at 60 µl/min for (c) HET-1A and (d) OE19. Images were taken during the analysis process. A purple outline defines the parameters of the cell, whilst the green box outlines the overall size of the cell during deformation at the stagnation point.

Figure 5.4 illustrates the changes in HET-1A and OE19 single cells as they undergo maximum deformation. Greyscale images are displayed, with purple outlines highlighting the membrane of each cell, and green boxes encompassing the entire cell area, acting as

segmentation boundaries for analysis. At a flow rate of 5 µL/ min, HET-1A (Figure 5.4a) exhibit a well-rounded shape, with well-defined circular outlines, suggestive of a structurally consistent and less deformable population. The rounded shape of HET-1A single cells become slightly elongated during maximum deformation when the flow rate is increased to 60 µL / min (Figure 5.4b). Consistent with the data provided in Figure 5.3, the OE19 cells shown signs of deformability even at lower flow rates (5 µL / min), illustrated in Figure 5.4c. An increase in flow rate (60 µL / min) led to an increase in deformability where the cell shapes become more elastic and compressed when exposed to larger stress.

5.3 Biomechanics of Barrett's oesophagus

Whilst significant differences between healthy and cancerous cell lines are expected due to their distinct nature, distinguishing subtle variations between closely related cell lines, such as Barrett's CP- cell lines, proves to be more challenging. Having already detailed their biochemical similarities in Chapter 3 and Chapter 4, deformation cytometry experiments were carried out with the aim of determining whether the mechanical properties of Barrett's single cells would be enough to determine disease progression.

In a similar trend to the OE19 dataset (Figure 5.3c), CP-A cells exhibit an increase in D.I. mean and error as the flow rate also rises. This is also depicted across all four CP- cell lines. This can be determined from Figure 5.5a, where the error rises from 0.07, to 0.33 between 5 and 60 µL/ min. The mean D.I. steadily rises with increasing flow rate, until 60 µL/ min, where there appears to be a plateau. This indicates that flow rates above 40 µL/ min are non-desirable for distinguishing CP-A single cells. Interestingly, the smaller-sized cells from this population (14.82 ± 0.647 µm) were calculated to have the largest variance of data at 60 µL/ min. This is illustrated in their respective density scatter plot (Figure 5.5b), where a larger proportion of smaller-sized cells are plotted across a D.I. variance of 2.29 ± 0.033. The cells exhibit increased malleability at higher flow rates (40 - 60 µL/ min); however it is also possible that the intensity of the flow is forcibly compressing the cells, suggesting that they may be increasingly sensitive to external mechanical stressors.

The presented data for the CP-B cell line (Figure 5.5c) highlights the small variation in cell diameter parameters (± 0.4 µm) across all flow rates, indicative of cellular rigidity and stable structural composition under pressure. There is relatively small D.I. error at the lowest flow

rate of 5 µL/ min (2.04 ± 0.07), also presented in Figure 5.5d, where the data values are tightly grouped together and clustered with few outliers. CP-B displays a similar trend of mean D.I. error overlap, with flow rates 40 and 60 µL / min largely overlapping, suggesting that a maximum deformation rate of 40 µL / min would be enough for maximum deformation.

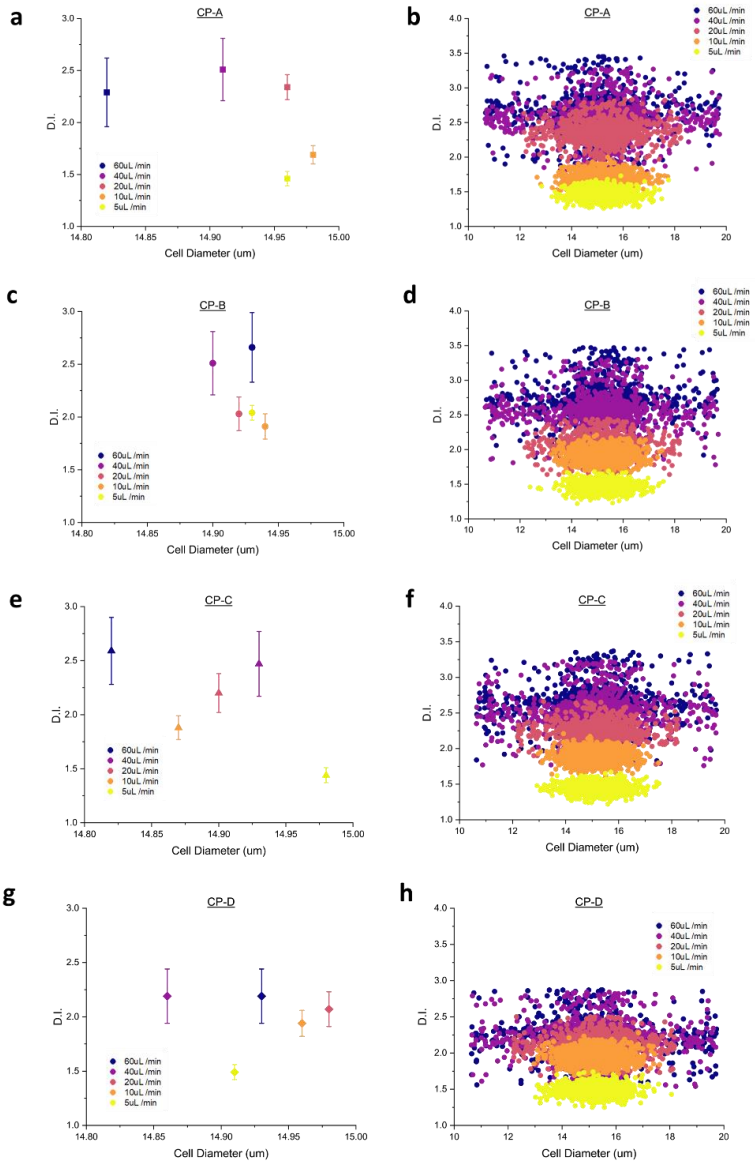

Figure 5.5: Mean deformation index (D.I.) and density scatter plots for Barrett's oesophagus single cells across flow rates (5 – 60 µL / min). : Scatter plots and corresponding mean deformation indices (D.I.) of CP- cell lines across varying flow rates. The left-hand panel depicts the mean D.I. values and their standard deviations for CP-A (a), CP-B (c), CP-C (e) and CP-D (g), plotted against cell size. The scatter plots (right-hand panel) show the relationship between cell size (µm) and DI for each flow rate. Each chart is colour coded to distinguish between flow rates: yellow (5 µL/min), orange (10 µL/min), pink (20 µL/min), purple (40 µL/min), and navy (60 µL/min).

At a lower flow rate of 5 µL/ min, the mean D.I. for CP-C single cells is calculated as 1.44 (± 0.07), with the largest average cell diameter (14.98 ± 0.06 µm). An increase in the flow rate impacts cellular diameter due to due to greater compression of the cell, with diameter at the highest flow rate of 60 µL/ min averaged as 14.82 µm. When relating this analysis to the corresponding scatter plot (Figure 5.5f), each population of cells is grouped together, with large variation seen only once the flow rate reaches a minimum of 40 µL/ min.

When examining the data from CP-D single cells, it is evident that the flow rates 40 and 60 µL/ min yield the exact same D.I. value and distribution across the dataset (2.19 ± 0.25). the scatter of data from this dataset is also heavily overlapped, seen in Figure 5.5h, where lower flow rates (5 and 10 µL/ min) have clustered populations with minimal D.I. error (± 0.07 – 0.12, respectively), and faster flow rates (20 - 60 µL/ min) leading to a larger spread of scattering. The results from Figure 5.5g and 5.5h further enforces that the use of a 40 µL/ min flow rate is sufficient for producing maximum deformation, as the D.I. values begin to plateau above this, as shown in Figure 5.6.

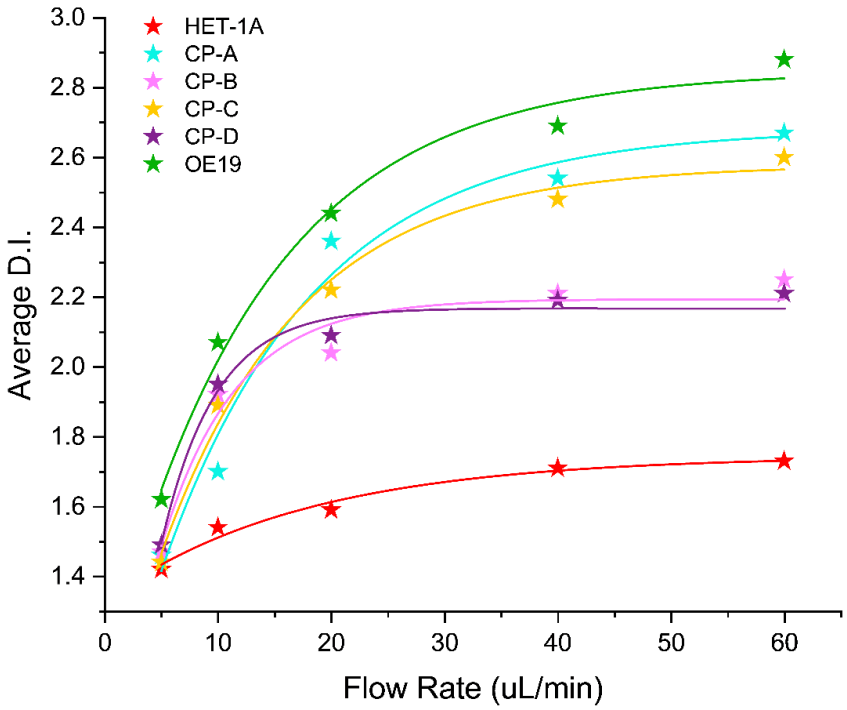

Figure 5.6: Average deformation index (D.I.) values for HET-1A (red), CP-A (cyan), CP-B (pink), CP-C (yellow), CP-D (purple), and OE19 (green) across flow rates of 5, 10, 20, 40, and 60 µL/min. Each data point represents the mean of 1000 individual D.I. measurements, with an exponential fit applied to each dataset to illustrate trends.

Figure 5.6 illustrates the relationship between flow rate and deformation index (D.I.) for cell populations involved in the development of oesophageal adenocarcinoma. The general trend observed across all populations is the linear relationship between the increase in average D.I. value and increase in flow rate. This would indicate that a higher flow rate would induce a higher level of deformation. Despite this, the rates of increase and the maximum D.I. values between each cell line differ largely, highlighting distinct mechanical properties.

The HET1A population (red) displays the slowest increase in D.I over each flow rate, supporting the theory that the dataset contains stiffer cells which resist deformation under mechanical stress. In contrast to this, as earlier elaborated, the OE19 cell line (green) shows the steepest increase and the highest overall D.I. values, reflective of a highly deformable cell population. CP-A (cyan) and CP-C (yellow) demonstrate increased deformability and similar trends when compared against the CP-B (pink) and CP-D (purple) cell populations, which show gradual increases in D.I. and lower maximum values than OE19.

Whilst mechanical differences alone were unable to distinguish between the Barrett's cell lines, clear trends emerge relating to similarities across datasets. CP-A and CP-C show similar trends regarding a steady increase in D.I. with increasing flow rate. Conversely, CP-D and CP-B show a more gradual increase and exhibit a plateau after a flow rate of 40 µL/ min. Each curve plateaus around the highest flow rate of 60 µL/ min, indicating that the cells may reach a deformation limit beyond a certain threshold. This is suggestive of a mechanical tolerance level or structural limit for these cells types under increasing flow-induced stress.

5.4 Linear Discriminant Analysis

LDA was performed with model performance assessed using metrics such as accuracy, recall, and precision, all defined in the methodology chapter (2.7.2 CNN Analysis), along with a confusion matrix. These values were subsequently calculated as 60.34 %, 55.51 % and 65.56 %, respectively. To handle any imbalances in class representation, random sampling was applied to ensure all cell lines were equally represented. The confusion matrix for this work is shown in Figure 5.7.

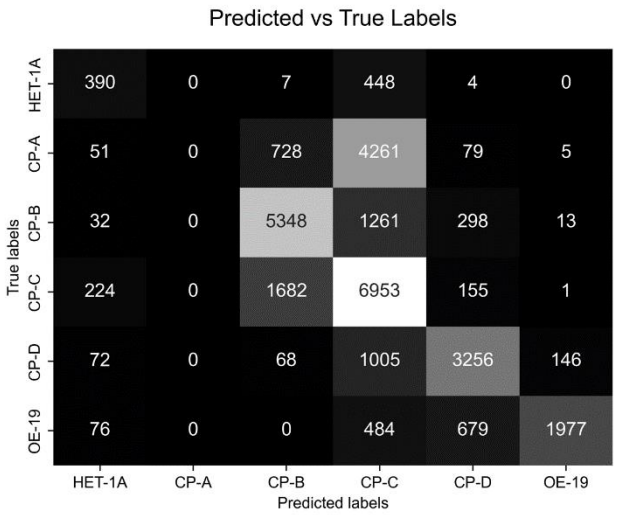

Figure 5.7: Confusion matrix illustrating the performance of the model in predicting the labels of six cell lines (HET-1A, CP-A, CP-B, CP-C, CP-D, and OE-19). Each cell represents the number of instances classified into the corresponding predicted class, with true labels on the y-axis and predicted labels on the x-axis. Misclassifications are shown as off-diagonal elements, highlighting areas of poor model performance.

The overall accuracy, recall and precision appear high even as the models perform poorly on specific classes due to the calculation methods used for these metrics. Accuracy is used to reflect the proportion of correctly classified samples (true positives + true negatives) out of all the samples, regardless of class. If the model performs well for dominant classes with larger proportions of data (such as CP-B and CP-C), this can increase the overall accuracy, even if it fails for a smaller or more challenging class (CP-A). Overall recall and overall precision were computed as macro-averages, where the recall or precision for each class was calculated separately, and those values then averaged equally across all classes, regardless of the number of the samples within that class. In this case, the classes with high true positive counts (detailed in Table) such as CP-B and CP-C, heavily influence these metrics, compensating for poor performance in smaller classes such as CP-A.

True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) are calculated from the confusion matrix to evaluate classification performance. TP represents the number of instances correctly classified as belonging to a specific class (diagonal entries of the confusion matrix). FP is the sum of all instances incorrectly classified as the target class, calculated as the column total for that class minus the TP. FN is the number of instances from the target class that were misclassified into other classes, calculated as the row total for that class minus the TP. TN includes all correctly classified instances that do not belong to the

target class, calculated as the sum of all remaining entries in the matrix excluding the target class's row and column.

Cell Line	True Positives (TP)	False Positives (FP)	False Negatives (FN)	True Negatives (TN)
HET-1A	390	455	459	28399
CP-A	0	0	5124	24579
CP-B	5348	2485	1604	20266
CP-C	6953	7459	2062	13229
CP-D	3256	1215	1291	23941
OE19	1977	165	1239	26322

Table 8: Table showing class-specific performance metrics for the model, including True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) for the HET-1A, CP-B, and CP-D cell lines.

For HET-1A single cells, 390 samples were correctly classified, however a larger number (448) were misclassified as CP-C cells. this significant miscalculation suggests that there is a overlap in defining features, such as cell circularity or D.I., between both cell lines, making it difficult for the LDA model to draw any clear conclusion. There were other minor misclassifications into other classes, however the confusion with CP-C was most prominent. In comparison, OE19 appears to perform better, with 1977 samples correctly identified. There were still a notable number of samples still misclassified, as CP-C (484) or CP-D (679), indicating that OE19 shares overlapping characteristics with both cell lines. OE19 achieved a higher classification accuracy than HET-1A, however the high rates of misclassification point towards significant overlap in feature distributions, highlighting the challenges of separating similar cell lines using a linear method such as LDA. Both cell lines has considerable misclassifications, with the most confusing occurring across the CP- cell lines.

CP-A was challenging to classify, with no samples correctly predicted, and the majority (4261) misclassified as CP-C. It can be proposed that there are significantly similar features across both cell lines, such as D.I. or cell perimeter, and that the lack of CP-A data within the training dataset directly affected classification accuracy. Additionally, smaller numbers of CP-A were also misclassified across CP-B and CP-D, illustrating the difficulty of isolating CP-A. CP-B had a better classification rate, with 5348 samples correctly identified. However, there were still 1261 misclassifications as CP-C and 298 as CP-D. Whilst CP-B showed less classification overlap with CP-A, the overlap with CP-C still significantly impacted its performance. This suggests that CP-B single cells had more distinct features, however the overlap with CP-C is too much of a challenge for this model. CP-C had relatively strong classification performance, with 6953 samples correctly classified. There were lower misclassification results when compared against CP-A and CP-B, however 1682 samples were still incorrectly classified as CP-B. The high number of classifications indicates that CP-C is more separable from the other cell lines overall. CP-D performed the worst out of all the CP- cell lines , with the exception of CP-A, with 3256 samples correctly classified. There were however, slight misclassifications into CP-C (1005) and OE19 (146). This suggests that CP-D shares morphological and mechanical characteristics with the other CP- cell lines, whilst still remaining distinct.

The model achieved an overall accuracy of 60.34 % ± 0.2, which indicates that whilst LDA could classify some of the cell lines reasonably well, there were still significant miscalculations. The recall, at 55.51 % ± 0.8, is reflective of the model's struggle to capture all true samples for certain classes, in particular HET-1A and CP-A. When looking at the precision, at 65.56 % ± 0.3, the presence of many false positives are highlighted, particular in CP-C and CP-D. These combined metrics show that LDA is capable of delivering a semi-accurate performance, however, the overlapping feature distributions across similar classes limited its overall effectiveness. When analysing the confusion matrix, there is substantial misclassification between HET-1A and OE19, as well as between each CP- cell line, suggesting that the chosen features may not be enough to fully differentiate between classes. Additionally the linear assumptions of LDA are unable to capture the more complex non-linear relationships which appear to be present within the datasets. Amongst the CP- cell lines, CP-C and CP-B performed best, indicating that their features are most distinct, whilst HET-1A and CP-A had a poorer performance, likely due to the significant overlap found with CP-C cells.

5.5 Convolutional Neural Network Analysis

Convolutional Neural Networks (CNNs) are a specialised type of deep learning model tailored for analysing image data, making them particularly powerful for tasks that involve detailed pattern recognition. Images contain thousands of data points (pixels), with each pixel representing a small unit of information, especially when considering their dimensions (size) and colour channels (e.g., RGB for red, green, and blue). Traditional deep neural networks are computationally inefficient for handling such large datasets due to the large number of connections required between neurons.

CNNs address this challenge through convolutional layers, which connect each neuron to a small, specific region of the image known as a receptive field. This structure reduces the number of computations while preserving spatial relationships within the image. Filters, also called kernels, are applied to these regions to extract specific features and generate feature maps, which are representations of detected patterns such as edges, textures, or shapes. By stacking multiple convolutional layers, each with different filters, CNNs progressively learn complex patterns, making them efficient and effective for image-based analysis, while minimising computational demands.

The CNN architecture tested for this work was Xception (331), with a focus on understanding how data augmentation and balancing influenced the models' performance, which was evaluated using accuracy, recall and precision (2.7.2 CNN Analysis). The network were trained over 10 epochs with a learning rate of 0.1, using Stochastic Gradient Descent (SGD) as the optimiser, without any pre-initialised weights (Table 6). Epochs refer to the number of complete passes the training data makes through the model, so for this analysis the data was passed through the network a total of 10 times, allowing it to learn and adjust its parameters (weights and biases) based on previous errors. The learning rate is a hyperparameter, determining the size of the adjustments made to the weights in each iteration. A learning rate which is too high may overshoot the optimal weights, whilst a low rate may cause a longer computational time. SGD was used to optimise the weights iteratively based on the loss. This training configuration suggests a standard approach to neural network training, with the intent of learning patterns in the data directly from scratch, rather than relying on pre-existing knowledge.

For Xception, without any data augmentation or balancing, the model achieved a high accuracy of 99.74 %, with recall at 98.3 % and precision at 99.8 %. This suggest that the model was already able to detect spatial features within the datasets. After applying data augmentation, there was a slight improvement in accuracy to 99.77 %, with recall peaking at 99.87 %, which shows that the model was better at recognising variations within the data. Precision remained at 99.83 %, meaning that the augmentation didn't appear to add any false positives. Interestingly, the addition of both balancing and augmentation caused a slight dip in accuracy (99.7 %) and recall (97.75 %)., possibly due to the introduction of noise or synthetic data during balancing. Precision remained high at 99.63 %, suggesting that the model was able to reliably classify samples, even after loss of sensitivity.

Architecture	Data Augmentation	Data Balancing	Accuracy	Recall	Precision
Xception	No	No	99.74	98.3	99.8
Xception	Yes	No	99.77	99.87	99.83
Xception	Yes	Yes	99.7	97.75	99.63

Table 9: Table illustrating the classification performance metrics for different CNN model configurations under varying conditions, with and without pre-initialised weights and specific modifications. The values represent performance accuracy (%) for distinct experimental setups, highlighting areas of improvement (green) and underperformance (red) across configurations.

Data augmentation helped improve the recall by exposing the model to a larger variety of inputs, whilst balancing was used to reduce the effects of class imbalance, especially for smaller classes.

Recall improved significantly with data balancing, suggesting that the model was able to better identify HET-1A samples that were previously overshadowed by larger classes such as CP-C or CP-D in the unbalanced dataset. Precision remained high across all configurations, indicating that there were very few false positives. In comparison, the OE19 dataset performed with a lower baseline than HET-1A, possibly due to overlapping features with the CP- cell lines, especially CP-C and CP-D. Data augmentation significantly impacted recall, improving the models ability to generalise and detect OE19 samples through exposure to different patterns during training. Balancing also played a critical role by reducing false

positives through addressing the class imbalance. Overall, OE19 benefitted more from balancing than HET-1A due to underrepresentation in the dataset.

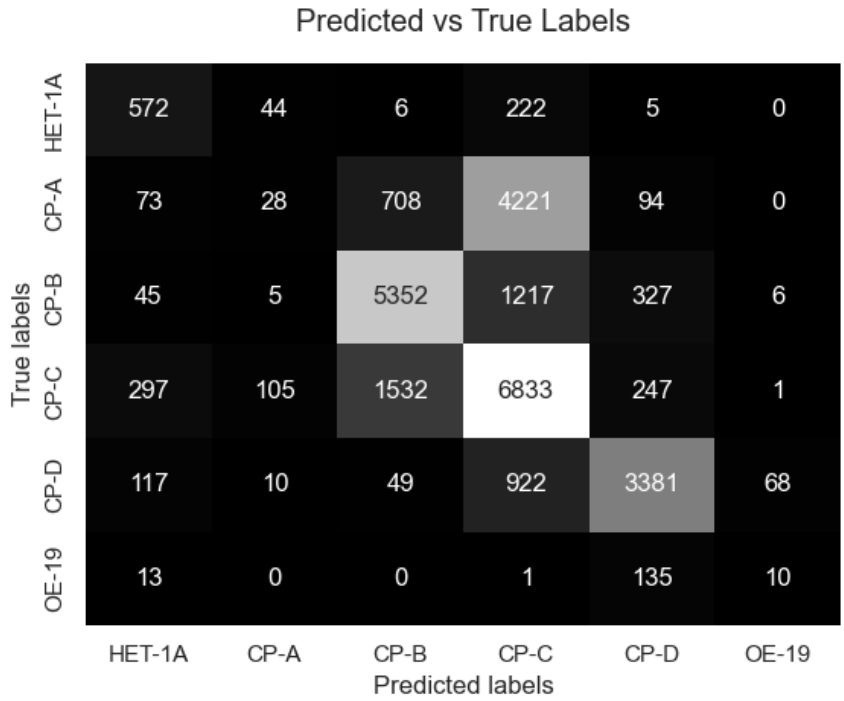

Figure 5.8: Xception Confusion matrix showing the final performance of the classification model across six cell lines (HET-1A, CP-A, CP-B, CP-C, CP-D, and OE-19). True labels are represented on the y-axis, and predicted labels are shown on the x-axis. Diagonal values indicate correct classifications, while off-diagonal values represent misclassifications, highlighting areas of model strength and areas requiring improvement.

CP-A proved to be the most difficult dataset to classify, with many samples misclassified as CP-B or CP-C. This is likely due to overlapping features such as strain and perimeter, with may be less distinct for CP-A. Data augmentation helped improve recall through diversifying the training data, which enabled the models to generalise better. This however, came at a cost to precision, as augmentation led to the introduction of noise, which then led to increased false positives. CP-B performed better, with consistently high recall and precision across all configurations. The features from this dataset appeared more distinct, which allowed the models to effectively separate it from other classes. Even without augmentation or balancing, CP-B showed strong performance, this was then enhanced when these techniques

were applied. Data augmentation didn't lead to much improvement for CP-C, however data balancing led to a noticeable difference in recall. CP-C already had strong baseline performance, with features well-represented within the dataset, but balancing helped reduced the number of false negatives for underrepresented CP-C samples. Overall CP-C was easier to distinguish than CP-A. CP-D benefitted most from the combination of augmentation and balancing, achieving the highest recall and precision across all CP- lines. This suggests that balancing the data led to a distinction in CP-D's features, and reducing confusion with CP-C and OE19, which previously cause misclassifications.

The results from this work show that the addition of data augmentation and balancing both play key roles in improving CNN performance. Amongst all the cell lines, HET-1A and CP-D were easier to classify due to their distinct features. OE19 and CP-A presented more challenges due to overlapping features with other classes, but balancing helped mitigate these issues. CP-C's features were generally well represented, with CP-B's distinctiveness making it easier to classify.

5.6 Chapter Summary

While previous studies have explored the cellular biochemistry and membrane composition of these cell lines (91, 178, 316, 405), research specifically addressing changes in the visual biomechanics throughout the progression of oesophageal adenocarcinoma remains absent. This study applies well-established techniques for assessing disease progression to investigate these changes, offering new insights into the cellular alterations associated with oesophageal adenocarcinoma.

5.6.1 Biomechanics in Barrett's Oesophagus

The presented data in this chapter provides valuable insights into how biomechanical properties evolve in different cell states. Notably, there is an observed steady increase in deformation with rising flow rates for each cell line, as shown in Figure 5.6, each of which eventually reach a plateau, highlighting how deformation behaviour responds to flow conditions through each stage of disease. This plateau effect may indicate a threshold in cell membrane elasticity, beyond which further changes in flow rate yield minimal deformation, potentially reflecting inherent structural limits within the cellular membrane composition of each cell line. This can be further supported by the comparison of increasing flow rates shown in Figure 5.5, where it is shown that a flow rate of 40 µl/min was able to induce maximum deformation for the CP- cell lines, and separate the healthy vs cancerous cell datasets into distinct clusters with minimal overlap (Figure 5.3).

These results identify a discernible increase in deformability when comparing healthy, diseased, and cancerous cell states (Figure 4.6). Healthy HET-1A cells display a stronger ability to retain their shape under compressive force, visually contrasting with the notable deformation seen in cancerous OE19 cells. This visual distinction supports the theory that as cells transition from healthy to malignant states, there is a reduction in their structural rigidity and resistance to deformation, likely reflecting progressive changes in membrane composition and cytoskeletal integrity (178, 406).

Further analysis of individual disease stages revealed that, while early and late disease stages exhibit lower membrane elasticity (Figure 4.2a, 4.2e), intermediate dysplastic stages present less rigidity (Figure 4.2b, 4.2d), suggesting a possible remodelling phase during disease progression (46, 183). This variation within stages implies that different disease stages may

involve specific cellular adaptations that affect elasticity and deformability, perhaps as a response to changes in the cellular microenvironment or mechanical forces within the tissue.

5.6.2 A.I. for Barrett's Progression

The first set of presented analysis used an LDA model on a preliminary, imbalanced dataset and found that deformation properties alone were insufficient for accurate cell classification. The LDA model struggled with data imbalance, leading to moderate accuracy and low precision for certain cell lines. This highlighted the need for additional data and balancing to improve model performance, as the LDA results were limited by the dataset's small size and uneven distribution. In the second set of analysis, the Xception model was trained on an augmented version of the preliminary dataset, leading to improved accuracy and recall compared to the LDA model. While data augmentation enhanced the model's overall precision, classification for some cell lines (like CP-A) remained challenging due to remaining imbalances. This analysis demonstrated that while augmentation improved model performance, balancing the data further would be essential for optimal classification accuracy.

CNNs excelled in analysing high-dimensional image-based data, effectively capturing spatial patterns and distinguishing features such as cell area and D.I. across different cell types. Their ability to generalise was enhanced by data augmentation and balancing, particularly for underrepresented classes (CP-A). Class imbalance and overlapping feature spaces were addressed, leading to improvements in recall and precision for smaller or less distinct cell lines, such as OE19. While CNNs demonstrated high accuracy and precision for dominant classes like CP-C and CP-B, they highlighted areas requiring further refinement, particularly in distinguishing classes with overlapping features.

In contrast, LDA provided a simpler and interpretable approach for classification by projecting the data into a reduced-dimensional space that maximally separated predefined classes. While LDA effectively identified key features distinguishing certain cell types, its limitations became apparent when dealing with highly non-linear and high-dimensional data. The results suggest that while LDA is a useful tool for initial data exploration and hypothesis testing, it lacks the robustness and scalability required for more complex biological data compared to CNNs.

Some of the confusion matrices in this analysis showed zero TP values for CP-A (Figure 5.7) and OE19 (Figure 5.8) , indicating that the model struggled to correctly classify any samples from those groups. This can arise from several factors, including class imbalance, where underrepresented classes are overshadowed by dominant ones, leading the model to ignore smaller classes. Overlapping features between classes can further confuse the model, causing misclassifications. Additionally, insufficient or poor-quality data, such as small sample sizes or noisy inputs, may prevent the model from learning key patterns. Other contributors include poorly tuned hyper-parameters, non-appropriate model architectures, or an inadequate feature selection, which may limit the model's ability to capture distinctions between classes. Errors during data pre-processing or labelling can also mislead the model during training. The lack of data augmentation or an unsuitable loss function may reduce the model's ability to generalise or handle imbalanced data. Addressing these issues through dataset balancing, feature refinement, augmentation, and model optimisation could help resolve the observed performance gaps.

Overall, the combination of CNNs and LDA provided complementary insights. CNNs demonstrated superior performance in extracting features and classifying complex patterns, while LDA offered interpretability and a framework for understanding key features driving separations. These findings underscore the value of integrating multiple analytical approaches to address diverse challenges in biological data analysis. Future work should focus on optimising CNN architectures for enhanced feature learning and exploring hybrid approaches that combine the interpretability of LDA with the predictive power of CNNs, ensuring comprehensive and accurate classification in complex datasets.

6 Thesis Reflections

The research presented builds upon foundational work in Raman spectroscopy to advance our understanding of oesophageal adenocarcinoma and Barrett's oesophagus. Raman spectroscopy revealed biochemical changes across the Barrett's progression spectrum, including alterations in lipid and nucleic acid content, which are consistent with known molecular features of squamous–columnar transformation (407). Initial findings demonstrated the efficacy of Raman spectroscopy in identifying biochemical changes associated with disease progression, achieving a high classification accuracy of 97 %. This laid the groundwork for the acid-bile study, which introduced acidic bile salt solution exposure (ABSS) to simulate the hostile microenvironment associated with disease development. This revealed critical biochemical and morphological adaptations in Barrett's oesophagus cells under stress, transitioning Raman spectroscopy from a diagnostic tool to a potential platform for monitoring therapeutic responses and disease progression.

The acid bile study highlighted significant cellular behaviours, with CP-C cells demonstrating a heightened sensitivity to acidic stress and CP-B cells exhibiting greater resilience. These findings underscore the heterogeneity within Barrett's oesophagus cell lines, with potential implications for understanding early markers of transformation. The use of heatmaps to visualise cumulative biochemical changes provided a novel approach to tracking stress responses, adding depth to single-cell analyses. However, the study's reliance on cell lines, rather than patient-derived samples or 3D organoid models, limits its translational relevance, emphasizing the need for future validation in more physiologically representative systems.

The incorporation of deformation cytometry in this research provides valuable insights into the mechanical properties of Barrett's oesophagus cells, highlighting their relationship with disease progression and stress response. The findings suggest that cellular deformability may serve as a potential biomarker for assessing resilience to acidic stress, with decreased stiffness and increased deformability in more advanced cell types, likely reflecting cytoskeletal remodelling associated with epithelial plasticity . However, the reliance on simplified *in vitro* conditions limits the broader applicability of the results, as interactions with extracellular matrix components and neighbouring cells, which influence mechanical behaviour, are absent. Whilst deformation metrics offer promising diagnostic potential, the lack of integration with complementary biochemical or genetic data restricts the interpretability of the observed mechanical changes. Future studies could enhance the utility

of deformability analyses by linking them to cytoskeletal or membrane alterations, providing a more mechanistic understanding of cellular responses to microenvironmental stress.

A key strength of this research lies in its interdisciplinary approach, integrating high-resolution Raman spectroscopy, deformation cytometry, and AI-based classification. The incorporation of CNN models for data analysis offers a pathway to more precise diagnostics, with potential scalability into portable, point-of-care devices. The techniques used here have translational potential, such as identifying early-stage markers of dysplasia or phenotyping residual/recurrent disease after therapy. Although both datasets describe disease progression, they were collected independently and cannot be directly correlated without integrated or co-localised measurement techniques. Experiments were conducted on immortalised cell lines in suspension, which lack tissue context and may not fully reflect in vivo behaviour. Cell lines are easy to use and reproducible but genetically altered; primary cells are more physiologically accurate but harder to maintain, whilst measuring cells in isolation ignores cell–cell and matrix interactions that are critical in Barrett's microenvironment.

Proposed multi-modal approaches, combining Raman spectroscopy with techniques such as fluorescence imaging or genomic profiling, could further enhance the understanding of cellular states by correlating biochemical, morphological, and mechanical data. These techniques could be adapted for structured models like monolayers or organoids to improve biological relevance. Combined with transcriptomics or immunostaining, Raman and biomechanical tools could deepen understanding of BO heterogeneity and predict progression risk. These integrations would not only refine classification models but also provide mechanistic insights into disease progression. Future studies could use platforms like AFM-Raman or live-cell co-imaging to relate biochemical and biomechanical changes at the single-cell level. With further development, these technologies may support real-time or in vivo applications through miniaturised, label-free endoscopic tools.

Appendix 1

DatasetCreationScript_1.mat

```
%% Preparation of the datasets to load onto the other script

% Specify locations of .wdf data files, Matlab scripts, and output path.
List all of the folders that have data to be processed. Just to for this
% script; can add new ones following the same format. Where there is an
% edit required I have added an X.

% directory where the data is
dirPath={...
    'X';...
};

% Matlab scripts required for this analysis. Should be in a folder
labelled
% MATLAB additional scripts
addpath( "X" );

% The path used to save the output from this script.
savePath = 'X';



% Specify each of the data sets.

% For each of repeats, need to provide ('i' refers to the data set;
% starting from 1, and ad one for each new data set):
% - start and end of all filenames as Name{i,1} and Name{i,2}, such that
the
%   actual filename is: "<Name{i,1}><repeat number><Name{i,2}>.wdf
% - the range of indices for the repeats as Index{i}=first:last;
% - start and end of all filenames for the background data, as BName{i,1}
%   and BName{i,2}, used similar to Name{} above.
% - the range of start and end of the background data runs, sim. to above.
% - the prefix for the output filename as "NameSaving{i}".
% - a "calibration peak" as CalibrationPeak{i}".

%%save each repeat with the same beginning (e.g. cpa / cpb),
%%add an underscore _  and then each repeat/new spectrum is just given a
%%number in ascending order. All spectrum names must start the same
%%so just the number at the end changes.

%Set 1:
Name{1,1}='X';
Name{1,2}='';      % No text after the repeat index for this set of runs.
Index{1}=[X:X];       % i.e. repeat indices 0, 1, 2, 3, 4 and 5.
NameSaving{1}='test1';  % For the saved output files.
CalibrationPeak{1}=520.5; % For this data sets
selectedDataSets = [1];       % Starts from 1. Can do ranges as
'=[start,end];'
```

```
%% Load each of the selected data sets
for j=selectedDataSets

    %
    % Load all repeats for this data set.
    %
    fprintf( "Loading data from data set '%s' (index %i)\n", dirPath{j}, j
);

    % Move to the directory for this dataset, and keep a record of where
we
    % were.
    oldPath = cd( dirPath{j} );

    % At the end of the loops below, m will be the number of data sets,
    % and b will be the number of background repeats. They don't seem to
    % be used anywhere, though - maybe this was orignally meant as a
check?
    m=1;
    b=1;

    % Loop over all repeats for this data set.
    for i=Index{j}

        % Get the filename for a single repeat.
        FileName{j,m}=strcat(Name{j,1},num2str(i),Name{j,2},'.wdf');
        fprintf( "- loading repeat: %s\n", FileName{j,m} ); % Check.

        % Import the ".wdf" file for this repeat.
        wdf=WdfReader(FileName{j,m});
        RamanSpectra{m}(:,1)=wdf.GetXList.';
        RamanSpectra{m}(:,2)=wdf.GetSpectra(1,1).';
        wdf.Close()

        m=m+1;      % Count the number of data repeats.
    end

    % Get the calibration peak for this dataset.
    CalibrationPeakTemp = CalibrationPeak{j};

    % Save the data. Move back to the original path first of all, i.e.
    % prior to cd'ing to the data directory.
    cd( oldPath );
    oldPath = cd( savePath );
    save([NameSaving{j} 'datasetraw'],...
       'RamanSpectra','CalibrationPeakTemp')

    % Move back to the original directory before finishing.
    cd( oldPath );
%% X alignment of the spectra respect to the Calibration peak position in
Si

% Every day of measuring a calibrating spectra is taken and it is aligned
% as good as it can to the 520.5 cm^-1 peak of Silicon. However, sometimes
% the peak seems to be around 520.3 or 520.8. This peak will be used to
% align the shift of the spectra and reduce the variability between days.
```

```
% As the PCA only considers the Y-values for analysis,
% a spline function should be fit and the Y-values should be guaranteed to
% have exactly the same X.
fprintf('Aligning spectra and fitting to the selected X values using a
spline \n')

tic   % Start timing from here.

Xint_st = round(RamanSpectra{1}(1,1))
Xintvalues = Xint_st:-0.5:round(RamanSpectra{1}(end,1))-4;
% Xintvalues = Xint_st + Xint_en;
% Xintvalues = round(RamanSpectra{1}(1,1)) +
4:0.5:round(RamanSpectra{1}(end,1))-4;
    % Array of X values for the interpolation. RamanSpectra{1}(1,1) is the
    % start point, and 5:0.5:round(...)-4 gives values separated by 0.5
    % (the rounding ensures all points are an integer or half-integer,
    % and evenly spaced),

% Spline interpolation of the actual data.
for k=1:length(RamanSpectra)
   RamanSpectraInt{k}=spline(RamanSpectra{k}(:,1)+(520.5-
CalibrationPeak{j}),RamanSpectra{k}(:,2),Xintvalues);
end

toc   % 'tic', 'toc' used for timing how long operations take; toc
displays the result.

%% Subtract the baseline
fprintf('Subtracting the baseline \n')

tic   % Start timing from here.

RamanSpectraBaseline = cell( size(RamanSpectraInt) );

% Two sets of parameters here for the two baseline corrections; see below.
smoothwidth1=40;
smoothwidth2=60;

frame1=500;
frame2=800;

iterations1=15;
iterations2=3;

for i=1:size(RamanSpectra,2)
    disp(['Correcting spectrum ' num2str(i) ' of '
num2str(length(RamanSpectraBaseline))])

    % 1. Remove the maximum possiible of the baseline; parameters ending
'1'
    [RamanSpectraB{i}, varargout{i}] = f_baseline_corr(Xintvalues,...
        RamanSpectraInt{i},smoothwidth1,frame1,iterations1);
        % By adding a final argument 'plot2D' as suggested by the comment,
        % will generate plots showing the baseline corrections.

    % 2. Help remove the zero baseline ; parameters ending '2'.
```

```
    [RamanSpectraB2{i}, varargout{i}] = f_baseline_corr(Xintvalues,...
        RamanSpectraB{i},smoothwidth2,frame2,iterations2);

    RamanSpectraBaseline{i}=RamanSpectraB2{i}.';
end

toc   % Display time from the previous 'tic' command.


%% Smoothing of the spectra
fprintf('Smoothing the spectra \n')

tic

%%can change the iterations and polynomial here if need be (from 2,17)
for i=1:size(RamanSpectraBaseline,2)
    RamanSpectraSmooth{i}=sgolayfilt(RamanSpectraBaseline{i},2,17);
end

toc

%% Truncating the spectra
%Could truncate down to fingerprint region here. May be easier than
zapping
%in the next section (zapping gets rid of section from 1800 - 2800,
keeping
%the high freq region - not sure if this is necessary, i find it hassle.)
fprintf('Truncating the spectra \n')

tic

lowerlim  = 500;  % units of cm^{-1}
higherlim = 3200;

% Get the array indices corresponding to the above values.
%May need to switch these round depending on if xdata is ascending or
%descending
higherlimIndex  = find(floor(Xintvalues)==lowerlim,1);
lowerlimIndex = find(floor(Xintvalues)==higherlim,1);

% Remove values outside of the above index range.
XvaluesTrunc = [Xintvalues(lowerlimIndex:higherlimIndex)];

% Same for all of the smoothed spectra.
for i=1:length(RamanSpectraSmooth)

RamanSpectraTrunc{i}=RamanSpectraSmooth{i}(lowerlimIndex:higherlimIndex);
end

toc


%% Normalization of the spectra
fprintf('Normalizing the spectra \n')

tic
```

```
% Intensity of the peaks may vary due to different setups, so only
relative
% intensity between peaks can be considered real data.
for i=1:length(RamanSpectraTrunc)

      % For normalization to the amide peak
     x2=find(round(XvaluesTrunc)==1600);
      x1=find(round(XvaluesTrunc)==1700);



      % If we assume that the baseline subtraction is good, then we can
      % just normalize by the maximum
      RamanSpectraNorm{i}=RamanSpectraTrunc{i}/... %was RamanSpectraZap
but trying out no zap
          max(sgolayfilt(RamanSpectraTrunc{i}(x1(1):x2(1)),2,31));
end

%%%toc

%% Data Matrix
% Get the colums of intensities to analyse in a matrix:
fprintf('Creating the Data Matrix \n')

tic

DataMatrix=[];
for i=1:size(RamanSpectraNorm,2)
    DataMatrix(:,end+1)=RamanSpectraNorm{i};

end

toc

%% Calculate the Average and STD
Average = mean(DataMatrix,2);
Error   = std (DataMatrix,0,2);


%% Plot figures
fprintf('Plotting Figure \n')
figure = figure
subplot(2,1,1)

plot(XvaluesTrunc, Average)
xlabel('Wavenumber / cm^{-1}')
ylabel('Raman Intensity ')

%hold on
subplot(2,1,2)
plot(XvaluesTrunc,Average)
xlim([500 1800])

xlabel('Wavenumber / cm^{-1}')
ylabel('Raman Intensity ')
xticks(500:100:1800)
%hold on
```

```
%% Saving Data
    disp('saving data')
    cd(savePath)
    filename = [NameSaving{j} '.tif']
    save([NameSaving{j} 'datasetanalysed'],...

'DataMatrix','Average','Error','XvaluesTrunc','RamanSpectra','CalibrationP
eakTemp')
    saveas(figure, filename)
%waitforbuttonpress

% Clear data for next dataset
    clear RamanSpectraB RamanBackgroundSpectra ...
        RamanBSpectraInt RamanSpectraBaseline RamanSpectraNorm ...
        RamanSpectraQuartz RamanSpectraSmooth RamanSpectraTrunc ...
        RamanBSpectraInt RamanSpectraInt RamanSpectraZero...
        BackgroundAverage BackgroundAverageZero...
        DataMatrix RamanSpectraZap

    %%
end
```

EMSC Analysis_2.mat

```
%% Preparation of the datasets to load into data analysis script


%% WDF FILE IMPORTING
clear all
close all
clc

% All lines to be edited have been labelled with an X

% add path for emsc folder & raw data
dirPath={'X'...
              };

% Add MATLAB additional files
addpath('X')
% EMSC function path
addpath('X')
% For the background, add the path where the background data is (reference
% dataset and analyseddataset from previous matlab script is also in this
folder!!):
addpath('X')


savingPath='X';

% The file names are always an identifier + a number: (I'm assuming this
% means namesaving)

%raw data path for testing EMSC function
cd('X');
wdf_files = dir('*.wdf');

Name{1,1}='X';
Name{1,2}='';      % No text after the repeat index for this set of runs.
Index{1}=[X:X];
CalibrationPeak{1}=520.5;
NameSaving{1}='X'; %the analysed file you want to run has to start with
this name also and keep it in the same folder as raw data


selectedDataset=1;


%% Optimizing the parameters
% This section gives a summary plot of what the spectra will look like
% after the different stages of dataset creation.
% This should help you choose the right parameters before you run the
whole
% thing

datasettesting=1; % Choose which dataset you are testing (if only 1 then
change to 1)
cd(dirPath{datasettesting});
```

```
indextesting=1; % Choose one spectrum to test on.
% Creating the structure file with the map inside
MapTesting=WdfReader([Name{datasettesting} '.wdf']);
% Obtaining the spectra from the structure file, where each row is a
% spectrum
RamanSpectratesting=MapTesting.GetSpectra(1,MapTesting.Count);
% Changing it to columns because that is how I usually work with them
RamanSpectratesting=RamanSpectratesting';
Wavenumbers=MapTesting.GetXList'; %added ' at the end to flip values to
1948x1
CalibrationPeaktesting=CalibrationPeak{datasettesting};
% Choose the final Xintvalues that you want to use for the spline
% interpolation.
% Make sure you choose values withing the range you have acquired, or the
% spline will go very weird. Typically go for 750-1800 for FP region
Xintvalues=[500:0.5:3200];


RamanSpectraInttesting=spline(Wavenumbers+(520.5-
CalibrationPeak{datasettesting}),...
    RamanSpectratesting(:,indextesting),Xintvalues);



% Import the background and reference data for the correction with EMSC
% For background (b)
load('X_datasetraw.mat')
% Variables called "Raman Spectra","Wavenumbers" and "CalibrationPeakTemp"
for k=1:size(RamanSpectra,2)
   b(:,k)=spline(Wavenumbers +(520.5-
CalibrationPeakTemp),RamanSpectra{k}(:,2),Xintvalues);
end

% For biological reference (r)
load('X_datasetraw')
% Variables called "Ramtest3=test2'an Spectra","Wavenumbers" and
"CalibrationPeakTemp"
for k=1:size(RamanSpectra,2)
   r(:,k)=spline(Wavenumbers+(520.5-
CalibrationPeakTemp),RamanSpectra{k}(:,2),Xintvalues);
end

% Choose the parameters for the EMSC correction

N=1;
r_EMSC=r.';
b_EMSC=b.';
[RamanSpectraEMSCtesting,background,c_r,c_b,B_N] =
EMSC(RamanSpectraInttesting,r_EMSC,b_EMSC,N);

% Choose the parameters for the baseline subtraction
addpath('X')
smoothwidth=10;
bwidth=400;
iterations=10;
[RamanSpectraBaselinetesting, baseline]=f_baseline_corr(Xintvalues,...
        RamanSpectraEMSCtesting,smoothwidth,bwidth,iterations);
```

```
% Plot the whole thing so you get to change the parameters :)
figure

subplot(2,2,1)
title('Raw data')
hold on
plot(Wavenumbers,RamanSpectratesting(:,indextesting))
xlabel('Wavenumber / cm^{-1}')
ylabel('Raman Intensity / counts')
axis tight

subplot(2,2,2)
title('EMSC corrected data')
hold on
plot(Xintvalues,RamanSpectraInttesting,'k')
plot(Xintvalues,RamanSpectraEMSCtesting,'b')
plot(Xintvalues,background,'r')
legend('Average Raw Raman Spectra','EMSC-corrected Raman Spectra',...
    'Subtracted PDMS Background')
xlabel('Wavenumber / cm^{-1}')
ylabel('Raman Intensity / counts')
axis tight

subplot(2,2,3)
title('Baseline')
hold on
plot(Xintvalues,RamanSpectraEMSCtesting)
plot(Xintvalues,baseline,'r')
xlabel('Wavenumber / cm^{-1}')
ylabel('Raman Intensity / counts')
legend('EMSC-corrected Raman Spectra','Baseline fitted')
axis tight

subplot(2,2,4)
title('Baseline subtracted data')
hold on
plot(Xintvalues,RamanSpectraBaselinetesting)
xlabel('Wavenumber / cm^{-1}')
ylabel('Raman Intensity / counts')
axis tight

MapTesting.Close
clear MapTesting

%% Loading the data
fprintf('Loading the Data \n')
tic

for j=selectedDataset
    cd(savingPath)
    try load([NameSaving{j} 'datasetraw' '.mat']) %move from output to the
same folder as the PDMS and biological reference data
    catch
        warning('Raw dataset could not be found, proceeding to load raw
wdf files');
    end
```

```
    fprintf(['Loading data from dataset' num2str(j) '\n']);
    cd(dirPath{j});

       m=1;


    % Loop over all repeats for this data set.
    for i=Index{j}

        % Get the filename for a single repeat.
        FileName{j,m}=strcat(Name{j,1},num2str(i),Name{j,2},'.wdf');
        fprintf( "- loading repeat: %s\n", FileName{j,m} ); % Check.

        % Import the ".wdf" file for this repeat.
        wdf=WdfReader(FileName{j,m});
        RamanSpectra{m}(:,1)=wdf.GetXList.';
        RamanSpectra{m}(:,2)=wdf.GetSpectra(1,1).';
        wdf.Close()

        m=m+1;      % Count the number of data repeats.
    end

    % Creating the structure file with the map inside
    Map=WdfReader([Name{j} '.wdf']);
    if exist('RamanSpectra')==0
        % Obtaining the spectra from the structure file, where each row is
a
        % spectrum
        RamanSpectra=Map.GetSpectra(1,Map.Count);
        % Changing it to columns because that is how I usually work with
them
        RamanSpectra=RamanSpectra';
        Wavenumbers=Map.GetXList;
        CalibrationPeakTemp=CalibrationPeak{j};
        cd(savingPath)
        save([NameSaving{j} '_datasetraw_emsc9623'],...
            'RamanSpectra','Wavenumbers','CalibrationPeakTemp')
    end
    toc

%% X alignment of the spectra respect to the Calibration peak position in
Si

% Prior to each experiment, the system should be aligned as closely as
possible
% to the 520.5 cm^-1 peak of Silicon. Sometimes the peak can be around
520.3 or 520.8.
% This peak will be used to align the Shift of the spectra and reduce the
% variability between days.

% As the PCA only considers the Y-values for analysis,
% a spline function should be fit and the Y-values should be guarantee to
% have exactly the same X.
fprintf('Aligning spectra and fitting to the selected X values using a
spline \n')
tic
RamanSpectraInt=NaN(length(Xintvalues),Map.Count);
```

```
for k=1:length(RamanSpectra)
  RamanSpectraInt(:,k)=spline(RamanSpectra{k}(:,1)+(520.5-
CalibrationPeak{j}),RamanSpectra{k}(:,2),Xintvalues);
end
toc

clear RamanSpectra Wavenumbers


%% Full correction of all spectra with EMSC

% Usign the EMSC function for correction with multiple background spectra.

disp('EMSC correcting')

tic
r_EMSC=r.';
b_EMSC=b.';

% smoothing of the reference and background files to reduce noise in the
% EMSC output
r_EMSC=[sgolayfilt(r_EMSC.',2,17)].';
b_EMSC=[sgolayfilt(b_EMSC.',2,17)].';

RamanSpectraEMSC=NaN(size(RamanSpectraInt));

for i=1:size(RamanSpectraEMSC,2)
    [RamanSpectraEMSC(:,i),background,c_r,c_b,B_N] =
EMSC(RamanSpectraInt(:,i).',r_EMSC,b_EMSC,N);
end
toc

%% Subtract the baseline
fprintf('Subtracting the baseline \n')

tic
RamanSpectraBaseline=NaN(size(RamanSpectraEMSC));
for i=1:size(RamanSpectraEMSC,2)
    if mod(i,20)==0
        disp(['Baseline correcting spectrum ' num2str(i) ' of '
num2str(size(RamanSpectraEMSC,2))])
    end
    % First step to remove the maximum possible of the baseline
    if i==1
        tic
    end
    [RamanSpectraBB, ~]=f_baseline_corr(Xintvalues,...
        RamanSpectraEMSC(:,i),smoothwidth,bwidth,iterations);%,'plot2D');
    RamanSpectraBaseline(:,i)=RamanSpectraBB.';
    if i==10
        elapsedTime = toc;
    end
    if mod(i,20)==0
        disp(['Time left ' num2str(round(elapsedTime*(Map.Count-i)/600))
'min'])
    end
```

```
    clear RamanSpectraBB
end
toc
clear RamanSpectraB RamanSpectraEMSC


%% Data Matrix
fprintf('Creating the Data Matrix \n')
tic
DataMatrix=RamanSpectraBaseline;

%clear RamanSpectraBaseline

%% Calculate the Average and STD

Average=mean(DataMatrix,2);
Error=std(DataMatrix,0,2);
Average2=flip(Average, 2);

%% Plot figures
fprintf('Plotting Figure \n')
figure
subplot(2,1,1)

plot(Xintvalues, Average2)
xlabel('Wavenumber / cm^{-1}')
ylabel('Raman Intensity ')

%hold on
subplot(2,1,2)
plot(Xintvalues,Average2)
xlim([500 1800])

xlabel('Wavenumber / cm^{-1}')
ylabel('Raman Intensity ')
xticks(500:100:1800)
%hold on

%% Saving Data
    cd(savingPath)
    save([NameSaving{j} 'X'],...
        'DataMatrix','Average','Error',...
        'CalibrationPeakTemp','Xintvalues')


end

disp('Done! :)')
```

CsvConversion.py

"""

Created on Mon Apr 17 20:22:56 2023

Mat->CSV

@author: stephenevans

"""

```
import numpy as np
import sys
import scipy.io as sio
from scipy import interpolate
import pandas as pd
#import pyqtgraph as pg
#from PyQt5.QtCore import Qt
#from pyqtgraph.Qt import QtCore, QtGui,QtWidgets
from PyQt5 import uic
from PyQt5 import QtWidgets
from PyQt5.QtWidgets import ( QFileDialog,QMainWindow)
from sklearn.metrics import auc
import matplotlib
matplotlib.use('Qt5Agg')

from matplotlib.figure import Figure
from matplotlib.backends.backend_qt5agg import (NavigationToolbar2QT as
NavigationToolbar)
from matplotlib.backends.backend_qt5agg import FigureCanvasQTAgg as FigureCanvas
```

```
# set-up graphs - 2 figures

class MplCanvas(FigureCanvas):
    def __init__(self,parent=None,width=5,height=4,dpi=150):

        fig = Figure(figsize=(width, height), dpi=dpi)
        self.axes = fig.add_subplot(211)
        self.axes2 = fig.add_subplot(212)



        fig.tight_layout(pad=5)
        #print ('in canvas')
        super().__init__(fig)
        #super(MplCanvas(),self).__init__(fig)

class MainWindow(QMainWindow):

    def __init__(self):
        super().__init__()
        QMainWindow.__init__(self)
        self.ui=uic.loadUi('Mat_to_CSV_GUI.ui',self)
        self.show()



        global fileCount,MessString,df2,df3,dframe,classDict, cols, cols2,color_by_name
        color_by_name=['tab:blue','tab:orange','tab:green','tab:red','tab:purple',
                'tab:brown','tab:pink','tab:gray','tab:olive','tab:cyan']
        classDict={} # used to store data and labels
```

```
fileCount=1 # number of classes / different tyoes of cell / material
MessString="Report:"
self.spinBox.setSingleStep(1) # set in gui set-up next time compile
self.setWindowTitle("PCA Analysis")
self.pushButton_3.clicked.connect(self.getfiles)
self.pushButton_5.clicked.connect(self.end)
self.pushButton_6.clicked.connect(self.clearAll)
self.pushButton_9.clicked.connect(self.save)

self.pushButton_8.setCheckable(True)
self.pushButton_8.toggle()

spc=self.spinBox.value()
self.pushButton_8.clicked.connect(self.spectrum_to_display)
self.pushButton_8.toggle()

self.spinBox.valueChanged.connect(self.spectrum_to_display)
#self.spinBox_3.valueChanged.connect(self.num_outliers)

# Message box
MessString="Messages\n"
self.textEdit.insertPlainText(MessString)



# plot set up
self.canvas=MplCanvas(self, width=8, height=9, dpi=100)
self.ui.verticalLayout_2.addWidget(self.canvas)
self.navi_toolbar = NavigationToolbar(self.canvas, self) #createa navigation toolbar for
our plot canvas
```

```
        self.verticalLayout_2.addWidget(self.navi_toolbar)


        # it is functional but certain elements not yet implemented - eg dark mode, toggling grids, saving outputs
        Message="\n Step1. Open *.mat file\n Step2. Use spinbox to select outlier (red curve) on lower graph\n Step3. Hit the Zap button\
      button to remove outlier\n Step4. Repeat steps 2 & 3 \n Step5. Press OK to save, spline interpolated, spectra *.csv\n\n"
        self.textEdit.insertPlainText(Message)



    def save(self):
        #print("old file name" ,fnam)
        temp=fnam.replace('.mat', '.csv')
        fname=temp
        print("new file name:",fname)
        print("df3=",len(df3.columns))
        #remove mean and std before saving
        if "mean" in df3.columns:
            del df3['mean']
            del df3['std']
        print("df3=",len(df3.columns),ncols3)


        col_name3=[]
        # for Raman spectra - then repeat for background
        #rename columns to account for those removed
        for spec in range(ncols3):
          if spec==0:
              col_name3=["Wavenumber"]
          else:
              colName="Data"+str(spec)
```

```
            col_name3.append(colName)
        df3.columns=col_name3
        df3.to_csv(fname,index=False)
        print("saved data")

def spectrum_to_display(self):
    self.spinBox.setMaximum(ncols)
    spc=self.spinBox.value()
    self.zap(spc)



def zap(self,spc):
    global df3, col_name3,ncols3

    #self.spinBox.setMaximum(ncols)
    #spc=self.spinBox.value()




    if "mean" in df3.columns:
        del df3['mean']
        del df3['std']

    #
    x=df2["'Wavenumber"]
    #df2['mean'] = df2.iloc[:, 1:ncols].mean(axis=1)
    #df2['std'] = df2.iloc[:, 1:ncols].std(axis=1)

    y=df2['mean']
```

```
dy=df2["std"]
#self.canvas.axes.cla()

self.dataplot(x,y,dy,fileCount)



num=spc-1



outlier=sortedArr[[num]]
outlier=int(np.take(outlier,0))
outlierColumn="Data"+str(outlier)

print("sb value=",spc," df2 column =",outlierColumn)

#MessString="spectrum ("+str(spc)+")="+outlierColumn+"/"+str(ncols)+"\n"
#self.textEdit.insertPlainText(MessString)



y2=df2[outlierColumn]
self.dataplot2(x,y2,fileCount)



ncols3=len(df3.columns)
print("df3 length", ncols3,"\n")
if self.pushButton_8.isChecked():

  self.pushButton_8.toggle()

  #print('zapping column',outlier)
```

```
        df3=df3.drop(columns=outlierColumn)
        ncols3=len(df3.columns)
        print('zapping column',outlier," df3 length  = ",ncols3," columns")
        MessString="Zapping "+outlierColumn +"\n"
        self.textEdit.insertPlainText(MessString)



    x=df3["Wavenumber"]
    #df3['mean'] = df3.iloc[:, 1:ncols3].mean(axis=1)
    #df3['std'] = df3.iloc[:, 1:ncols3].std(axis=1)

    #y=df3['mean']
    #dy=df3["std"]
    y= df3.iloc[:, 1:ncols3].mean(axis=1)
    dy=df3.iloc[:, 1:ncols3].std(axis=1)
    self.dataplot3(x,y,dy)




def getfiles(self):
    global col_name,ncols,fnam, corrFactor, dataArray,dataSplineArray,sortedArr,df2,df3
    dlg = QFileDialog()
    dlg.setNameFilter('(*.mat)')
    dlg.setFileMode(QFileDialog.AnyFile)
```

```
df2 = pd.DataFrame()
if dlg.exec_():

    filenames = dlg.selectedFiles()
    fnam=filenames[0]
    print("open",fnam)

    sep="/"
    fnamLen=len(fnam)
    lastSep=fnam.rfind(sep)
    #pathName=fnam[:lastSep+1]  # identify path
    fileName=fnam[lastSep+1:fnamLen]
    MessString=fileName+"\n"
    self.textEdit.insertPlainText(MessString)



    self.pushButton_3.setEnabled(False)
    mat_contents = sio.loadmat(fnam)
    print(mat_contents.keys())

    for key in mat_contents.keys():

       if key=="__header__":
          header=mat_contents['__header__']
          print("_____________header_____________")
          #print(header)
          print("     ")
       if key=="__version__":
```

```
    version=mat_contents['__version__']
    print("_____version_____")
    #print(version)
    print("    ")
if key=="__globals__":
    glob=mat_contents['__globals__']
    print("_____globals_____")
    #print(glob)
    print("    ")
if key=='CalibrationPeakTemp':
    cal_peak=mat_contents['CalibrationPeakTemp']
    cal_peak_val=cal_peak[0,0]
    print("_____calibration peak_____")
    print(cal_peak_val)
    print("    ")
if key=='RamanBackgroundSpectra':
    RamanBack=mat_contents[key]
    print("_____Background_____")
    numBackSpectra=RamanBack.size
    print("numberRamanBack",numBackSpectra)
    print("    ")
if key=='RamanSpectra':
    RamanSpectra=mat_contents[key]
    numSpectra=RamanSpectra.size

    print("_____Spectra_____")
    print("number of spectra=",numSpectra)
    print("    ")
```

```
corrFactor=round((520.5-cal_peak_val),1)
print("correction factor", corrFactor)
col_name=["Wavenumber"]
# for Raman spectra - then repeat for background
for spec in range(numSpectra):
  data=RamanSpectra[0,spec]
  lnData=len(data)
  colName="Data"+str(spec+1)

  col_name.append(colName)
  data_x=data[:,0]+corrFactor
  data_y=data[:,1]

  #translate in y  so min value=0
  data_y_min=np.amin(data_y)
  data_y=data_y-data_y_min

  if data_x[0]>data_x[-1]:
    #print("inverting data",spec) to make low to high wavenumber
    data_x=data_x[::-1]
    data_y=data_y[::-1]

 #take x data to 1 dp
  test_x=np.round(data_x,decimals=1)


  if spec==0:
  #write x,y
     dataArray=np.column_stack((data_x,data_y))
     xnew,ynew=self.spliInt(spec)
```

```
      dataSplineArray=np.column_stack((xnew,ynew))
    else:
    #write y
      dataArray=np.column_stack((dataArray,data_y))
      xnew,ynew=self.spliInt(spec)
      dataSplineArray=np.column_stack((dataSplineArray,ynew))

#  create data frame at end from the arrays - this stops fragmentation
df= pd.DataFrame(dataArray)
df2= pd.DataFrame(dataSplineArray)
df.columns=col_name
df2.columns=col_name
df2=df2.round(1)
#print(df)



df3 = pd.DataFrame()
df3=df2.copy(deep=True)
flag="showAll"
ncols=numSpectra

x=df2["Wavenumber"]
df2['mean'] = df2.iloc[:, 1:ncols].mean(axis=1)
df2['std'] = df2.iloc[:, 1:ncols].std(axis=1)

y=df2['mean']
dy=df2["std"]
self.dataplot(x,y,dy,fileCount)
print(df2)
```

```
#determine outliers
col_name=df2.columns
#print(col_name)
# calc area under curve using skilearn auc function
meanArea=round(auc(x,y),2)
MessString="Area under mean = "+str(meanArea)+"\n"
self.textEdit.insertPlainText(MessString)

#find outlier spectra
colArea=[]
iCount=[]
I=0
for col in col_name:
   if col=="Wavenumber":
      x=df2.Wavenumber
   elif col=="mean":
      pass
   elif col=="std":
      pass
   else:
      I=I+1
      y=df2[col]
      newArea=abs(round(auc(x,y)-meanArea,2))
      colArea.append(newArea)
      iCount.append(I)
areaArray=np.column_stack((iCount,colArea))
#print (areaArray)
print(" ")
sortedArr = areaArray[areaArray[:,1].argsort()]
```

```
        print("_____ Sorted_____")
        sortedArr=sortedArr[::-1]
        print (sortedArr)

        #show 3 largest to zap

        out_first=sortedArr[[0]]
        out_first=int(np.take(out_first,0))

        print("first=",out_first)
        first="Data"+str(out_first)
        y2=df2[first]
        self.dataplot2(x,y2,fileCount)
        MessString="Outlier = "+str(out_first)+"\n"
        self.textEdit.insertPlainText(MessString)




def spliInt(self,spect):
    spect=spect+1
    #Spline Interpolation
    x=dataArray[:,0]
    y=dataArray[:,spect]
    xStart=2*round(x[0]/2);xEnd=2*round(x[-1]/2)#1947
   # print("xstart,xstop",xStart,xEnd)
    s=interpolate.InterpolatedUnivariateSpline(x, y)
    xnew=np.arange(xStart,xEnd,0.5) #+ corrFactor
    ynew=s(xnew)
```

```
        xnew=xnew[::-1]
        ynew=ynew[::-1]
        return xnew,ynew

#23456
    def dataplot(self,x,y,dy,fileCount):
        #Average Data

        if fileCount==1:
            self.canvas.axes.cla()
            self.canvas.axes.set_xlabel("Wavenumber (cm-1)",fontsize=12)
            self.canvas.axes.set_ylabel("Intensity (arb. units)",fontsize=12)
            self.canvas.axes.set_title("Raman Spectra",fontsize=14)
        #offset_ratio=1.5
        #y=y+(fileCount*offset_ratio)    # offset spectra
        pcol=color_by_name[fileCount-1]
        self.canvas.axes.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1.0)
        self.canvas.axes.fill_between(x,y+dy,y-dy,facecolor=pcol,linewidth=1.5,alpha=0.4)

        cb2=self.checkBox_2.isChecked()
        if cb2==True:
            self.canvas.axes.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
        #self.canvas.axes.legend()
        self.canvas.draw ()

    def dataplot2(self,x,y,fileCount):
        #Average Data

        if fileCount==1:
            self.canvas.axes2.cla()
```

```
        self.canvas.axes2.set_xlabel("Wavenumber (cm-1)",fontsize=12)
        self.canvas.axes2.set_ylabel("Intensity (arb. units)",fontsize=12)
        self.canvas.axes2.set_title("Raman Spectra",fontsize=14)


    pcol=color_by_name[3]
    self.canvas.axes.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1.0)


    cb2=self.checkBox_2.isChecked()
    if cb2==True:
        self.canvas.axes2.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
    #self.canvas.axes.legend()
    self.canvas.draw ()

def dataplot3(self,x,y,dy):
    #Average Data

    self.canvas.axes2.cla()
    self.canvas.axes2.set_xlabel("Wavenumber (cm-1)",fontsize=12)
    self.canvas.axes2.set_ylabel("Intensity (arb. units)",fontsize=12)
    self.canvas.axes2.set_title("Raman Spectra after Zapping",fontsize=14)

    pcol=color_by_name[fileCount-1]
    self.canvas.axes2.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1.0)
    self.canvas.axes2.fill_between(x,y+dy,y-dy,facecolor=pcol,linewidth=1.5,alpha=0.4)

    cb2=self.checkBox_2.isChecked()
    if cb2==True:
        self.canvas.axes2.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
```

```
        #self.canvas.axes.legend()
        self.canvas.draw ()




#23456
    def clearAll(self):
        print("Quit")
        app.quit()

    def end(self):
        print("Quit")
        app.quit()



if __name__=="__main__":
    app=QtWidgets.QApplication(sys.argv)
    w=MainWindow()
    w.show()
    app.exec_()
```

RamanPreProcessing_4.py

```
"""
import numpy as np

import sys
import pandas as pd

from PyQt5 import uic
from PyQt5 import QtWidgets
from PyQt5.QtWidgets import ( QFileDialog,QMainWindow)

import matplotlib
matplotlib.use('Qt5Agg')
from scipy.signal import savgol_filter
from matplotlib.figure import Figure
from matplotlib.backends.backend_qt5agg import (NavigationToolbar2QT as
NavigationToolbar)
from matplotlib.backends.backend_qt5agg import FigureCanvasQTAgg as FigureCanvas

from BaselineRemoval import BaselineRemoval
from scipy import sparse
from scipy.sparse.linalg import spsolve



class MplCanvas(FigureCanvas):
   def __init__(self,parent=None,width=5,height=4,dpi=300):

      fig = Figure(figsize=(width, height), dpi=dpi)
```

```
        self.axes = fig.add_subplot(211)
        self.axes2 = fig.add_subplot(212)
        fig.tight_layout(pad=5)
        #print ('in canvas')
        super().__init__(fig)
        #super(MplCanvas(),self).__init__(fig)

class MainWindow(QMainWindow):

    def __init__(self):
        super().__init__()
        QMainWindow.__init__(self)
        self.ui=uic.loadUi('Raman_Processing_GUI.ui',self)
        self.show()


        global fileCount,MessString,df2,dframe,classDict, cols, cols2,color_by_name
        color_by_name=['tab:blue','tab:orange','tab:green','tab:red','tab:purple',
                'tab:brown','tab:pink','tab:gray','tab:olive','tab:cyan']
        classDict={} # used to store data and labels

        fileCount=0 # number of classes / different tyoes of cell / material
        MessString="Report:"

        self.setWindowTitle("Raman Processing")
        self.pushButton_3.clicked.connect(lambda:self.getfiles())

        #limits for truncate

        self.pushButton_8.clicked.connect(lambda:self.truncate())
```

```
        self.pushButton_7.clicked.connect(lambda:self.preview_baseline())
        self.pushButton_5.clicked.connect(lambda:self.apply_to_all())
        self.pushButton_6.clicked.connect(lambda:self.preview_smooth())
        self.pushButton_9.clicked.connect(lambda:self.smooth_all())
        self.spinBox_3.valueChanged.connect(lambda:self.preview_smooth())
        self.spinBox_8.valueChanged.connect(lambda:self.preview_smooth())
        self.pushButton.clicked.connect(lambda:self.normalise())
        self.pushButton_2.clicked.connect(lambda:self.mean_sd())
        self.pushButton_4.clicked.connect(lambda:self.save())



        # Message box
        MessString="Files Loaded (and colour)\n"
        self.textEdit.setPlainText(MessString)


        # plot set-up
        self.canvas=MplCanvas(self, width=8, height=9, dpi=100)
        self.ui.verticalLayout_2.addWidget(self.canvas)
        self.navi_toolbar = NavigationToolbar(self.canvas, self) #createa navigation toolbar for our plot canvas
        self.verticalLayout_2.addWidget(self.navi_toolbar)


        # it is functional but certain elements not yet implemented - eg dark mode, toggling grids, saving outputs



    def save(self):
    #fname=fnam+"analysed"
        temp=fnam
```

```
        print("old file name",temp)
        temp2=temp.replace('raw', 'analysed')
        print("new file name",temp2)

        df4.to_csv(temp2,index=False)
        text=str(temp2)
        self.textEdit.setPlainText(str(text))




    def mean_sd(self):
        df4['mean'] = df4.iloc[:, 1:ncols].mean(axis=1)
        df4['SD1'] = df4.iloc[:, 1:ncols].std(axis=1)
        x=df4.Wavenumber
        y=df4['mean']
        dy=df4['SD1']
        flag=3
        self.dataplot3(x,y,dy,flag)



    def normalise (self):
        # zoom in to noise and to sharp peak to optimize smooth without reducing peak height
significantly
        print("in normalise")
        xval=self.doubleSpinBox_2.value() # get value to normalise
        xpos=(df4[df4["Wavenumber"]==xval].index.values)
        xpos=np.take(xpos,0)
```

```
        text="Normalised against peak at "+ str(xval)+" (cm-1) \n"
        print(text)
        self.textEdit.setPlainText(text)



        for coln in col_name:
           if coln=='Wavenumber':
              pass
           else:
              y=df4[coln]
              ynorm=y[xpos]
              #print("x,ynorm",xpos,ynorm)

              ynew=y/ynorm
              df4[coln]=ynew
        self.dataplot(df4)




    def preview_smooth(self):
        # zoom in to noise and to sharp peak to optimize smooth without reducing peak height significantly
        print("in smooth preview")
        polynom=self.spinBox_3.value()     # polynomial
        wind=self.spinBox_8.value()        # window
        print("poly= ",polynom,"window= ",wind, "(w/p)=", str(wind/polynom))

        x=df3["Wavenumber"]
        y=df3["Data1"]
```

```
        flag=0
        self.dataplot2(x,y,flag)

        Iny_SGS = savgol_filter(y, wind, polynom) # window size 51, polynomial order 3
        flag=1
        self.dataplot2(x,Iny_SGS,flag)

    def smooth_all(self):
        global df4
        print("in smooth all")
        polynom=self.spinBox_3.value()     # polynomial
        wind=self.spinBox_8.value()  # window
        text="Smooth: poly= "+str(polynom)+" window= "+str(wind)+ " (w/p)= "+
str(wind/polynom)+"\n"
        print(text)

        self.textEdit.setPlainText(text)

        df4 = pd.DataFrame()# put smoothed data here


        for col in col_name:
          if col=="Wavenumber":
             x=df3.Wavenumber
             df4[col]=x
          else:
             y=df3[col]
             Iny_SGS = savgol_filter(y, wind, polynom) # typical window size 15, polynomial
order 2  note higher w/p ratio gives stronger smoothing
             df4[col]=Iny_SGS
```

```
    #show effect for smoothing on a single spectrum
    x=df4["Wavenumber"]
    y=df4["Data1"]
    flag=0
    self.dataplot2(x,y,flag)
    flag=1
    x=df3["Wavenumber"]
    y=df3["Data1"]
    self.dataplot2(x,y,flag)




def preview_baseline(self):
    #first pick method from combo box
    choice=self.comboBox.currentIndex()
    choice_txt=self.comboBox.currentText()
    txt="baseline choice= "+ choice_txt+"\n"
    print(txt)
    self.textEdit.insertPlainText(txt)



    # plot average and work out best base line then apply to all


    x=df2["Wavenumber"]
    y = df2.iloc[:, 1:ncols].mean(axis=1)
```

```
flag=0
self.dataplot2(x,y,flag)

if choice==0:
   #ALS fit
   #Asymmetric Least Squares Smoothing" by P. Eilers and H. Boelens in 2005.

   txt="Asymmetric Least Square \n"
   print(txt)
   self.textEdit.insertPlainText(txt)

   self.textEdit.insertPlainText("uses spin box values N, lambda, and p \n")



   lam=self.spinBox_6.value()
   p=self.doubleSpinBox.value()
   niter=self.spinBox_5.value()
   lam=(lam*1000)+1E3
   p=p/1000
   print("n. iterations=", niter," lambda=",lam, " p= ",p)
   y_base=self.baseline_als_optimized(y, lam, p, niter)
   y_res=y-y_base
   flag=1
   self.dataplot2(x,y_base,flag)
   flag=2
   self.dataplot2(x,y_res,flag)

if choice==1:
   #IMOD poly
   #IMOD fit
```

```
            p=int(self.doubleSpinBox.value())
            if p<2:
               p=2
               #dont let p be less than 2
            niter=10*self.spinBox_5.value()



            grad=0.001
            txt="IMOD Poly (degree, iterations, gradient) = "+str(p)+" "+str(niter)+" " +str(grad)+"
\n"
            print(txt)
            self.textEdit.insertPlainText(txt)
            baseObj=BaselineRemoval(y)
            y_res=baseObj.IModPoly(degree=p,repitition=niter,gradient=grad)
            y_base=y-y_res
            flag=1
            self.dataplot2(x,y_base,flag)
            flag=2
            self.dataplot2(x,y_res,flag)



        if choice==2:
            #Zhang
            txt="Zhang \n"
            print(txt)
            self.textEdit.insertPlainText(txt)

            self.textEdit.insertPlainText("uses spin box values N, lambda \n")

            niter=self.spinBox_5.value()
```

```
        lam=self.spinBox_6.value()

        baseObj=BaselineRemoval(y)
        y_res=baseObj.ZhangFit(lambda_=lam, porder=1, repitition=niter)
        y_base=y-y_res
        flag=1
        self.dataplot2(x,y_base,flag)
        flag=2
        self.dataplot2(x,y_res,flag)
```

```
#2345
    def apply_to_all(self):
        global df3
        temp_array=df2.to_numpy()
        x=df2["Wavenumber"]

        choice=self.comboBox.currentIndex()
        if choice==0: #ALS fit
            lam=self.spinBox_6.value()
            p=self.doubleSpinBox.value()
            niter=self.spinBox_5.value()
            lam=(lam*1000)+1E3
            p=p/1000

            #x=df2["Wavenumber"]
```

```
            temp_array[:,0]=x
            for i in range(1,ncols):
                y=temp_array[:,i]
                yb=self.baseline_als_optimized(y, lam, p, niter)
                yr=y-yb
                temp_array[:,i]=yr
            #create new dataframe of baseline corrected spectra
            df3= pd.DataFrame(temp_array)
            df3.columns=col_name
            outText="Parameters: N;lambda; poly "+str(niter)+";"+ str(lam)+"; "+str(p)+" \n"

            self.textEdit.setPlainText(outText)
            self.dataplot(df3)

        if choice==1:
            #IMOD poly
            #IMOD fit
            p=int(self.doubleSpinBox.value())
            if p<2:
               p=2
               #dont let p be less than 2
            niter=10*self.spinBox_5.value()



            grad=0.001
            txt="IMOD Poly (degree, iterations, gradient) = "+str(p)+" "+str(niter)+" " +str(grad)+"
\n"
            print(txt)
            self.textEdit.insertPlainText(txt)
            temp_array[:,0]=x
```

```
    for i in range(1,ncols):
            y=temp_array[:,i]
            baseObj=BaselineRemoval(y)
            yr=baseObj.IModPoly(degree=p,repitition=niter,gradient=grad)
            temp_array[:,i]=yr
        #create new dataframe of baseline corrected spectra
    df3= pd.DataFrame(temp_array)
    df3.columns=col_name
    self.dataplot(df3)

if choice==2: #Zhang
    niter=self.spinBox_5.value()
    lam=self.spinBox_6.value()
    temp_array[:,0]=x
    for i in range(1,ncols):
            y=temp_array[:,i]
            baseObj=BaselineRemoval(y)
            yr=baseObj.ZhangFit(lambda_=lam, porder=1,repitition=niter)
            temp_array[:,i]=yr
        #create new dataframe of baseline corrected spectra
    df3= pd.DataFrame(temp_array)
    df3.columns=col_name
    outText="Parameters: N;lambda; poly "+str(niter)+";"+ str(lam)+"; "+str(1)+" \n"

    outText="Baseline: Zhang  \n"
    self.textEdit.setPlainText(outText)
    self.dataplot(df3)
```

```
    def baseline_als_optimized(self, y, lam, p, niter):
        L = len(y)
        D = sparse.diags([1,-2,1],[0,-1,-2], shape=(L,L-2))
        D = lam * D.dot(D.transpose()) # Precompute this term since it does not depend on
`w`
        w = np.ones(L)
        W = sparse.spdiags(w, 0, L, L)
        for i in range(niter):
            W.setdiag(w) # Do not create a new matrix, just update diagonal values
            Z = W + D
            z = spsolve(Z, w*y)
            w = p * (y > z) + (1-p) * (y < z)
        return z




    def truncate(self):
        global df2,outText
        xmin=self.spinBox.value()
        xmax=self.spinBox_2.value()
        print("truncate between limits",xmin,xmax)
        MessString="Truncated between "+str(xmin)+" and "+str(xmax)+"\n"
```

```
        self.textEdit.setPlainText(str(MessString))

        imin=(df[df["Wavenumber"]==xmin].index.values)
        imax=(df[df["Wavenumber"]==xmax].index.values)
        imin=np.take(imin,0)
        imax=np.take(imax,0)
        x=df['Wavenumber']
        newData=df.loc[imax:imin]
        df2 = pd.DataFrame(newData)
        df2.reset_index(drop=True, inplace=True)

        self.dataplot(df2)



def getfiles(self):
    global data,col_name,df,ncols,fnam,flag,outText

    dlg = QFileDialog()
    dlg.setNameFilter('(*.csv)')
    dlg.setFileMode(QFileDialog.AnyFile)
    fileNameList=[]


    if dlg.exec_():
        filenames = dlg.selectedFiles()
        fnam=filenames[0]
```

```
        # get file & path name  (might only work for Mac i.e need to change separator?) i.e. if windows use "\"
        sep="/"
        fnamLen=len(fnam)
        lastSep=fnam.rfind(sep)
        #pathName=fnam[:lastSep+1]
        fileName=fnam[lastSep+1:fnamLen]



        outText="File: "+str(fileName)+"\n"
        self.textEdit.setPlainText(outText)


        data= pd.read_csv (filenames[0])
        ncols=len(data.columns)
        print("number of columns=",ncols)
        col_name=[]
        df = pd.DataFrame(data) #create pandas dataframe called data


       # name the columns as wavenumber or data : first create a list of names then add to data frame
    for i in range(ncols):
        if i==0:
           col_name.append('Wavenumber')
        else:
           col_name.append("Data"+str(i))


    df.columns=col_name  # add column names into df
    print(df)


    self.dataplot(df)
```

```
    x=df["Wavenumber"]
    xmax=int(max(x))
    xmin=int(min(x))
    print(xmin,xmax)

    self.spinBox.setValue(xmin)
    self.spinBox_2.setValue(xmax)




def dataplot2(self,x,y,flag):

     pcol=color_by_name[flag]
     if flag==0:
        print("in flag ==0")
        self.canvas.axes.cla()
        self.canvas.axes.set_xlabel("Wavenumber (cm-1)",fontsize=12)
        self.canvas.axes.set_ylabel("Intensity (arb. units)",fontsize=12)
        self.canvas.axes.set_title("Raman Spectra",fontsize=14)
        pcol=color_by_name[flag]
        self.canvas.axes.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1)

     elif flag==1:
        print("in flag ==1")
        pcol=color_by_name[flag]
        self.canvas.axes.plot( x,y,color=pcol,linewidth=1,linestyle='--',alpha=1)
     elif flag==2:
```

```
        pcol=color_by_name[flag-1]
        self.canvas.axes.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1)


    cb2=self.checkBox_2.isChecked()
    if cb2==True:
        self.canvas.axes.grid(color = 'gray', linestyle = '--', linewidth = 0.5)

    self.canvas.draw ()



def dataplot(self,df):
    #Average Data

    self.canvas.axes.cla()
    self.canvas.axes.set_xlabel("Wavenumber (cm-1)",fontsize=12)
    self.canvas.axes.set_ylabel("Intensity (arb. units)",fontsize=12)
    self.canvas.axes.set_title("Raman Spectra",fontsize=14)

    pcol=color_by_name[0]
    for col in col_name:
        if col=="Wavenumber":
            x=df.Wavenumber
        else:
            y=df[col]
            self.canvas.axes.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=.5)


    cb2=self.checkBox_2.isChecked()
```

```
        if cb2==True:
            self.canvas.axes.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
        #self.canvas.axes.legend()
        self.canvas.draw ()


    def dataplot3(self,x,y,dy,flag):
        #Average Data
        pcol=color_by_name[flag]

        self.canvas.axes2.cla()
        self.canvas.axes2.set_xlabel("Wavenumber (cm-1)",fontsize=12)
        self.canvas.axes2.set_ylabel("Intensity (arb. units)",fontsize=12)
        self.canvas.axes2.set_title("Raman Spectra ",fontsize=14)
        self.canvas.axes2.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1.0)
        self.canvas.axes2.fill_between(x,y+dy,y-dy,facecolor=pcol,linewidth=1.5,alpha=0.4)


        cb2=self.checkBox_2.isChecked()
        if cb2==True:
            self.canvas.axes2.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
        #self.canvas.axes.legend()
        self.canvas.draw ()




if __name__=="__main__":
    app=QtWidgets.QApplication(sys.argv)
```

```
w=MainWindow()
w.show()
app.exec_()
```

RamanPCA_5.py

"""

Created on 15/4/2023 - PCA Analysis of Raman Spectra

Files input are analysed.csv files

1) read in multiple CSV files - one at a time (though could be changed to read all at once quite easily) these should be files that are "analyzed"

2) display data - the 'mean' data column for each data class displayed

3) truncate data so that all spectra cover the same wavenumber regime and cover the regions of interest

4) perform SNV

5) PCA anaylsis using sk learn

6) display outputs

7) save in way suitable for publication - can produce tiff or other format output directly clicking on navigation bar

at the bottom of the plots - this also allows change colours. lines, etc

Use PyQT & pyqtgrpah as framework for program & displays - could be altered for Matplot lib which would give better histograms

@author: stephenevans

"""

```
import numpy as np

import sys

import pandas as pd

#import pyqtgraph as pg
```

```
#from PyQt5.QtCore import Qt
#from pyqtgraph.Qt import QtCore, QtGui,QtWidgets
from PyQt5 import uic
from PyQt5 import QtWidgets
from PyQt5.QtWidgets import ( QFileDialog,QMainWindow)

import matplotlib
matplotlib.use('Qt5Agg')
import scipy.stats as st

from matplotlib.figure import Figure
from matplotlib.backends.backend_qt5agg import (NavigationToolbar2QT as
NavigationToolbar)
from matplotlib.backends.backend_qt5agg import FigureCanvasQTAgg as FigureCanvas
from matplotlib.pyplot import cm

matplotlib.rcParams['savefig.dpi'] = 600  # sets resolution for saving file (nb tiff can be very
large - but high quality at this setting)
#from PyQt5.QtWidgets import QApplication, QMainWindow


from sklearn.decomposition import PCA
#from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
#from sklearn.preprocessing import StandardScaler
#from sklearn import svm
#from sklearn.cluster import KMeans
#from sklearn.model_selection import train_test_split, cross_val_score
#from collections import Counter
#import matplotlib
```

```
class MplCanvas(FigureCanvas):
    def __init__(self,parent=None,width=5,height=4,dpi=300):

        fig = Figure(figsize=(width, height), dpi=dpi)
        self.axes = fig.add_subplot(221)

        self.axes2 = fig.add_subplot(222)
        self.axes3 = fig.add_subplot(223,projection='3d')    # for 3D plot
        #self.axes3 = fig.add_subplot(223)     # for 2D plot
        self.axes4 = fig.add_subplot(224)

        fig.tight_layout(pad=5)
        #print ('in canvas')
        super().__init__(fig)
        #super(MplCanvas(),self).__init__(fig)

class MainWindow(QMainWindow):

    def __init__(self):
        super().__init__()
        QMainWindow.__init__(self)
        self.ui=uic.loadUi('Raman_PCA_GUI.ui',self)
        self.show()


        global fileCount,MessString,df2,dframe,classDict, cols,
cols2,color_by_name,c_map_by_name
```

```
color_by_name=['cyan','deeppink','orange','purple','green',
    'deepred','deeppink','sienna']
c_map_by_name=["Reds","Greens", "Blues","Purples","Greys","RdPu"]
classDict={} # used to store data and labels

fileCount=0 # number of classes / different tyoes of cell / material
MessString="Report:"

global pointSize,transp,barWidth
pointSize=100
transp=0.4
barWidth=0.25

self.setWindowTitle("PCA Analysis")
self.pushButton_3.clicked.connect(lambda:self.getfiles())
self.pushButton_5.clicked.connect(lambda:self.pca_fit())
self.pushButton_2.clicked.connect(lambda:self.clearAll())
self.pushButton_8.clicked.connect(lambda:self.truncate())


self.spinBox.valueChanged.connect(lambda:self.valueChange)
self.spinBox_2.valueChanged.connect(self.valueChange)
self.spinBox_3.valueChanged.connect(self.num_comp_PCA)


# changed axis so PC1 is on x-axis
self.spinBox_5.setValue(2)
self.spinBox_6.setValue(1)
self.spinBox_5.valueChanged.connect(self.pca_fit)
self.spinBox_6.valueChanged.connect(self.pca_fit)
```

```
        # Message box
        MessString="Files Loaded (and colour)\n"
        self.textEdit.insertPlainText(MessString)

        self.spinBox_8.valueChanged.connect(self.plotParams)
        self.doubleSpinBox.valueChanged.connect(self.plotParams)
        self.doubleSpinBox_2.valueChanged.connect(self.plotParams)

        #self.checkBox_6.setChecked(False)
        #self.checkBox_6.toggled.connect(self.bgndChange)

        # plot set up
        self.canvas=MplCanvas(self, width=8, height=9, dpi=100)
        self.ui.verticalLayout_2.addWidget(self.canvas)
        self.navi_toolbar = NavigationToolbar(self.canvas, self) #createa navigation toolbar for
our plot canvas
        self.verticalLayout_2.addWidget(self.navi_toolbar)

        # it is functional but certain elements not yet implemented - eg dark mode, toggling
grids, saving outputs



    def plotParams(self):
        global pointSize,transp,barWidth
        barWidth=self.doubleSpinBox_2.value()
        pointSize=self.spinBox_8.value()
        transp=self.doubleSpinBox.value()
```

```
    print("point=",pointSize," : Bar width=",barWidth," aplha=",transp)


def clearAll(self):
    self.canvas.axes.cla()
    self.canvas.axes2.cla()
    self.canvas.axes3.cla()
    self.canvas.axes4.cla()
    self.canvas.draw ()
    self.pushButton_8.setEnabled(True)
    dfSNV=[]
    df2=[]
    df_temp=[]
    fileCount=0
    fileNameList=[]
    df=[]
    # need to clear all relevant data - at present best to simply re-run program


def truncate(self):
     global XvaluesTrunc
     xlowerLimit=self.spinBox.value()
     xupperLimit=self.spinBox_2.value()
     print("truncate between limits",xlowerLimit,xupperLimit)

     #data prep truncate for all data have same stop and start
     #print("Dictionary=\n",classDict.keys())
     df_temp=pd.DataFrame()

     for ky in classDict:
        df_temp = classDict[ky]
        alldataArray=[]
```

```
    col_name=df_temp.columns
    wnbr="Wavenumber"

    #find indices of upper and lower limits
    imin=(df_temp[df_temp[wnbr]==xlowerLimit].index.values)
    imax=(df_temp[df_temp[wnbr]==xupperLimit].index.values)

    imin=np.take(imin,0)
    imax=np.take(imax,0)
    print("")
    #print("data set:",ky,imin,imax)

    # create series / array of truncated x values
    XvaluesTrunc=xFullSpectrum.truncate(imax,imin-1,copy=False);
    XvaluesTrunc.reset_index(drop=True, inplace=True)
    newData=df_temp.iloc[imax:imin,:]
    newData=newData.drop(['Wavenumber'], axis=1)
    # could use iloc[imax:imin,1:] instead of dropping Wavenumber

    newData.reset_index(drop=True, inplace=True)
    #print("new Data.head",newData.head)

    #replace the old dataframe with new truncated dataframe
    classDict[ky]=newData

#plot new truncated dataframe - plots mean of the SNV normalised data for each class
counter=0
for ky in classDict:
    counter=counter+1
```

```
        df_temp_new = classDict[ky]
        ncols=len(df_temp_new)
        x= XvaluesTrunc
        y = df_temp_new.iloc[:, 1:ncols].mean(axis=1)# calc mean for plotting
        dy=df_temp_new.iloc[:, 1:ncols].std(axis=1)
        self.dataplot(x,y,dy,counter)
    self.pushButton_8.setEnabled(False)   # truncate only allowed once
   # could add reset button and allow re-truncate?

def valueChange(self):
   # values for truncate
    global xlowerLimit,xupperLimit
    s=self.spinBox.value()
    sb1_val=s
    xlowerLimit=sb1_val
    sb2_val=self.spinBox_2.value()
    xupperLimit=sb2_val
    #print("truncate region limits",xlowerLimit,xupperLimit)

def num_comp_PCA(self):
        global numPCA_components
        sb3_val=self.spinBox_3.value()
        numPCA_components=sb3_val



def pca_fit(self):
   global XvaluesTrunc,alldataArray
   numPCA_components=self.spinBox_3.value()
```

```
print("In PCA")
print("number of components=",numPCA_components)
alldataArray=[]
count=0;counter=0

for ky in classDict:
   count=count+1
   df_temp_new = classDict[ky]
   col_name=df_temp_new.columns

   ncols=len(col_name)

   lbl=count*np.ones((ncols,1),dtype=int) #add labels to array to allow classification
   #print(lbl.shape)
   temp_array=df_temp_new.to_numpy() # convert from df to np array

   temp_array=temp_array.T   #transpose array so spectra are in rows

   #add labels (nb (shape= (num spectra, no spectral point + 1 point for lable)))
   temp_array=np.hstack((lbl, temp_array))
   print("Label / class",ky," shape",temp_array.shape)

   # add all data in one large array for PCA & LDA
   if count==1:
      alldataArray=temp_array
   else:
      alldataArray=np.concatenate((alldataArray, temp_array))

   lab = alldataArray[:,0].astype('uint8') # get labels
```

```
# perfprm PCA with top n_components
skpca1=PCA(n_components=numPCA_components)
X1=skpca1.fit(alldataArray[:,1:]) # ignore labels in col 0


expl_var_1=X1.explained_variance_ratio_

# plot variance explained (from PCA)
xpos=range(1,numPCA_components+1,1)
counter=1# set colour
self.dataplot2(xpos,expl_var_1,counter)


# plot cummulative varience
counter=2 # just to change colour
ypos=np.cumsum(expl_var_1)
self.dataplot2(xpos,ypos,counter)



# find unique labels (classes of data)
unique=list(set(lab))
#print("class identifiers (labels)",unique)

# obtain PCA components to make scatter plot
#num_components=4
skpca2 = PCA(n_components=numPCA_components)
Xt2 = skpca2.fit_transform(alldataArray[:,1:])
```

```
#print("shape of Xt2=", Xt2.shape)



xi=[];yi=[]
#choose PCAs to show
pca_x=(self.spinBox_6.value())-1
pca_y=(self.spinBox_5.value() )-1
pca_z=(self.spinBox_7.value() )-1

y_lbl="PC"+str(pca_y+1)
x_lbl="PC"+str(pca_x+1)
z_lbl="PC"+str(pca_z+1)
print("PC show=",pca_x,pca_y, pca_z)

#column in Xt2[[j,0]    provides the PCA componet - colm 0 shown here is PCA 1
# these are plotted as a function of class /label



for i,u in enumerate(unique):

   xi = [Xt2[j,pca_x] for j in range(len(Xt2[:,0])) if lab[j] == u]  #xi = PCA 1

   yi = [Xt2[j,pca_y] for j in range(len(Xt2[:,1])) if lab[j] == u]   #yi= PCA2

   zi = [Xt2[j,pca_z] for j in range(len(Xt2[:,2])) if lab[j] == u]   #yi= PCA3



   # plot 3D
   self.scattplot2(xi,yi,zi,i,x_lbl,y_lbl,z_lbl,u)
```

```
    # plot 2D
    fig4_plot=self.comboBox.currentIndex()
    if fig4_plot==2:
        if self.checkBox_4.isChecked()==True:
            self.scattplot(xi,yi,i,x_lbl,y_lbl,u)

        if self.checkBox_3.isChecked()==True:
            self.contourplot(xi,yi,i,x_lbl,y_lbl)
    elif fig4_plot==1:
        bw=barWidth
        bins=int((max(xi) - min(xi))/bw)
        #print("Bins=",bins,bw)
        freq1,x1 = np.histogram(xi,bins=bins)
        x11=x1[:-1]
        #print("bins",x11)
        self.histplot(x11,freq1,i)

    xi=[];yi=[];zi=[]


# repeat PCA but convert into loading - i.e see each componet versus wavelength
skpca3 = PCA(n_components=numPCA_components)
components = skpca3.fit_transform(alldataArray[:,1:])
loadings=skpca3.components_.T * np.sqrt(skpca3.explained_variance_)
#print("loadings.shape",loadings.shape)

fig4_plot=self.comboBox.currentIndex()
if fig4_plot==0:
    numPCA_to_plot=self.spinBox_4.value()
    for i in range(numPCA_to_plot):
```

```
            yli = [loadings[j,i] for j in range(len(loadings[:,i]))]
            self.dataplot3(XvaluesTrunc,yli,i,numPCA_to_plot)




    def getfiles(self):
        global x,y,dataArray,fileCount,MessString,df2,classDict,dfSNV,xFullSpectrum


        fileCount=fileCount+1  # number of different classes / sample types
        dfname="df"+str(fileCount) #create sequential names for dataframes as read in  eg
df1,df2,...
        dfnameSNV="df"+str(fileCount) +"SNV"   # dataframe names for Standard Normal
Variate eg, df1SNV,...
        print("file number",fileCount," dataframe name",dfname)


        # open file (show .csv files)
        dlg = QFileDialog()
        dlg.setNameFilter('(*.csv)')
        dlg.setFileMode(QFileDialog.AnyFile)
        fileNameList=[]



        minX_value=self.spinBox.value()   # low wavenumber limit
        maxX_value=self.spinBox_2.value() # high wavenumber limit



        if dlg.exec_():
```

```
        filenames = dlg.selectedFiles()

        fnam=filenames[0] # could put all data analysed files in a folder and open in one go

        data= pd.read_csv (filenames[0])

        df = pd.DataFrame(data) #create pandas dataframe called df (reading analysed.csv files - as produced by Raman processing program)



        # get file & path name  (might only work for Mac i.e need to change separator?) i.e. if windows use "\"

        sep="/"

        fnamLen=len(fnam)

        lastSep=fnam.rfind(sep)

        #pathName=fnam[:lastSep+1]  # identify path

        fileName=fnam[lastSep+1:fnamLen]   # identify file name

        fileNameList.append((fileName))


        #print(df.columns)

        # set spin box values to values that are appropriate for the data

        x=df["Wavenumber"]

        if x.max()<maxX_value:

           maxX_value = int(x.max())

           self.spinBox_2.setValue(maxX_value)

        if x.min()>minX_value:

           minX_value=int(x.min())

           self.spinBox.setValue(minX_value)


        # plot x,y summary data just to show representation of loaded data (nb could have also plotted smoothed data)

        xFullSpectrum=df["Wavenumber"] # in case you need to know original spectral length
```

```
        y=df["mean"]  # get the mean value column from the data file (could also get the smoothed mean)

        y_sd=df["SD1"]


        #y=df["mean_smoothed"]  # get the mean value column from the data file (could also get the smoothed mean)

        #y_sd=df["SD_smoothed"]

        self.dataplot(x,y,y_sd,fileCount) # plot average spectrum from data analysed


        #put file names in message window - to show which files have been loaded

        pcol=color_by_name[fileCount-1]    #colours for plotting

        MessString=fileName+" ("+ pcol+")"+"\n"

        self.textEdit.insertPlainText(MessString)


        # drop unused columns

        df.drop('mean', axis=1, inplace=True)

        df.drop('SD1', axis=1, inplace=True)

        #df.drop('mean_smoothed', axis=1, inplace=True)

        #df.drop('SD_smoothed', axis=1, inplace=True)

        #print(df.columns)




        # create SNV

        col_name=df.columns

        col_nameSNV=["Wavenumber"]


        dataSNV=[]  #new dataframe

        #calc standard normal variance
```

```
        # create numpy array of data and column name header then create new SNV
dataframe
        # this is a longwinded way and could be done using StandardScalar
        # feat_SNV = StandardScaler().fit_transform(feat)



        for col in col_name:
           if col=="Wavenumber":
              x=df[col]
           elif col=="mean":
              pass
           elif col=="SD1":
              pass
           elif col=="mean_smoothed":
              pass
           elif col=="SD_smoothed":
              pass
           else:
              newcolName=col+"_SNV"


              col_nameSNV.append(newcolName)


              #SNV is (data-average)/stdev  (column wise)
              dataSNV=(df[col]-df[col].mean())/df[col].std()
              if col=="Data1":
                 dataArray_SNV=np.column_stack((x,dataSNV))
              else:
                 dataArray_SNV=np.column_stack((dataArray_SNV,dataSNV))


        #create new data frame - without average and with SNV normalised data
```

```
        dfSNV = pd.DataFrame(dataArray_SNV)
        dfSNV.columns=col_nameSNV



        #calc mean of SNV data as a check
        ncols=len(dfSNV.columns)
        meanSNV =dfSNV.iloc[:, 1:ncols].mean(axis=1)
        stdSNV= dfSNV.iloc[:, 1:ncols].std(axis=1)

        # create dataframe for average/ SDev
        """
        dfSNV_Average=pd.DataFrame()
        dfSNV_Average["Wavenumber"]=x
        dfSNV_Average["mean"]=meanSNV
        dfSNV_Average["std"]=stdSNV

        #self.dataplot(x,meanSNV,fileCount) #plots the SNV spectrum
        """

        #add name data frame to dictionary of dataframes,  name=dfnameSNV
        classDict[dfnameSNV]=dfSNV.copy()
        #print("Dictionary=\n",classDict)



def scattplot2(self,xi,yi,zi,colcount,x_lbl,y_lbl,z_lbl,u):
    # PCA Scores

    pcol=color_by_name[colcount]
    if colcount==0:
       self.canvas.axes3.cla()
```

```
            self.canvas.axes3.set_xlabel(x_lbl,fontsize=12)
            self.canvas.axes3.set_ylabel(y_lbl,fontsize=12)
            self.canvas.axes3.set_zlabel(z_lbl,fontsize=12)
            self.canvas.axes3.set_title("PCA Scores",fontsize=14)

        self.canvas.axes3.scatter( xi,yi,zi,color=pcol, s=100,edgecolors='k',  alpha=transp,
label=str(u))


        if self.checkBox.isChecked()==True:
            self.canvas.axes3.legend()

        if self.checkBox_2.isChecked()==True:
            self.canvas.axes3.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
        self.canvas.draw ()



    def scattplot(self,xi,yi,colcount,x_lbl,y_lbl,u):
        # PCA Scores

        pcol=color_by_name[colcount]
        if colcount==0:
            self.canvas.axes4.cla()
            self.canvas.axes4.set_xlabel(x_lbl,fontsize=12)
            self.canvas.axes4.set_ylabel(y_lbl,fontsize=12)
            self.canvas.axes4.set_title("PCA Scores",fontsize=14)

        self.canvas.axes4.scatter( xi,yi,color=pcol, s=pointSize,edgecolors='k',  alpha=transp,
label=str(u))
```

```
    if self.checkBox.isChecked()==True:
        self.canvas.axes4.legend()



    if self.checkBox_2.isChecked()==True:
        self.canvas.axes4.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
    self.canvas.draw ()



def contourplot(self,xi,yi,colcount,x_lbl,y_lbl):
    # LDA Scores

    if self.checkBox_4.isChecked()==False and colcount==0:
        self.canvas.axes4.cla()
        self.canvas.axes4.set_xlabel(x_lbl,fontsize=12)
        self.canvas.axes4.set_ylabel(y_lbl,fontsize=12)
        # if not showing points clear figure
    xmin=min(xi)
    xmax=max(xi)
    ymin=min(yi)
    ymax=max(yi)



    xx, yy = np.mgrid[xmin:xmax:100j, ymin:ymax:100j]
    positions = np.vstack([xx.ravel(), yy.ravel()])
    values = np.vstack([xi, yi])
    kernel = st.gaussian_kde(values)
    f = np.reshape(kernel(positions).T, xx.shape)
```

```
    #pcol=color_by_name[colcount]
    cmap=c_map_by_name[colcount]
    """f colcount==0:
        self.canvas.axes2.cla()
        self.canvas.axes2.set_xlabel(x_lbl,fontsize=12)
        self.canvas.axes2.set_ylabel(y_lbl,fontsize=12)
        #self.canvas.axes2.set_title("LDA Scores",fontsize=14)
    """
    #step = 0.1

    m = np.amax(f)
    step = m/10
    print("m=",m)
    #levels = np.arange(0.0, m, step) + step
    levels = np.arange(step, m, step) + step
    self.canvas.axes4.contourf(xx, yy, f, levels, cmap=cmap, alpha=transp)
    if self.checkBox.isChecked()==True:
        self.canvas.axes4.legend()
    self.canvas.draw ()


def histplot(self,x,freq,colcount):
    pcol=color_by_name[colcount]

    if colcount==0:
        self.canvas.axes4.cla()
        self.canvas.axes4.set_xlabel("PCA Scores",fontsize=12)
        self.canvas.axes4.set_ylabel("Number of Cells",fontsize=12)
        #self.canvas.axes4.set_title("LD Score",fontsize=14)
```

```
            self.canvas.axes4.bar(x,freq,width=barWidth,bottom=None,
align='edge',color=pcol,edgecolor='w',alpha=transp+0.1, label=str(colcount+1))
        else:
            #self.canvas.axes4.set_xlabel("PCA Scores",fontsize=12)
            self.canvas.axes4.bar(x,freq,width=barWidth,bottom=None,
align='edge',color=pcol,edgecolor='w',alpha=transp+0.1, label=str(colcount+1))




        if self.checkBox_2.isChecked()==True:
            self.canvas.axes4.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
        if self.checkBox.isChecked()==True:
             self.canvas.axes4.legend()
        self.canvas.draw ()




    def dataplot2(self,x,y,colcount):
        # plot varience and variance explained
        self.canvas.axes2.set_xlabel("Principal Component",fontsize=12)
        self.canvas.axes2.set_ylabel("Variance",fontsize=12)
        self.canvas.axes2.set_title("Variance Explained/ Cummulative Variance",fontsize=14)
        if self.checkBox.isChecked()==True:
            self.canvas.axes2.legend()
```

```
        if colcount==1:
            self.canvas.axes2.cla()
            self.canvas.axes2.plot( x,y,marker='o',markerfacecolor="limegreen",
markersize=8,color='k',  linestyle='-',alpha=1,label='Var. Expl.')
        else:
            self.canvas.axes2.plot( x,y,marker='o',markerfacecolor="gold",
markersize=8,color='k',  linestyle='-',alpha=1,label='Cum. Var.')



        self.canvas.axes2.set_ylim([-0.05, 1.05])


        if self.checkBox_2.isChecked()==True:
            self.canvas.axes2.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
        if self.checkBox.isChecked()==True:
            self.canvas.axes2.legend()


        self.canvas.draw ()



    def dataplot3(self,x,y,colcount,numPCA_to_plot):
        #PC Loadings
        nme="PC"+str(colcount+1)


        if colcount==0:
            self.canvas.axes4.cla()
            self.canvas.axes4.set_xlabel("Wavenumber (cm-1)",fontsize=12)
            self.canvas.axes4.set_ylabel("PC",fontsize=12)
            self.canvas.axes4.set_title("PC Loadings",fontsize=14)


        offset_ratio=1.0
```

```
    #y = list(np.asarray(y) + col/5)
    y = list(np.asarray(y) + offset_ratio+colcount)
    colors=cm.Blues(np.linspace(0.5, 1, numPCA_to_plot))
    pcol=colors[colcount]

    #self.canvas.axes4.plot( x,y,linewidth=1.5,linestyle='-',alpha=1.0)
    self.canvas.axes4.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1.0)

    if self.checkBox_2.isChecked()==True:
       self.canvas.axes4.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
    #self.canvas.axes4.legend()
    self.canvas.draw ()



def dataplot(self,x,y,dy,fileCount):
    #Average Data

    if fileCount==1:
       self.canvas.axes.cla()
       self.canvas.axes.set_xlabel("Wavenumber (cm-1)",fontsize=12)
       self.canvas.axes.set_ylabel("Intensity (arb. units)",fontsize=12)
       self.canvas.axes.set_title("Raman Spectra",fontsize=14)
    offset_ratio=1.5
    y=y+(fileCount*offset_ratio)    # offset spectra
    pcol=color_by_name[fileCount-1]
    self.canvas.axes.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1.0)
    self.canvas.axes.fill_between(x,y+dy,y-dy,facecolor=pcol,linewidth=1.5,alpha=0.4)

    if self.checkBox_2.isChecked()==True:
```

```
            self.canvas.axes.grid(color = 'gray', linestyle = '--', linewidth = 0.5)


        #self.canvas.axes.legend()
        self.canvas.draw ()




if __name__=="__main__":
    app=QtWidgets.QApplication(sys.argv)
    w=MainWindow()
    w.show()
    app.exec_()
```

RamanLDA_6.py

"""

Created on 15/4/2023 - PCA Analysis of Raman Spectra

Files input are analysed.csv files

1) read in multiple CSV files - one at a time (though could be changed to read all at once quite easily) these should be files that are "analyzed"

2) display data - the 'mean' data column for each data class displayed

3) truncate data so that all spectra cover the same wavenumber regime and cover the regions of interest

4) perform SNV

5) PCA anaylsis using sk learn

6) display outputs

7) save in way suitable for publication - can produce tiff or other format output directly clicking on navigation bar

at the bottom of the plots - this also allows change colours. lines, etc

Use PyQT & pyqtgrpah as framework for program & displays - could be altered for Matplot lib which would give better histograms

@author: stephenevans

"""

```
import numpy as np

import sys

import pandas as pd

#import statistics
```

```
#import seaborn as sns


from PyQt5 import uic

from PyQt5 import QtWidgets

from PyQt5.QtWidgets import ( QFileDialog,QMainWindow)


import scipy.stats as st


import matplotlib

import matplotlib.pyplot as plt


matplotlib.rcParams['savefig.dpi'] = 600  # sets resolution for saving file (nb tiff can be very large - but high quality at this setting)

matplotlib.use('Qt5Agg')


from matplotlib.figure import Figure

from matplotlib.backends.backend_qt5agg import (NavigationToolbar2QT as NavigationToolbar)

from matplotlib.backends.backend_qt5agg import FigureCanvasQTAgg as FigureCanvas

from matplotlib import cm


from sklearn.decomposition import PCA

from sklearn.discriminant_analysis import LinearDiscriminantAnalysis

#from sklearn.preprocessing import StandardScaler

#from sklearn import svm

#from sklearn.cluster import KMeans

from sklearn.model_selection import train_test_split, cross_val_score

from sklearn.metrics import confusion_matrix,ConfusionMatrixDisplay, classification_report

#from sklearn.model_selection import RepeatedStratifiedKFold

#from collections import Counter
```

```
from sklearn.model_selection import StratifiedKFold
#from sklearn.datasets import make_classification



class MplCanvas(FigureCanvas):
    def __init__(self,parent=None,width=5,height=4,dpi=300):

        fig = Figure(figsize=(width, height), dpi=dpi)
        self.axes = fig.add_subplot(221)

        self.axes2 = fig.add_subplot(222)
        self.axes3 = fig.add_subplot(223,projection='3d')    # for 3D plot
        #self.axes3 = fig.add_subplot(223)     # for 2D plot
        self.axes4 = fig.add_subplot(224)

        fig.tight_layout(pad=5)
        #print ('in canvas')
        super().__init__(fig)
        #super(MplCanvas(),self).__init__(fig)

class MainWindow(QMainWindow):

    def __init__(self):
        super().__init__()
        QMainWindow.__init__(self)
        self.ui=uic.loadUi('Raman_LDA_GUI.ui',self)
        self.show()
```

```
        global fileCount,MessString,df2,dframe,classDict, cols,
cols2,color_by_name,c_map_by_name
        #plot parameters
        global pointSize,transp,barWidth
        pointSize=100
        transp=0.4
        barWidth=0.25

        color_by_name=['lightcoral','navajowhite','palegreen','lightskyblue','mediumpurple',
            'limegreen','sienna','darkred']# can add more colours
        c_map_by_name=["Reds","Greens", "Purples","Blues","Greys","RdPu"]
        classDict={} # used to store data and labels like deepskyblue and mediumblue and
crimson

        fileCount=0 # number of classes / different tyoes of cell / material
        MessString="Report:"

        self.setWindowTitle("LDA Analysis")
        self.pushButton_3.clicked.connect(lambda:self.getfiles())
        self.pushButton_5.clicked.connect(lambda:self.LDA_fit())
        self.pushButton_7.clicked.connect(lambda:self.PCA_LDA_fit())
        self.pushButton_9.clicked.connect(lambda:self.save())
        self.pushButton_8.clicked.connect(lambda:self.truncate())

        #limits for truncate
        self.spinBox.valueChanged.connect(lambda:self.valueChange)
        self.spinBox_2.valueChanged.connect(self.valueChange)
        # number of ocmponents to use
        self.spinBox_8.valueChanged.connect(self.num_comp_PCA)
        self.spinBox_3.valueChanged.connect(self.num_comp_LDA)
```

```
        # changed axis so PC1 is on x-axis
        #self.spinBox_5.setValue(2)
        #self.spinBox_6.setValue(1)
        self.spinBox_4.valueChanged.connect(self.plotParams)
        self.doubleSpinBox.valueChanged.connect(self.plotParams)
        self.doubleSpinBox_2.valueChanged.connect(self.plotParams)

        # Message box
        MessString="Files Loaded (and colour)\n"
        self.textEdit.insertPlainText(MessString)

        # plot set-up
        self.canvas=MplCanvas(self, width=8, height=9, dpi=100)

        self.ui.verticalLayout_2.addWidget(self.canvas)
        self.navi_toolbar = NavigationToolbar(self.canvas, self) #createa navigation toolbar for
our plot canvas
        self.verticalLayout_2.addWidget(self.navi_toolbar)

        # it is functional but certain elements not yet implemented - eg dark mode, toggling
grids, saving outputs



    # need to add button to implement clear all
    def clearAll(self):
        self.canvas.axes.cla()
```

```
        self.canvas.axes2.cla()
        self.canvas.axes3.cla()
        self.canvas.axes4.cla()
        self.canvas.draw ()
        self.pushButton_8.setEnabled(True)

        # need to clear all relevant data - at present best to simply re-run program

    def truncate(self):
        global XvaluesTrunc
        xlowerLimit=self.spinBox.value()
        xupperLimit=self.spinBox_2.value()
        print("truncate between limits",xlowerLimit,xupperLimit)

        MessString="Truncated between "+str(xlowerLimit)+" and "+str(xupperLimit)+"\n"
        self.textEdit.insertPlainText(MessString)



        #data prep truncate for all data have same stop and start
        #print("Dictionary=\n",classDict.keys())
        df_temp=pd.DataFrame()

        for ky in classDict:
            df_temp = classDict[ky]
            alldataArray=[]

            col_name=df_temp.columns
            wnbr="Wavenumber"
```

```
    #find indices of upper and lower limits
    imin=(df_temp[df_temp[wnbr]==xlowerLimit].index.values)
    imax=(df_temp[df_temp[wnbr]==xupperLimit].index.values)

    imin=np.take(imin,0)
    imax=np.take(imax,0)
    print("")
    #print("data set:",ky,imin,imax)

    # create series / array of truncated x values
    XvaluesTrunc=xFullSpectrum.truncate(imax,imin-1,copy=False);
    XvaluesTrunc.reset_index(drop=True, inplace=True)
    newData=df_temp.iloc[imax:imin,:]
    newData=newData.drop(['Wavenumber'], axis=1)
    # could use iloc[imax:imin,1:] instead of dropping Wavenumber

    newData.reset_index(drop=True, inplace=True)
    #print("new Data.head",newData.head)

    #replace the old dataframe with new truncated dataframe
    classDict[ky]=newData

#plot new truncated dataframe - plots mean of the SNV normalised data for each class
counter=0
for ky in classDict:
    counter=counter+1
    df_temp_new = classDict[ky]
    ncols=len(df_temp_new)
    x= XvaluesTrunc
    y = df_temp_new.iloc[:, 1:ncols].mean(axis=1)# calc mean for plotting
```

```
            dy=df_temp_new.iloc[:, 1:ncols].std(axis=1)
            self.dataplot(x,y,dy,counter)
        #self.pushButton_8.setEnabled(False)   # truncate only allowed once
       # could add reset button and allow re-truncate?

    def valueChange(self):
       # values for truncate
        global xlowerLimit,xupperLimit
        s=self.spinBox.value()
        sb1_val=s
        xlowerLimit=sb1_val
        sb2_val=self.spinBox_2.value()
        xupperLimit=sb2_val
        #print("truncate region limits",xlowerLimit,xupperLimit)

    def num_comp_PCA(self):
            global numPCA_components
            numPCA_components=self.spinBox_8.value()
    def num_comp_LDA(self):
            global numLDA_components
            numLDA_components=self.spinBox_3.value()

    def plotParams(self):
            global pointSize,transp,barWidth
            barWidth=self.doubleSpinBox.value()
            pointSize=self.spinBox_4.value()
            transp=self.doubleSpinBox_2.value()
            print("point=",pointSize," : Bar width=",barWidth," aplha=",transp)

#2345
```

```
def LDA_fit(self):
    print("In LDA")
    global XvaluesTrunc,alldataArray,lab,unique
    numLDA_components=self.spinBox_3.value()
    print("number of LDA components=",numLDA_components)

    alldataArray=[]
    count=0;counter=0

    for ky in classDict:
        count=count+1
        df_temp_new = classDict[ky]
        col_name=df_temp_new.columns

        ncols=len(col_name)

        lbl=count*np.ones((ncols,1),dtype=int) #add labels to array to allow classification
        temp_array=df_temp_new.to_numpy() # convert from df to np array
        temp_array=temp_array.T   #transpose array so spectra are in rows
        #add labels (nb (shape= (num spectra, no spectral point + 1 point for lable)))
        temp_array=np.hstack((lbl, temp_array))
        #print("Label / class",ky," shape",temp_array.shape)
        # add all data in one large array for PCA & LDA
        if count==1:
            alldataArray=temp_array
        else:
            alldataArray=np.concatenate((alldataArray, temp_array))

        lab = alldataArray[:,0].astype('uint8') # get labels
    # find unique labels (classes of data)
```

```
        unique=list(set(lab))
        y=lab



        # perform LDA with top n_components
        X=alldataArray[:,1:]
        sk_lda=LinearDiscriminantAnalysis(n_components= numLDA_components)
        X_r2=sk_lda.fit_transform(X,y)


        scores = cross_val_score(sk_lda, X, y, cv=4)
        print("Accuracy LDA only but using cross validation: %0.4f (+/- %0.4f)" %
(scores.mean(), scores.std() * 2))
        print("")


        Message="LDA Accuracy (using CV=4) = "+str(round(scores.mean(),3))+"(+/-
"+str(round(scores.std() * 2,3))+")\n"
        self.textEdit.insertPlainText(Message)


        lda_varExp=sk_lda.explained_variance_ratio_
        print("LDA variance explained=",lda_varExp,"\n")



        # plot variance explained (from LDA)
        xpos=range(1,numLDA_components+1,1)
        ypos=lda_varExp

        counter=1# set colour
        self.dataplot2(xpos,lda_varExp,counter) # plot
        ypos=np.cumsum(lda_varExp)
```

```
counter=2# set colour
self.dataplot2(xpos,ypos,counter) # plot



xi=[];yi=[];zi=[]
y_lbl="LD"+str(2)
x_lbl="LD"+str(1)
z_lbl="LD"+str(3)



lda_z=(self.spinBox_7.value())-1
lda_y=(self.spinBox_6.value())-1
lda_x=(self.spinBox_5.value() )-1
print("lda x,y",lda_x,lda_y)






for i,u in enumerate(unique):

   xi = [X_r2[j,lda_x] for j in range(len(X_r2[:,0])) if lab[j] == u]



   if fileCount>2:
      print("filecount",fileCount)
      yi = [X_r2[j,lda_y] for j in range(len(X_r2[:,1])) if lab[j] == u]
   if fileCount>3:
      zi = [X_r2[j,lda_z] for j in range(len(X_r2[:,2])) if lab[j] == u]
```

```
fig4_plot=self.comboBox.currentIndex()
if fig4_plot==0:
    pass
elif fig4_plot==1:
    freq1,x1 = np.histogram(xi)
    x11=x1[:-1]
    self.histplot(x11,freq1,i)
elif fig4_plot==2:
    freq1,x1 = np.histogram(yi)
    x11=x1[:-1]
    self.histplot(x11,freq1,i)
else:
    print("whatever")

x_lbl="LDA"+str(lda_x+1)
y_lbl="LDA"+str(lda_y+1)
z_lbl="LDA"+str(lda_z+1)


# plot scatter plots 2D and 3D

if fileCount>2:
    self.scattplot(xi,yi,i,x_lbl,y_lbl,u)
if fileCount>3:
    self.scattplot2(xi,yi,zi,i,x_lbl,y_lbl,z_lbl,u)

xi=[];yi=[];zi=[]
```

```
#2345
    def PCA_LDA_fit(self):
        global XvaluesTrunc,alldataArray
        numPCA_components=self.spinBox_8.value()
        numLDA_components=self.spinBox_3.value()



        print("In PCA_LDA")
        print("number of PCA components=",numPCA_components)
        print("number of LDA components=",numLDA_components)

        alldataArray=[]
        count=0;counter=0

        for ky in classDict:
            count=count+1
            df_temp_new = classDict[ky]
            col_name=df_temp_new.columns
            ncols=len(col_name)
            lbl=count*np.ones((ncols,1),dtype=int) #add labels to array to allow classification
            temp_array=df_temp_new.to_numpy() # convert from df to np array
            temp_array=temp_array.T   #transpose array so spectra are in rows
            temp_array=np.hstack((lbl, temp_array))
            if count==1:
                alldataArray=temp_array
            else:
                alldataArray=np.concatenate((alldataArray, temp_array))
                lab = alldataArray[:,0].astype('uint8') # get labels
```

```
        unique=list(set(lab))  # find unique labels (classes of data)
        y=lab

        #combined PCA / LDA
        X=alldataArray[:,1:]
        # use test train as alternative to k-fold cross validation - ueseful for prediction / generation of conusion matrix...
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=2)
        sk_lda=LinearDiscriminantAnalysis(n_components= numLDA_components)

        #cv= RepeatedStratifiedKFold(n_splits=10, n_repeats=3, random_state=1)
        cv=4

        #combined PCA / LDA
        sk_pca = PCA(n_components=numPCA_components) # or fixed value eg 15 components
        Xpc = sk_pca.fit_transform(X)

        scores = cross_val_score(sk_lda, Xpc, y, cv=cv)
        print("Accuracy PCA_LDA (cv=4): %0.4f (+/- %0.4f)" % (scores.mean(), scores.std() * 2))

        Message="PCA_LDA Accuracy (cv=4) = "+str(round(scores.mean(),3))+"(+/- "+str(round(scores.std() * 2,3))+")\n\n"
        self.textEdit.insertPlainText(Message)

        # for plotting
        lda_z=(self.spinBox_7.value())-1
        lda_y=(self.spinBox_6.value())-1
        lda_x=(self.spinBox_5.value() )-1
        #print("lda x,y",lda_x,lda_y)
```

```
        """
        X=sk_lda.fit_transform(Xpc,y)
        scores2 = cross_val_score(sk_lda, X, y, cv=cv)
        print("Accuracy PCA_LDA (cv=4): %0.4f (+/- %0.4f)" % (scores2.mean(), scores2.std() *
2))
        Message="PCA_LDA Accuracy (cv=4) = "+str(round(scores2.mean(),3))+"(+/-
"+str(round(scores2.std() * 2,3))+")\n"
        self.textEdit.insertPlainText(Message)
        """

        #X, y = make_classification(random_state=0)
        X=sk_lda.fit_transform(Xpc,y)
        lda_pca_varExp=sk_lda.explained_variance_ratio_




        print("LDA-PCA varience explained=",np.around(lda_pca_varExp),4)
        print("")

        # use test train split to calc sklearn to calc confusion matrix and prediction score -
based on single test /train set
        """
        sk_lda.fit(X_train,y_train)
        y_pred=sk_lda.predict(X_test)
        print("prediction: ",sk_lda.score(X_test,y_test))
        #confusion matrix
        cm = confusion_matrix(y_test, y_pred)
        print("Confusion Matrix: ")
```

```
print(cm)
print(classification_report(y_test,y_pred,digits=4 ))
"""


# as above but averaged over multple test train splits
conf_matrix_list_of_arrays = []

skf = StratifiedKFold(n_splits=4)

#normalised version - set to 'none to get raw data
for train_index, test_index in skf.split(X,y):

    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    #print("train/test",train_index, test_index)
    sk_lda.fit(X_train, y_train)
    y_pred=sk_lda.predict(X_test)
    conf_matrix = confusion_matrix(y_test, y_pred,labels=sk_lda.classes_)

    nconf_matrix = confusion_matrix(y_test,
y_pred,labels=sk_lda.classes_,normalize='true')
    #print(conf_matrix)



    conf_matrix_list_of_arrays.append(nconf_matrix)
mean_of_conf_matrix_arrays = np.mean(conf_matrix_list_of_arrays, axis=0)
ncm=mean_of_conf_matrix_arrays
ncm=np.around(ncm, decimals=3, out=None)
```

```
print("Confusion Matrix: Averaged & Normalised ")
print(ncm)
#plot confusion matrix - matplot
# disp = ConfusionMatrixDisplay(confusion_matrix=ncm,display_labels=sk_lda.classes_)
#disp.plot()



# calc preciscion, recall etc
true_pos = np.diag(ncm)
false_pos = np.sum(ncm, axis=0) - true_pos
false_neg = np.sum(ncm, axis=1) - true_pos
true_neg=np.sum(ncm)-true_pos-false_pos-false_neg


# Precision: The percentage of positive predictions made by the model that were correct. It is calculated as TP / (TP + FP).
# Recall: The percentage of actual positive cases that were correctly predicted by the model. It is calculated as TP / (TP + FN).
precision = np.around((true_pos / (true_pos + false_pos)),4)
Ave_precision=round(np.mean(precision),4)
recall =np.around((true_pos / (true_pos + false_neg)),4)
Ave_recall=round(np.mean(recall),4)
print("Precison  ",precision, "   Avereage Precision= ",Ave_precision)
print("Recall ",recall,"   Avereage Recall= ",Ave_recall)
#F1-score: The harmonic mean of precision and recall. It is a balanced measure that considers both precision and recall.
#It is calculated as 2 * (precision * recall) / (precision + recall).


F1_score=np.around(2 * (precision * recall) / (precision + recall),4)
Ave_F1=round(np.mean(F1_score),4)
print("F1-score ",F1_score, " Avearage F1-score= ",Ave_F1)
```

```
#Accuracy: The percentage of correct predictions made by the model. It is calculated as (TP + TN) / (TP + TN + FP + FN).

Acc=np.around((true_pos+true_neg)/(true_pos+true_neg+false_pos+false_neg),4)

Ave_Acc=round(np.mean(Acc),4)

print("Accuracy ",Acc, "  Average Accuracy",Ave_Acc)


#Message="Confusion Matrix: Averaged & Normalised \n"+ str(ncm)+"\n"

#self.textEdit.insertPlainText(Message)

Message="Precision \n"+ str(recall)+"\n"+"Average precision "+ str(Ave_precision)+"\n"

self.textEdit.insertPlainText(Message)



Message="Recall \n"+ str(recall)+"\n"+"Average recall "+ str(Ave_recall)+"\n"

self.textEdit.insertPlainText(Message)



Message="F1-Score \n"+ str(F1_score)+"\n"+"Average F1-score "+ str(Ave_F1)+"\n"

self.textEdit.insertPlainText(Message)

Message="Accuracyl \n"+ str(Acc)+"\n"+"Average Accuracy "+ str(Ave_Acc)+"\n"

self.textEdit.insertPlainText(Message)

xpos=range(1,numLDA_components+1,1)

ypos=lda_pca_varExp

counter=1# set colour

self.dataplot2(xpos,ypos,counter)

ypos=np.cumsum(lda_pca_varExp)

counter=2# set colour
```

```
self.dataplot2(xpos,ypos,counter)



xi=[];yi=[];zi=[]
for i,u in enumerate(unique):

   xi = [X[j,lda_x] for j in range(len(X[:,0])) if lab[j] == u]

   if fileCount>2:
      #print("filecount",fileCount)
      yi = [X[j,lda_y] for j in range(len(X[:,1])) if lab[j] == u]

   if fileCount>3:
      zi = [X[j,lda_z] for j in range(len(X[:,2])) if lab[j] == u]



   fig4_plot=self.comboBox.currentIndex()
   if fig4_plot==0:
      pass
   elif fig4_plot==1:
      bw=barWidth
      bins=int((max(xi) - min(xi))/bw)
      #print("Bins=",bins,bw)
      freq1,x1 = np.histogram(xi,bins=bins)
      x11=x1[:-1]
      #print("bins",x11)
      self.histplot(x11,freq1,i)
   elif fig4_plot==2:
      bw=barWidth
```

```
    bins=int((max(yi) - min(yi))/bw)
    freq1,x1 = np.histogram(yi,bins=bins)
    x11=x1[:-1]
    self.histplot(x11,freq1,i)
elif fig4_plot==3:
    # confusion matrix plot
    self.cmplot(ncm)
else:
    print("whatever")



x_lbl="PCA_LDA"+str(lda_x+1)
y_lbl="PCA_LDA"+str(lda_y+1)
z_lbl="PCA_LDA"+str(lda_z+1)

# plot scatter plots 2D and 3D
if fileCount>2:

    if self.checkBox_5.isChecked()==True:
        self.scattplot(xi,yi,i,x_lbl,y_lbl,u)
    if self.checkBox_3.isChecked()==True:
        self.contourplot(xi,yi,i,x_lbl,y_lbl)




if fileCount>3:
    self.scattplot2(xi,yi,zi,i,x_lbl,y_lbl,z_lbl,u)
```

```
        xi=[];yi=[];zi=[]
```

```
    def getfiles(self):
        global x,y,dataArray,fileCount,MessString,df2,classDict,dfSNV,xFullSpectrum


        fileCount=fileCount+1  # number of different classes / sample types
        dfname="df"+str(fileCount) #create sequential names for dataframes as read in  eg
df1,df2,...
```

```
    dfnameSNV="df"+str(fileCount) +"SNV"   # dataframe names for Standard Normal Variate eg, df1SNV,...
    #print("file number",fileCount," dataframe name",dfname)

    # open file (show .csv files)
    dlg = QFileDialog()
    dlg.setNameFilter('(*.csv)')
    dlg.setFileMode(QFileDialog.AnyFile)
    fileNameList=[]

    minX_value=self.spinBox.value()   # low wavenumber limit
    maxX_value=self.spinBox_2.value() # high wavenumber limit

    if dlg.exec_():

        filenames = dlg.selectedFiles()
        fnam=filenames[0] # could put all data analysed files in a folder and open in one go
        data= pd.read_csv (filenames[0])
        df = pd.DataFrame(data) #create pandas dataframe called df (reading analysed.csv files - as produced by Raman processing program)

        # get file & path name  (might only work for Mac i.e need to change separator?) i.e. if windows use "\"
        sep="/"
        fnamLen=len(fnam)
        lastSep=fnam.rfind(sep)
        #pathName=fnam[:lastSep+1]  # identify path
        fileName=fnam[lastSep+1:fnamLen]   # identify file name
```

```
        fileNameList.append((fileName))


        #print(df.columns)
        # set spin box values to values that are appropriate for the data
        x=df["Wavenumber"]
        if x.max()<maxX_value:
            maxX_value = int(x.max())
            self.spinBox_2.setValue(maxX_value)
        if x.min()>minX_value:
            minX_value=int(x.min())
            self.spinBox.setValue(minX_value)


        # plot x,y summary data just to show representation of loaded data (nb could have
also plotted smoothed data)
        xFullSpectrum=df["Wavenumber"] # in case you need to know original spectral
length
        y=df["mean"]  # get the mean value column from the data file (could also get the
smoothed mean)
        y_sd=df["SD1"]


        #y=df["mean_smoothed"]  # get the mean value column from the data file (could
also get the smoothed mean)
        #y_sd=df["SD_smoothed"]
        self.dataplot(x,y,y_sd,fileCount) # plot average spectrum from data analysed


        #put file names in message window - to show which files have been loaded
        pcol=color_by_name[fileCount-1]    #colours for plotting
        MessString="C"+ str(fileCount)+": "+fileName+" ("+ pcol+")"+"\n"
        self.textEdit.insertPlainText(MessString)
```

```
        if fileCount>1:
            self.spinBox_3.setValue(fileCount-1)
            fcount=self.spinBox_3.value()
            print("filecount=",fileCount,fcount)
            self.spinBox_3.setMinimum(1)
            #self.spinBox_3.setMaximum(fileCount-1)

        #self.spinBox_3.setMinimum(1)

        # drop unused columns
        df.drop('mean', axis=1, inplace=True)
        df.drop('SD1', axis=1, inplace=True)
        #df.drop('mean_smoothed', axis=1, inplace=True)
        #df.drop('SD_smoothed', axis=1, inplace=True)
        #print(df.columns)



        # create SNV
        col_name=df.columns
        col_nameSNV=["Wavenumber"]

        dataSNV=[]  #new dataframe
        #calc standard normal variance

        # create numpy array of data and column name header then create new SNV
dataframe
        # this is a longwinded way and could be done using StandardScalar
```

```
# feat_SNV = StandardScaler().fit_transform(feat)



for col in col_name:
    if col=="Wavenumber":
        x=df[col]
    elif col=="mean":
        pass
    elif col=="SD1":
        pass
    elif col=="mean_smoothed":
        pass
    elif col=="SD_smoothed":
        pass
    else:
        newcolName=col+"_SNV"


        col_nameSNV.append(newcolName)


        #SNV is (data-average)/stdev  (column wise)
        dataSNV=(df[col]-df[col].mean())/df[col].std()
        if col=="Data1":
            dataArray_SNV=np.column_stack((x,dataSNV))
        else:
            dataArray_SNV=np.column_stack((dataArray_SNV,dataSNV))

#create new data frame - without average and with SNV normalised data
dfSNV = pd.DataFrame(dataArray_SNV)
dfSNV.columns=col_nameSNV
```

```
        #calc mean of SNV data as a check
        #ncols=len(dfSNV.columns)
        #meanSNV =dfSNV.iloc[:, 1:ncols].mean(axis=1)
        #stdSNV= dfSNV.iloc[:, 1:ncols].std(axis=1)



        #add name data frame to dictionary of dataframes,  name=dfnameSNV
        classDict[dfnameSNV]=dfSNV.copy()
        #print("Dictionary=\n",classDict)
        #set spin box to be highest number allowed



def cmplot(self,ncm):

      self.canvas.axes4.cla()
      self.canvas.axes4.set_xlabel('Predictions',fontsize=14)
      self.canvas.axes4.set_ylabel("Actuals",fontsize=14)
      self.canvas.axes4.matshow(ncm, cmap=matplotlib.cm.Blues, alpha=0.3)
      for i in range(ncm.shape[0]):
         for j in range(ncm.shape[1]):
            self.canvas.axes4.text(x=j, y=i,s=ncm[i, j], va='center',ha='center', size='large')


      #self.canvas.axes4.set_title('Confusion Matrix', fontsize=18)
```

```
        self.canvas.draw ()


    def histplot(self,x,freq,colcount):
            pcol=color_by_name[colcount]


            if colcount==0:
               self.canvas.axes4.cla()
               self.canvas.axes4.set_xlabel("LDA Score",fontsize=12)
               self.canvas.axes4.set_ylabel("Number of Cells",fontsize=12)
               #self.canvas.axes4.set_title("LD Score",fontsize=14)
               self.canvas.axes4.bar(x,freq,width=barWidth,bottom=None,
align='edge',color=pcol,edgecolor='w',alpha=transp+0.1, label=str(colcount+1))
            else:
               self.canvas.axes4.set_xlabel("LD",fontsize=12)
               self.canvas.axes4.bar(x,freq,width=barWidth,bottom=None,
align='edge',color=pcol,edgecolor='w',alpha=transp+0.1, label=str(colcount+1))




            if self.checkBox_2.isChecked()==True:
               self.canvas.axes4.grid(color = 'gray', linestyle = '--', linewidth = 0.5)


            if self.checkBox.isChecked()==True:
               self.canvas.axes4.legend()


            self.canvas.draw ()



    def scattplot2(self,xi,yi,zi,colcount,x_lbl,y_lbl,z_lbl,u):
        # PCA Scores
```

```
        pcol=color_by_name[colcount]
        if colcount==0:
            self.canvas.axes3.cla()
            self.canvas.axes3.set_xlabel(x_lbl,fontsize=12)
            self.canvas.axes3.set_ylabel(y_lbl,fontsize=12)
            self.canvas.axes3.set_zlabel(z_lbl,fontsize=12)
            #self.canvas.axes3.set_title("LDA Scores",fontsize=14)

        self.canvas.axes3.scatter( xi,yi,zi,color=pcol, s=pointSize,edgecolors='k',  alpha=transp,
label=str(u))

        """
        xmin=min(xi)
        xmax=max(xi)
        ymin=min(yi)
        ymax=max(yi)


        xx, yy = np.mgrid[xmin:xmax:100j, ymin:ymax:100j]
        positions = np.vstack([xx.ravel(), yy.ravel()])
        values = np.vstack([xi, yi])
        kernel = st.gaussian_kde(values)
        f = np.reshape(kernel(positions).T, xx.shape)
        cmap=c_map_by_name[colcount]
        m = np.amax(f)
        step = m/10
        print("m=",m)
        #levels = np.arange(0.0, m, step) + step
        levels = np.arange(step, m, step) + step
```

```
        #self.canvas.axes2.contourf(xx, yy, f, levels, cmap=cmap, alpha=0.5)
        self.canvas.axes3.contourf(xx,yy,f, levels,  zdir='z',  offset=-5, cmap=cmap, alpha=0.5)
        self.canvas.axes3.contourf(xx,yy,f, levels,  zdir='y',  offset=-7, cmap=cmap, alpha=0.5)
        self.canvas.axes3.contourf(xx,yy,f, levels,  zdir='x',  offset=-7, cmap=cmap, alpha=0.5)
        """


        if self.checkBox_2.isChecked()==True:
            self.canvas.axes3.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
        if self.checkBox.isChecked()==True:
            self.canvas.axes3.legend()
        self.canvas.draw ()



    def scattplot(self,xi,yi,colcount,x_lbl,y_lbl,u):
        # LDA Scores

        pcol=color_by_name[colcount]
        if colcount==0:
            self.canvas.axes2.cla()
            self.canvas.axes2.set_xlabel(x_lbl,fontsize=12)
            self.canvas.axes2.set_ylabel(y_lbl,fontsize=12)
            #self.canvas.axes2.set_title("LDA Scores",fontsize=14)

        self.canvas.axes2.scatter( xi,yi,color=pcol, s=pointSize,edgecolors='k',  alpha=transp,
label=str(u))



        if self.checkBox_2.isChecked()==True:
```

```
        self.canvas.axes2.grid(color = 'gray', linestyle = '--', linewidth = 0.5)
    if self.checkBox.isChecked()==True:
        self.canvas.axes2.legend()

    self.canvas.draw ()

def contourplot(self,xi,yi,colcount,x_lbl,y_lbl):
    # LDA Scores

    if self.checkBox_5.isChecked()==False and colcount==0:
        self.canvas.axes2.cla()
        self.canvas.axes2.set_xlabel(x_lbl,fontsize=12)
        self.canvas.axes2.set_ylabel(y_lbl,fontsize=12)
        # if not showing points clear figure
    xmin=min(xi)
    xmax=max(xi)
    ymin=min(yi)
    ymax=max(yi)


    xx, yy = np.mgrid[xmin:xmax:100j, ymin:ymax:100j]
    positions = np.vstack([xx.ravel(), yy.ravel()])
    values = np.vstack([xi, yi])
    kernel = st.gaussian_kde(values)
    f = np.reshape(kernel(positions).T, xx.shape)


    #pcol=color_by_name[colcount]
    cmap=c_map_by_name[colcount]
    """f colcount==0:
```

```
        self.canvas.axes2.cla()
        self.canvas.axes2.set_xlabel(x_lbl,fontsize=12)
        self.canvas.axes2.set_ylabel(y_lbl,fontsize=12)
        #self.canvas.axes2.set_title("LDA Scores",fontsize=14)
    """
    #step = 0.1

    m = np.amax(f)
    step = m/10
    print("m=",m)
    #levels = np.arange(0.0, m, step) + step
    levels = np.arange(step, m, step) + step
    self.canvas.axes2.contourf(xx, yy, f, levels, cmap=cmap, alpha=0.5)
    self.canvas.draw ()



  def dataplot2(self,x,y,colcount):
    # plot varience and variance explained
    self.canvas.axes4.set_xlabel("LD Component",fontsize=12)
    self.canvas.axes4.set_ylabel("Variance",fontsize=12)
    self.canvas.axes4.set_title("Variance Explained/ Cummulative Variance",fontsize=14)



    if colcount==1:
        self.canvas.axes4.cla()
        self.canvas.axes4.plot( x,y,marker='o',markerfacecolor="limegreen",
markersize=8,color='k',  linestyle='-',alpha=1,label='Var. Expl.')
    else:
```

```
        self.canvas.axes4.plot( x,y,marker='o',markerfacecolor="gold",
markersize=8,color='k',  linestyle='-',alpha=1,label='Cum. Var.')

    if self.checkBox_2.isChecked()==True:
        self.canvas.axes4.grid(color = 'gray', linestyle = '--', linewidth = 0.5)

    if self.checkBox.isChecked()==True:
        self.canvas.axes4.legend()

    self.canvas.axes4.set_ylim([-0.05, 1.05])
    self.canvas.draw ()



  def dataplot(self,x,y,dy,fileCount):
    #Average Data

    if fileCount==1:
        self.canvas.axes.cla()
        self.canvas.axes.set_xlabel("Wavenumber (cm-1)",fontsize=12)
        self.canvas.axes.set_ylabel("Intensity (arb. units)",fontsize=12)
        self.canvas.axes.set_title("Raman Spectra",fontsize=14)



        #matplotlib.style.use('_mpl-gallery')
        #fig, ax = matplotlib.subplots()

        #ax.fill_between(x, y-dy, y+dy, alpha=.5, linewidth=0)
        #ax.plot(x, y, linewidth=2)
```

```
        #ax.set(xlim=(0, 8), xticks=np.arange(1, 8),
         #      ylim=(0, 8), yticks=np.arange(1, 8))

        #matplotlib.show()




    offset_ratio=1.5
    y=y+(fileCount*offset_ratio)    # offset spectra
    pcol=color_by_name[fileCount-1]
    self.canvas.axes.plot( x,y,color=pcol,linewidth=1.5,linestyle='-',alpha=1.0)
    self.canvas.axes.fill_between(x,y+dy,y-dy,facecolor=pcol,linewidth=1.5,alpha=0.4)


    if self.checkBox_2.isChecked()==True:
        self.canvas.axes.grid(color = 'gray', linestyle = '--', linewidth = 0.5)



    self.canvas.draw ()



def bgndChange(self):
    global styles, ax_col,cols
```

```
if __name__=="__main__":
    app=QtWidgets.QApplication(sys.argv)
    w=MainWindow()
    w.show()
    app.exec_()
```

References

1. Martinez-Uribe, O., Becker, T.C. and Garman, K.S. Promises and Limitations of Current Models for Understanding Barrett's Esophagus and Esophageal Adenocarcinoma. *Cell Mol Gastroenterol Hepatol.* 2024, **17**(6), pp.1025-1038.
2. Grey, K. *Summary of the Three Major Types of Structural Filaments.* 2021.
3. Gala de Pablo, J. *Biochemical phenotyping of live single cells using confocal Raman spectroscopy*. PhD thesis, University of Leeds, 2019.
4. Menezes, M.A. and Herbella, F.A.M. Pathophysiology of Gastroesophageal Reflux Disease. 2017, **41**(7), p.1.
5. Aldridge, H. *AI FOR SINGLE CELL ANALYSIS*. Unpublished, 2023.
6. Armistead, F. *Mechanical phenotyping of single cells using shear and inertial microfluidic*. PhD thesis, University of Leeds, 2019.
7. Falk, G.W. and Katzka, D.A. 140 - Diseases of the Esophagus. In: Goldman, L. and Schafer, A.I. eds. *Goldman's Cecil Medicine (Twenty Fourth Edition)*. Philadelphia: W.B. Saunders, 2012, pp.874-886.
8. Gossett, D.R., Tse, H.T., Lee, S.A., Ying, Y., Lindgren, A.G., Yang, O.O., Rao, J., Clark, A.T. and Di Carlo, D.J.P.o.t.N.A.o.S. Hydrodynamic stretching of single cells for large population mechanical phenotyping. 2012, **109**(20), pp.7630-7635.
9. Fitzgerald, R.C. Barrett's oesophagus and oesophageal adenocarcinoma: how does acid interfere with cell proliferation and differentiation? 2005, **54**(suppl 1), pp.i21-i26.
10. Morgan, E., Soerjomataram, I., Rumgay, H., Coleman, H.G., Thrift, A.P., Vignat, J., Laversanne, M., Ferlay, J. and Arnold, M. The Global Landscape of Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma Incidence and Mortality in 2020 and Projections to 2040: New Estimates From GLOBOCAN 2020. *Gastroenterology.* 2022, **163**(3), pp.649-658.e642.
11. Xu, J., Yu, T., Zois, C.E., Cheng, J.-X., Tang, Y., Harris, A.L. and Huang, W.E. Unveiling Cancer Metabolism through Spontaneous and Coherent Raman Spectroscopy and Stable Isotope Probing. *Cancers.* 2021, **13**(7), p.1718.
12. Armistead, F.J., Gala De Pablo, J., Gadêlha, H., Peyman, S.A. and Evans, S.D. Cells Under Stress: An Inertial-Shear Microfluidic Determination of Cell Behavior. *Biophys J.* 2019, **116**(6), pp.1127-1135.
13. Quante, M., Graham, T.A. and Jansen, M. Insights Into the Pathophysiology of Esophageal Adenocarcinoma. *Gastroenterology.* 2018, **154**(2), pp.406-420.
14. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I. and Jemal, A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024, **74**(3), pp.229-263.
15. Kamangar, F., Nasrollahzadeh, D., Safiri, S., Sepanlou, S.G., Fitzmaurice, C., Ikuta, K.S., Bisignano, C., Islami, F., Roshandel, G., Lim, S.S., Abolhassani, H., Abu-Gharbieh, E., Adedoyin, R.A., Advani, S.M., Ahmed, M.B., Aichour, M.T.E., Akinyemiju, T., Akunna, C.J., Alahdab, F., Alipour, V., Almasi-Hashiani, A., Almulhim, A.M., Anber, N.H., Ansari-Moghaddam, A., Arabloo, J., Arab-Zozani, M., Awedew, A.F., Badawi, A., Berfield, K.S.S., Berhe, K., Bhattacharyya, K., Biondi, A., Bjørge, T., Borzì, A.M., Bosetti, C., Carreras, G., Carvalho, F., Castro, C., Chu, D.-T., Costa, V.M., Dagnew, B., Darega Gela, J., Daryani, A., Demeke, F.M., Demoz, G.T., Dianatinasab, M., Elbarazi, I., Emamian, M.H., Etemadi, A., Faris, P.S., Fernandes, E., Filip, I.,

Fischer, F., Gad, M.M., Gallus, S., Gebre, A.K., Gebrehiwot, T.T., Gebremeskel, G.G., Gebresillassie, B.M., Ghasemi-kebria, F., Ghashghaee, A., Ghith, N., Golechha, M., Gorini, G., Gupta, R., Hafezi-Nejad, N., Haj-Mirzaian, A., Harvey, J.D., Hashemian, M., Hassen, H.Y., Hay, S.I., Henok, A., Hoang, C.L., Hosgood, H.D., Househ, M., Ilesanmi, O.S., Ilic, M.D., Irvani, S.S.N., Jain, C., James, S.L., Jee, S.H., Jha, R.P., Joukar, F., Kabir, A., Kasaeian, A., Kassaw, M.W., Kaur, S., Kengne, A.P., Kerboua, E., Khader, Y.S., Khalilov, R., Khan, E.A., Khoja, A.T., Kocarnik, J.M., Komaki, H., Kumar, V., La Vecchia, C., Lasrado, S., Li, B., Lopez, A.D., Majeed, A., Manafi, N., Manda, A.L., Mansour-Ghanaei, F., Mathur, M.R., Mehta, V., Mehta, D., Mendoza, W., Mithra, P., Mohammad, K.A., Mohammadian-Hafshejani, A., Mohammadpourhodki, R., Mohammed, J.A., Mohebi, F., Mokdad, A.H., Monasta, L., Moosavi, D., Moosazadeh, M., Moradi, G., Moradpour, F., Moradzadeh, R., Naik, G., Negoi, I., Nggada, H.A., Nguyen, H.L.T., Nikbakhsh, R., Nixon, M.R., Olagunju, A.T., Olagunju, T.O., Padubidri, J.R., Pakshir, K., Patel, S., Pathak, M., Pham, H.Q., Pourshams, A., Rabiee, N., Rabiee, M., Radfar, A., Rafiei, A., Ramezanzadeh, K., Rath, G.K., Rathi, P., Rawaf, S., Rawaf, D.L., Rezaei, N., Roro, E.M., Saad, A.M., Salimzadeh, H., Samy, A.M., Sartorius, B., Sarveazad, A., Sekerija, M., Sha, F., Shamsizadeh, M., Sheikhbahaei, S., Shirkoohi, R., Siddappa Malleshappa, S.K., Singh, J.A., Sinha, D.N., Smarandache, C.-G., Soshnikov, S., Suleria, H.A.R., Tadesse, D.B., Tesfay, B.E., Thakur, B., Traini, E., Tran, K.B., Tran, B.X., Ullah, I., Vacante, M., Veisani, Y., Vujcic, I.S., Weldesamuel, G.T., Xu, R., Yazdi-Feyzabadi, V., Yuce, D., Zadnik, V., Zaidi, Z., Zhang, Z.-J., Malekzadeh, R. and Naghavi, M. The global, regional, and national burden of oesophageal cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology & Hepatology.* 2020, **5**(6), pp.582-597.

16. Liu, C.Q., Ma, Y.L., Qin, Q., Wang, P.H., Luo, Y., Xu, P.F. and Cui, Y. Epidemiology of esophageal cancer in 2020 and projections to 2030 and 2040. *Thorac Cancer.* 2023, **14**(1), pp.3-11.
17. Arnold, M., Laversanne, M., Brown, L.M., Devesa, S.S., Bray, F.J.O.j.o.t.A.C.o.G. and ACG. Predicting the future burden of esophageal cancer by histological subtype: international trends in incidence up to 2030. 2017, **112**(8), pp.1247-1255.
18. Santucci, C., Mignozzi, S., Malvezzi, M., Collatuzzo, G., Levi, F., La Vecchia, C. and Negri, E. Global trends in esophageal cancer mortality with predictions to 2025, and in incidence by histotype. *Cancer Epidemiology.* 2023, **87**, p.102486.
19. Sachdeva, K., Natarajan, K. and Iyer, P.G. Improving esophageal cancer screening across the globe: Translating knowledge into action. *Indian Journal of Gastroenterology.* 2024.
20. El-Serag, H.B., Sweet, S., Winchester, C.C. and Dent, J. Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut.* 2014, **63**(6), pp.871-880.
21. Nirwan, J.S., Hasan, S.S., Babar, Z.U., Conway, B.R. and Ghori, M.U. Global Prevalence and Risk Factors of Gastro-oesophageal Reflux Disease (GORD): Systematic Review with Meta-analysis. *Sci Rep.* 2020, **10**(1), p.5814.
22. Marley, J., Nicholl, B.I., Macdonald, S., Mair, F.S. and Jani, B.D. Associations between long-term conditions and upper gastrointestinal cancer incidence: A prospective population-based cohort of UK Biobank participants. *J Multimorb Comorb.* 2021, **11**, p.26335565211056136.

23. Solaymani–Dodaran, M., Card, T.R. and West, J. Cause-Specific Mortality of People With Barrett's Esophagus Compared With the General Population: A Population-Based Cohort Study. *Gastroenterology.* 2013, **144**(7), pp.1375-1383.e1371.
24. Edgren, G., Adami, H.O., Weiderpass, E. and Nyrén, O. A global assessment of the oesophageal adenocarcinoma epidemic. *Gut.* 2013, **62**(10), pp.1406-1414.
25. Vaughan, T.L. and Fitzgerald, R.C. Precision prevention of oesophageal adenocarcinoma. *Nat Rev Gastroenterol Hepatol.* 2015, **12**(4), pp.243-248.
26. Shipp, D., Sinjab, F. and Notingher, I. Raman spectroscopy: Techniques and applications in the life sciences. *Advances in Optics and Photonics.* 2017, **9**, p.315.
27. Lagergren, J., Bergström, R., Lindgren, A. and Nyrén, O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med.* 1999, **340**(11), pp.825-831.
28. Singh, S., Sharma, A.N., Murad, M.H., Buttar, N.S., El-Serag, H.B., Katzka, D.A. and Iyer, P.G. Central adiposity is associated with increased risk of esophageal inflammation, metaplasia, and adenocarcinoma: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol.* 2013, **11**(11), pp.1399-1412.e1397.
29. Cook, M.B., Kamangar, F., Whiteman, D.C., Freedman, N.D., Gammon, M.D., Bernstein, L., Brown, L.M., Risch, H.A., Ye, W., Sharp, L., Pandeya, N., Webb, P.M., Wu, A.H., Ward, M.H., Giffen, C., Casson, A.G., Abnet, C.C., Murray, L.J., Corley, D.A., Nyrén, O., Vaughan, T.L. and Chow, W.H. Cigarette smoking and adenocarcinomas of the esophagus and esophagogastric junction: a pooled analysis from the international BEACON consortium. *J Natl Cancer Inst.* 2010, **102**(17), pp.1344-1353.
30. Arnold, M., Soerjomataram, I., Ferlay, J. and Forman, D. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut.* 2015, **64**(3), pp.381-387.
31. Cook, M.B., Wild, C.P. and Forman, D. A systematic review and meta-analysis of the sex ratio for Barrett's esophagus, erosive reflux disease, and nonerosive reflux disease. *Am J Epidemiol.* 2005, **162**(11), pp.1050-1061.
32. Lagergren, J. and Lagergren, P. Recent developments in esophageal adenocarcinoma. *CA Cancer J Clin.* 2013, **63**(4), pp.232-248.
33. Tramacere, I., Pelucchi, C., Bagnardi, V., Rota, M., Scotti, L., Islami, F., Corrao, G., Boffetta, P., La Vecchia, C. and Negri, E. A meta-analysis on alcohol drinking and esophageal and gastric cardia adenocarcinoma risk. *Ann Oncol.* 2012, **23**(2), pp.287-297.
34. Rokkas, T., Pistiolas, D., Sechopoulos, P., Robotis, I. and Margantinis, G. Relationship between Helicobacter pylori infection and esophageal neoplasia: a meta-analysis. *Clin Gastroenterol Hepatol.* 2007, **5**(12), pp.1413-1417, 1417.e1411-1412.
35. Statista. *Number of esophageal cancer cases in England 1995-2018, by age.* [Online]. 2024. [Accessed 24.7.24].
36. Maitra, I., Date, R.S. and Martin, F.L. Towards screening Barrett's oesophagus: current guidelines, imaging modalities and future developments. *Clin J Gastroenterol.* 2020, **13**(5), pp.635-649.
37. Ronkainen, J., Aro, P., Storskrubb, T., Johansson, S.E., Lind, T., Bolling–Sternevald, E., Vieth, M., Stolte, M., Talley, N.J. and Agréus, L.J.G. Prevalence of Barrett's esophagus in the general population: an endoscopic study. 2005, **129**(6), pp.1825-1831.
38. Zagari, R.M., Fuccio, L., Wallander, M.-A., Johansson, S., Fiocca, R., Casanova, S., Farahmand, B.Y., Winchester, C.C., Roda, E. and Bazzoli, F.J.G. Gastro-oesophageal

reflux symptoms, oesophagitis and Barrett's oesophagus in the general population: the Loiano–Monghidoro study. 2008, **57**(10), pp.1354-1359.
39. Shiota, S., Singh, S., Anshasi, A., El-Serag, H.B.J.C.G. and Hepatology. Prevalence of Barrett's esophagus in Asian countries: a systematic review and meta-analysis. 2015, **13**(11), pp.1907-1918.
40. Sharma, N. and Ho, K.Y. Risk Factors for Barrett's Oesophagus. *Gastrointest Tumors.* 2016, **3**(2), pp.103-108.
41. Fitzgerald, R.C., di Pietro, M., Ragunath, K., Ang, Y., Kang, J.Y., Watson, P., Trudgill, N., Patel, P., Kaye, P.V., Sanders, S., O'Donovan, M., Bird-Lieberman, E., Bhandari, P., Jankowski, J.A., Attwood, S., Parsons, S.L., Loft, D., Lagergren, J., Moayyedi, P., Lyratzopoulos, G. and de Caestecker, J. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut.* 2014, **63**(1), pp.7-42.
42. Jankowski, J., Barr, H., Wang, K. and Delaney, B. Diagnosis and management of Barrett's oesophagus. *Bmj.* 2010, **341**, p.c4551.
43. Thrift, A.P., Kramer, J.R., Qureshi, Z., Richardson, P.A. and El-Serag, H.B. Age at onset of GERD symptoms predicts risk of Barrett's esophagus. *Am J Gastroenterol.* 2013, **108**(6), pp.915-922.
44. To, H., Clemons, N.J., Duong, C.P., Trainer, A.H. and Phillips, W.A. The Genetics of Barrett's Esophagus: A Familial and Population-Based Perspective. *Dig Dis Sci.* 2016, **61**(7), pp.1826-1834.
45. Findlay, J.M., Middleton, M.R. and Tomlinson, I. Genetic susceptibility to Barrett's oesophagus: Lessons from early studies. *United European Gastroenterol J.* 2016, **4**(4), pp.485-492.
46. Contino, G., Vaughan, T.L., Whiteman, D. and Fitzgerald, R.C. The Evolving Genomic Landscape of Barrett's Esophagus and Esophageal Adenocarcinoma. *Gastroenterology.* 2017, **153**(3), pp.657-673.e651.
47. Krishnamoorthi, R., Singh, S., Ragunathan, K., Visrodia, K., Wang, K.K., Katzka, D.A. and Iyer, P.G. Factors Associated With Progression of Barrett's Esophagus: A Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol.* 2018, **16**(7), pp.1046-1055.e1048.
48. Olsen, C.M., Pandeya, N., Green, A.C., Webb, P.M. and Whiteman, D.C. Population attributable fractions of adenocarcinoma of the esophagus and gastroesophageal junction. *Am J Epidemiol.* 2011, **174**(5), pp.582-590.
49. Chow, W.H., Finkle, W.D., McLaughlin, J.K., Frankl, H., Ziel, H.K. and Fraumeni, J.F., Jr. The relation of gastroesophageal reflux disease and its treatment to adenocarcinomas of the esophagus and gastric cardia. *Jama.* 1995, **274**(6), pp.474-477.
50. Biswas, S., Quante, M., Leedham, S. and Jansen, M. The metaplastic mosaic of Barrett's oesophagus. *Virchows Arch.* 2018, **472**(1), pp.43-54.
51. (NICE), N.I.f.H.a.C.E. *Dyspepsie - proven GORD.* [Online]. 2022. [Accessed 24/7/24].
52. Eusebi, L.H., Cirota, G.G., Zagari, R.M. and Ford, A.C.J.G. Global prevalence of Barrett's oesophagus and oesophageal cancer in individuals with gastro-oesophageal reflux: a systematic review and meta-analysis. 2021, **70**(3), pp.456-463.
53. Chandrasekar, V.T., Hamade, N., Desai, M., Rai, T., Gorrepati, V.S., Jegadeesan, R., Sathyamurthy, A. and Sharma, P.J.E. Significantly lower annual rates of neoplastic progression in short-compared to long-segment non-dysplastic Barrett's esophagus: a systematic review and meta-analysis. 2019, **51**(07), pp.665-672.

54. Westhoff, B., Brotze, S., Weston, A., McElhinney, C., Cherian, R., Mayo, M.S., Smith, H.J. and Sharma, P. The frequency of Barrett's esophagus in high-risk patients with chronic GERD. *Gastrointest Endosc.* 2005, **61**(2), pp.226-231.
55. Parasa, S., Desai, M., Vittal, A., Chandrasekar, V.T., Pervez, A., Kennedy, K.F., Gupta, N., Shaheen, N.J. and Sharma, P. Estimating neoplasia detection rate (NDR) in patients with Barrett's oesophagus based on index endoscopy: a systematic review and meta-analysis. *Gut.* 2019, **68**(12), pp.2122-2128.
56. Enestvedt, B.K., Lugo, R., Guarner-Argente, C., Shah, P., Falk, G.W., Furth, E. and Ginsberg, G.G. Location, location, location: does early cancer in Barrett's esophagus have a preference? *Gastrointest Endosc.* 2013, **78**(3), pp.462-467.
57. Cassani, L., Sumner, E., Slaughter, J.C. and Yachimski, P. Directional distribution of neoplasia in Barrett's esophagus is not influenced by distance from the gastroesophageal junction. *Gastrointest Endosc.* 2013, **77**(6), pp.877-882.
58. Singh, S., Garg, S.K., Singh, P.P., Iyer, P.G. and El-Serag, H.B. Acid-suppressive medications and risk of oesophageal adenocarcinoma in patients with Barrett's oesophagus: a systematic review and meta-analysis. *Gut.* 2014, **63**(8), pp.1229-1237.
59. Nguyen, D.M., Richardson, P. and El-Serag, H.B. Medications (NSAIDs, statins, proton pump inhibitors) and the risk of esophageal adenocarcinoma in patients with Barrett's esophagus. *Gastroenterology.* 2010, **138**(7), pp.2260-2266.
60. Kambhampati, S., Tieu, A.H., Luber, B., Wang, H. and Meltzer, S.J. Risk Factors for Progression of Barrett's Esophagus to High Grade Dysplasia and Esophageal Adenocarcinoma. *Sci Rep.* 2020, **10**(1), p.4899.
61. Lagergren, J., Bergström, R., Adami, H.O. and Nyrén, O. Association between medications that relax the lower esophageal sphincter and risk for esophageal adenocarcinoma. *Ann Intern Med.* 2000, **133**(3), pp.165-175.
62. Fortuny, J., Johnson, C.C., Bohlke, K., Chow, W.H., Hart, G., Kucera, G., Mujumdar, U., Ownby, D., Wells, K., Yood, M.U. and Engel, L.S. Use of anti-inflammatory drugs and lower esophageal sphincter-relaxing drugs and risk of esophageal and gastric cancers. *Clin Gastroenterol Hepatol.* 2007, **5**(10), pp.1154-1159.e1153.
63. Ranka, S., Gee, J.M., Johnson, I.T., Skinner, J., Hart, A.R. and Rhodes, M. Non-steroidal anti-inflammatory drugs, lower oesophageal sphincter-relaxing drugs and oesophageal cancer. A case-control study. *Digestion.* 2006, **74**(2), pp.109-115.
64. Jankowski, J.A.Z., de Caestecker, J., Love, S.B., Reilly, G., Watson, P., Sanders, S., Ang, Y., Morris, D., Bhandari, P., Brooks, C., Attwood, S., Harrison, R., Barr, H. and Moayyedi, P. Esomeprazole and aspirin in Barrett's oesophagus (AspECT): a randomised factorial trial. *Lancet.* 2018, **392**(10145), pp.400-408.
65. Chak, A., Ochs-Balcom, H., Falk, G., Grady, W.M., Kinnard, M., Willis, J.E., Elston, R. and Eng, C. Familiality in Barrett's esophagus, adenocarcinoma of the esophagus, and adenocarcinoma of the gastroesophageal junction. *Cancer Epidemiol Biomarkers Prev.* 2006, **15**(9), pp.1668-1673.
66. Collen, M.J., Lewis, J.H. and Benjamin, S.B. Gastric acid hypersecretion in refractory gastroesophageal reflux disease. *Gastroenterology.* 1990, **98**(3), pp.654-661.
67. Mulholland, M.W., Reid, B.J., Levine, D.S. and Rubin, C.E. Elevated gastric acid secretion in patients with Barrett's metaplastic epithelium. *Dig Dis Sci.* 1989, **34**(9), pp.1329-1334.
68. Taylor, J.B. and Rubenstein, J.H. Meta-analyses of the effect of symptoms of gastroesophageal reflux on the risk of Barrett's esophagus. *Am J Gastroenterol.* 2010, **105**(8), pp.1729, 1730-1727; quiz 1738.

69. Rubenstein, J.H. and Taylor, J.B. Meta-analysis: the association of oesophageal adenocarcinoma with symptoms of gastro-oesophageal reflux. 2010, **32**(10), pp.1222-1227.
70. Peters, Y., Al-Kaabi, A., Shaheen, N.J., Chak, A., Blum, A., Souza, R.F., Di Pietro, M., Iyer, P.G., Pech, O., Fitzgerald, R.C. and Siersema, P.D. Barrett oesophagus. *Nat Rev Dis Primers.* 2019, **5**(1), p.35.
71. Gastroenterology, B.S.o. *Diagnosis and management of Barrett esophagus: European*

Society of Gastrointestinal Endoscopy (ESGE) Guideline. [Online]. 2013. [Accessed 24.7.24].
72. Ronkainen, J., Aro, P., Storskrubb, T., Johansson, S.E., Lind, T., Bolling-Sternevald, E., Vieth, M., Stolte, M., Talley, N.J. and Agréus, L. Prevalence of Barrett's esophagus in the general population: an endoscopic study. *Gastroenterology.* 2005, **129**(6), pp.1825-1831.
73. Marques de Sá, I., Marcos, P., Sharma, P. and Dinis-Ribeiro, M. The global prevalence of Barrett's esophagus: A systematic review of the published literature. *United European Gastroenterol J.* 2020, **8**(9), pp.1086-1105.
74. Marques de Sá, I., Leal, C., Silva, J., Falcão, D., Felix, C., Nascimento, C., Boal Carvalho, P., Vasconcelos, H., Pedroto, I., Chagas, C., Cravo, M., Cotter, J., Sharma, P. and Dinis-Ribeiro, M. Prevalence of Barrett's esophagus in a Southern European country: a multicenter study. *Eur J Gastroenterol Hepatol.* 2021, **33**(1S Suppl 1), pp.e939-e943.
75. Zagari, R.M., Fuccio, L., Wallander, M.-A., Johansson, S., Fiocca, R., Casanova, S., Farahmand, B.Y., Winchester, C.C., Roda, E. and Bazzoli, F. Gastro-oesophageal reflux symptoms, oesophagitis and Barrett's oesophagus in the general population: the Loiano–Monghidoro study. 2008, **57**(10), pp.1354-1359.
76. Desai, T.K., Krishnan, K., Samala, N., Singh, J., Cluley, J., Perla, S. and Howden, C.W. The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut.* 2012, **61**(7), pp.970-976.
77. Klaver, E., Bureo Gonzalez, A., Mostafavi, N., Mallant-Hent, R., Duits, L.C., Baak, B., Böhmer, C.J.M., van Oijen, A., Naber, T., Scholten, P., Meijer, S.L., Bergman, J. and Pouw, R.E. Barrett's esophagus surveillance in a prospective Dutch multi-center community-based cohort of 985 patients demonstrates low risk of neoplastic progression. *United European Gastroenterol J.* 2021, **9**(8), pp.929-937.
78. Kastelein, F., van Olphen, S.H., Steyerberg, E.W., Spaander, M.C. and Bruno, M.J. Impact of surveillance for Barrett's oesophagus on tumour stage and survival of patients with neoplastic progression. *Gut.* 2016, **65**(4), pp.548-554.
79. Hamel, C., Ahmadzai, N., Beck, A., Thuku, M., Skidmore, B., Pussegoda, K., Bjerre, L., Chatterjee, A., Dennis, K., Ferri, L., Maziak, D.E., Shea, B.J., Hutton, B., Little, J., Moher, D. and Stevens, A. Screening for esophageal adenocarcinoma and precancerous conditions (dysplasia and Barrett's esophagus) in patients with chronic gastroesophageal reflux disease with or without other risk factors: two systematic reviews and one overview of reviews to inform a guideline of the Canadian Task Force on Preventive Health Care (CTFPHC). *Syst Rev.* 2020, **9**(1), p.20.
80. Rubenstein, J.H. and Inadomi, J.M. Cost-Effectiveness of Screening, Surveillance, and Endoscopic Eradication Therapies for Managing the Burden of Esophageal Adenocarcinoma. *Gastrointest Endosc Clin N Am.* 2021, **31**(1), pp.77-90.
81. Sami, S.S., Moriarty, J.P., Rosedahl, J.K., Borah, B.J., Katzka, D.A., Wang, K.K., Kisiel, J.B., Ragunath, K., Rubenstein, J.H. and Iyer, P.G. Comparative Cost Effectiveness of

Reflux-Based and Reflux-Independent Strategies for Barrett's Esophagus Screening. *Am J Gastroenterol.* 2021, **116**(8), pp.1620-1631.
82. Qumseya, B.J., Bukannan, A., Gendy, S., Ahemd, Y., Sultan, S., Bain, P., Gross, S.A., Iyer, P. and Wani, S. Systematic review and meta-analysis of prevalence and risk factors for Barrett's esophagus. *Gastrointest Endosc.* 2019, **90**(5), pp.707-717.e701.
83. Sawas, T., Zamani, S.A., Killcoyne, S., Dullea, A., Wang, K.K., Iyer, P.G., Fitzgerald, R.C. and Katzka, D.A. Limitations of Heartburn and Other Societies' Criteria in Barrett's Screening for Detecting De Novo Esophageal Adenocarcinoma. *Clin Gastroenterol Hepatol.* 2022, **20**(8), pp.1709-1718.
84. Nguyen, T.H., Thrift, A.P., Rugge, M. and El-Serag, H.B. Prevalence of Barrett's esophagus and performance of societal screening guidelines in an unreferred primary care population of U.S. veterans. *Gastrointest Endosc.* 2021, **93**(2), pp.409-419.e401.
85. Desai, M., Saligram, S., Gupta, N., Vennalaganti, P., Bansal, A., Choudhary, A., Vennelaganti, S., He, J., Titi, M., Maselli, R., Qumseya, B., Olyaee, M., Waxman, I., Repici, A., Hassan, C. and Sharma, P. Efficacy and safety outcomes of multimodal endoscopic eradication therapy in Barrett's esophagus-related neoplasia: a systematic review and pooled analysis. *Gastrointest Endosc.* 2017, **85**(3), pp.482-495.e484.
86. van Vilsteren, F.G., Pouw, R.E., Seewald, S., Alvarez Herrero, L., Sondermeijer, C.M., Visser, M., Ten Kate, F.J., Yu Kim Teng, K.C., Soehendra, N., Rösch, T., Weusten, B.L. and Bergman, J.J. Stepwise radical endoscopic resection versus radiofrequency ablation for Barrett's oesophagus with high-grade dysplasia or early cancer: a multicentre randomised trial. *Gut.* 2011, **60**(6), pp.765-773.
87. Phoa, K.N., van Vilsteren, F.G., Weusten, B.L., Bisschops, R., Schoon, E.J., Ragunath, K., Fullarton, G., Di Pietro, M., Ravi, N., Visser, M., Offerhaus, G.J., Seldenrijk, C.A., Meijer, S.L., ten Kate, F.J., Tijssen, J.G. and Bergman, J.J. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *Jama.* 2014, **311**(12), pp.1209-1217.
88. Pasricha, S., Bulsiewicz, W.J., Hathorn, K.E., Komanduri, S., Muthusamy, V.R., Rothstein, R.I., Wolfsen, H.C., Lightdale, C.J., Overholt, B.F., Camara, D.S., Dellon, E.S., Lyday, W.D., Ertan, A., Chmielewski, G.W. and Shaheen, N.J. Durability and predictors of successful radiofrequency ablation for Barrett's esophagus. *Clin Gastroenterol Hepatol.* 2014, **12**(11), pp.1840-1847.e1841.
89. Gregson, E.M., Bornschein, J. and Fitzgerald, R.C. Genetic progression of Barrett's oesophagus to oesophageal adenocarcinoma. *British Journal of Cancer.* 2016, **115**(4), pp.403-410.
90. Chao, D.L., Sanchez, C.A., Galipeau, P.C., Blount, P.L., Paulson, T.G., Cowan, D.S., Ayub, K., Odze, R.D., Rabinovitch, P.S. and Reid, B.J. Cell Proliferation, Cell Cycle Abnormalities, and Cancer Outcome in Patients with Barrett's Esophagus: A Long-term Prospective Study. *Clinical Cancer Research.* 2008, **14**(21), pp.6988-6995.
91. Atherfold, P.A. and Jankowski, J.A. Molecular biology of Barrett's cancer. *Best Practice & Research Clinical Gastroenterology.* 2006, **20**(5), pp.813-827.
92. Inadomi, J., Alastal, H., Bonavina, L., Gross, S., Hunt, R.H., Mashimo, H., di Pietro, M., Rhee, H., Shah, M., Tolone, S., Wang, D.H. and Xie, S.-H. Recent advances in Barrett's esophagus. 2018, **1434**(1), pp.227-238.
93. Gindea, C., Birla, R., Hoara, P., Caragui, A. and Constantinoiu, S. Barrett esophagus: history, definition and etiopathogeny. *J Med Life.* 2014, **7 Spec No. 3**(Spec Iss 3), pp.23-30.

94. Jiang, M., Li, H., Zhang, Y., Yang, Y., Lu, R., Liu, K., Lin, S., Lan, X., Wang, H., Wu, H., Zhu, J., Zhou, Z., Xu, J., Lee, D.K., Zhang, L., Lee, Y.C., Yuan, J., Abrams, J.A., Wang, T.C., Sepulveda, A.R., Wu, Q., Chen, H., Sun, X., She, J., Chen, X. and Que, J. Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. *Nature.* 2017, **550**(7677), pp.529-533.
95. Minacapelli, C.D., Bajpai, M., Geng, X., Cheng, C.L., Chouthai, A.A., Souza, R., Spechler, S.J. and Das, K.M. Barrett's metaplasia develops from cellular reprograming of esophageal squamous epithelium due to gastroesophageal reflux. *Am J Physiol Gastrointest Liver Physiol.* 2017, **312**(6), pp.G615-g622.
96. Ling, F.C., Khochfar, J., Baldus, S.E., Brabender, J., Drebber, U., Bollschweiler, E., Hoelscher, A.H. and Schneider, P.M. HIF-1alpha protein expression is associated with the environmental inflammatory reaction in Barrett's metaplasia. *Dis Esophagus.* 2009, **22**(8), pp.694-699.
97. Colleypriest, B.J., Burke, Z.D., Griffiths, L.P., Chen, Y., Yu, W.Y., Jover, R., Bock, M., Biddlestone, L., Quinlan, J.M., Ward, S.G., Mark Farrant, J., Slack, J.M.W. and Tosh, D. Hnf4α is a key gene that can generate columnar metaplasia in oesophageal epithelium. *Differentiation.* 2017, **93**, pp.39-49.
98. O'Riordan, J.M., Abdel-latif, M.M., Ravi, N., McNamara, D., Byrne, P.J., McDonald, G.S., Keeling, P.W., Kelleher, D. and Reynolds, J.V. Proinflammatory cytokine and nuclear factor kappa-B expression along the inflammation-metaplasia-dysplasia-adenocarcinoma sequence in the esophagus. *Am J Gastroenterol.* 2005, **100**(6), pp.1257-1264.
99. Moons, L.M., Kusters, J.G., van Delft, J.H., Kuipers, E.J., Gottschalk, R., Geldof, H., Bode, W.A., Stoof, J., van Vliet, A.H., Ketelslegers, H.B., Kleinjans, J.C. and Siersema, P.D. A pro-inflammatory genotype predisposes to Barrett's esophagus. *Carcinogenesis.* 2008, **29**(5), pp.926-931.
100. Milano, F., van Baal, J.W., Buttar, N.S., Rygiel, A.M., de Kort, F., DeMars, C.J., Rosmolen, W.D., Bergman, J.J., J, V.A.M., Wang, K.K., Peppelenbosch, M.P. and Krishnadath, K.K. Bone morphogenetic protein 4 expressed in esophagitis induces a columnar phenotype in esophageal squamous cells. *Gastroenterology.* 2007, **132**(7), pp.2412-2421.
101. Chen, H., Fang, Y., Tevebaugh, W., Orlando, R.C., Shaheen, N.J. and Chen, X. Molecular mechanisms of Barrett's esophagus. *Dig Dis Sci.* 2011, **56**(12), pp.3405-3420.
102. Kong, J., Crissey, M.A., Stairs, D.B., Sepulveda, A.R. and Lynch, J.P. Cox2 and β-catenin/T-cell factor signaling intestinalize human esophageal keratinocytes when cultured under organotypic conditions. *Neoplasia.* 2011, **13**(9), pp.792-805.
103. Murata-Kamiya, N., Kurashima, Y., Teishikata, Y., Yamahashi, Y., Saito, Y., Higashi, H., Aburatani, H., Akiyama, T., Peek, R.M., Azuma, T. and Hatakeyama, M. Helicobacter pylori CagA interacts with E-cadherin and deregulates the β-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells. *Oncogene.* 2007, **26**(32), pp.4617-4626.
104. Wang, D.H., Clemons, N.J., Miyashita, T., Dupuy, A.J., Zhang, W., Szczepny, A., Corcoran-Schwartz, I.M., Wilburn, D.L., Montgomery, E.A., Wang, J.S., Jenkins, N.A., Copeland, N.A., Harmon, J.W., Phillips, W.A. and Watkins, D.N. Aberrant epithelial-mesenchymal Hedgehog signaling characterizes Barrett's metaplasia. *Gastroenterology.* 2010, **138**(5), pp.1810-1822.
105. Yamanaka, Y., Shiotani, A., Fujimura, Y., Ishii, M., Fujita, M., Matsumoto, H., Tarumi, K., Kamada, T., Hata, J. and Haruma, K. Expression of Sonic hedgehog (SHH) and

CDX2 in the columnar epithelium of the lower oesophagus. *Dig Liver Dis.* 2011, **43**(1), pp.54-59.
106. Wang, D.H., Tiwari, A., Kim, M.E., Clemons, N.J., Regmi, N.L., Hodges, W.A., Berman, D.M., Montgomery, E.A., Watkins, D.N., Zhang, X., Zhang, Q., Jie, C., Spechler, S.J. and Souza, R.F. Hedgehog signaling regulates FOXA2 in esophageal embryogenesis and Barrett's metaplasia. *J Clin Invest.* 2014, **124**(9), pp.3767-3780.
107. Vega, M.E., Giroux, V., Natsuizaka, M., Liu, M., Klein-Szanto, A.J., Stairs, D.B., Nakagawa, H., Wang, K.K., Wang, T.C., Lynch, J.P. and Rustgi, A.K. Inhibition of Notch signaling enhances transdifferentiation of the esophageal squamous epithelium towards a Barrett's-like metaplasia via KLF4. *Cell Cycle.* 2014, **13**(24), pp.3857-3866.
108. Morrow, D.J., Avissar, N.E., Toia, L., Redmond, E.M., Watson, T.J., Jones, C., Raymond, D.P., Litle, V. and Peters, J.H. Pathogenesis of Barrett's esophagus: bile acids inhibit the Notch signaling pathway with induction of CDX2 gene expression in human esophageal cells. *Surgery.* 2009, **146**(4), pp.714-721; discussion 721-712.
109. Menke, V., van Es, J.H., de Lau, W., van den Born, M., Kuipers, E.J., Siersema, P.D., de Bruin, R.W., Kusters, J.G. and Clevers, H. Conversion of metaplastic Barrett's epithelium into post-mitotic goblet cells by gamma-secretase inhibition. *Dis Model Mech.* 2010, **3**(1-2), pp.104-110.
110. Hamilton, S.R. and Yardley, J.H. Regnerative of cardiac type mucosa and acquisition of Barrett mucosa after esophagogastrostomy. *Gastroenterology.* 1977, **72**(4 Pt 1), pp.669-675.
111. Nakanishi, Y., Saka, M., Eguchi, T., Sekine, S., Taniguchi, H. and Shimoda, T. Distribution and significance of the oesophageal and gastric cardiac mucosae: a study of 131 operation specimens. 2007, **51**(4), pp.515-519.
112. Koak, Y. and Winslet, M. Changing role of in vivo models in columnar-lined lower esophagus. *Diseases of the Esophagus.* 2002, **15**(4), pp.271-277.
113. Yu, W.-Y., Slack, J.M.W. and Tosh, D. Conversion of columnar to stratified squamous epithelium in the developing mouse oesophagus. *Developmental Biology.* 2005, **284**(1), pp.157-170.
114. Barham, C.P., Jones, R.L., Biddlestone, L.R., Hardwick, R.H., Shepherd, N.A. and Barr, H. Photothermal laser ablation of Barrett's oesophagus: endoscopic and histological evidence of squamous re-epithelialisation. *Gut.* 1997, **41**(3), pp.281-284.
115. Shields, H.M., Rosenberg, S.J., Zwas, F.R., Ransil, B.J., Lembo, A.J. and Odze, R. Prospective evaluation of multilayered epithelium in Barrett's esophagus. *Am J Gastroenterol.* 2001, **96**(12), pp.3268-3273.
116. Quante, M., Bhagat, G., Abrams, J.A., Marache, F., Good, P., Lee, M.D., Lee, Y., Friedman, R., Asfaha, S., Dubeykovskaya, Z., Mahmood, U., Figueiredo, J.L., Kitajewski, J., Shawber, C., Lightdale, C.J., Rustgi, A.K. and Wang, T.C. Bile acid and inflammation activate gastric cardia stem cells in a mouse model of Barrett-like metaplasia. *Cancer Cell.* 2012, **21**(1), pp.36-51.
117. Wang, X., Ouyang, H., Yamamoto, Y., Kumar, P.A., Wei, T.S., Dagher, R., Vincent, M., Lu, X., Bellizzi, A.M., Ho, K.Y., Crum, C.P., Xian, W. and McKeon, F. Residual embryonic cells as precursors of a Barrett's-like metaplasia. *Cell.* 2011, **145**(7), pp.1023-1035.
118. Demicco, E.G., Farris, A.B., Baba, Y., Agbor-Etang, B., Bergethon, K., Mandal, R., Daives, D., Fukuoka, J., Shimizu, M., Dias-Santagata, D., Ogino, S., Iafrate, A.J., Gaissert, H.A. and Mino-Kenudson, M. The dichotomy in carcinogenesis of the distal esophagus and esophagogastric junction: intestinal-type vs cardiac-type mucosa-associated adenocarcinoma. *Modern Pathology.* 2011, **24**(9), pp.1177-1190.

119. Jankowski, J.A., Harrison, R.F., Perry, I., Balkwill, F. and Tselepis, C. Barrett's metaplasia. *Lancet.* 2000, **356**(9247), pp.2079-2085.
120. Alonso, L. and Fuchs, E. Stem cells of the skin epithelium. *Proc Natl Acad Sci U S A.* 2003, **100 Suppl 1**(Suppl 1), pp.11830-11835.
121. Ahnen, D.J., Poulsom, R., Stamp, G.W., Elia, G., Pike, C., Jeffery, R., Longcroft, J., Rio, M.C., Chambon, P. and Wright, N.A. The ulceration-associated cell lineage (UACL) reiterates the Brunner's gland differentiation programme but acquires the proliferative organization of the gastric gland. *J Pathol.* 1994, **173**(4), pp.317-326.
122. Coad, R.A., Woodman, A.C., Warner, P.J., Barr, H., Wright, N.A. and Shepherd, N.A. On the histogenesis of Barrett's oesophagus and its associated squamous islands: a three-dimensional study of their morphological relationship with native oesophageal gland ducts. *J Pathol.* 2005, **206**(4), pp.388-394.
123. Chang, C.L., Lao-Sirieix, P., Save, V., De La Cueva Mendez, G., Laskey, R. and Fitzgerald, R.C. Retinoic acid-induced glandular differentiation of the oesophagus. *Gut.* 2007, **56**(7), pp.906-917.
124. Sarosi, G., Brown, G., Jaiswal, K., Feagins, L.A., Lee, E., Crook, T.W., Souza, R.F., Zou, Y.S., Shay, J.W. and Spechler, S.J. Bone marrow progenitor cells contribute to esophageal regeneration and metaplasia in a rat model of Barrett's esophagus. *Dis Esophagus.* 2008, **21**(1), pp.43-50.
125. Souza, R.F. Reflux esophagitis and its role in the pathogenesis of Barrett's metaplasia. *J Gastroenterol.* 2017, **52**(7), pp.767-776.
126. Martinez, P., Timmer, M.R., Lau, C.T., Calpe, S., Sancho-Serra Mdel, C., Straub, D., Baker, A.M., Meijer, S.L., Kate, F.J., Mallant-Hent, R.C., Naber, A.H., van Oijen, A.H., Baak, L.C., Scholten, P., Böhmer, C.J., Fockens, P., Bergman, J.J., Maley, C.C., Graham, T.A. and Krishnadath, K.K. Dynamic clonal equilibrium and predetermined cancer risk in Barrett's oesophagus. *Nat Commun.* 2016, **7**, p.12158.
127. Stachler, M., Bao, C., Tourdot, R., Brunette, G., Stewart, C., Sun, L., Baba, H., Watanabe, M., Agoston, A., Jajoo, K., Davison, J., Nason, K., Getz, G., Wang, K., Imamura, Y., Odze, R., Bass, A. and Zhang, C.-Z. *Genomic signatures of past and present chromosomal instability in the evolution of Barrett's esophagus to esophageal adenocarcinoma*. bioRxiv. 2021. 2021. Available from: http://europepmc.org/abstract/PPR/PPR303967

https://www.biorxiv.org/content/biorxiv/early/2021/03/28/2021.03.26.437288.full.pdf

https://doi.org/10.1101/2021.03.26.437288
128. Li, X., Galipeau, P.C., Paulson, T.G., Sanchez, C.A., Arnaudo, J., Liu, K., Sather, C.L., Kostadinov, R.L., Odze, R.D., Kuhner, M.K., Maley, C.C., Self, S.G., Vaughan, T.L., Blount, P.L. and Reid, B.J. Temporal and Spatial Evolution of Somatic Chromosomal Alterations: A Case-Cohort Study of Barrett's Esophagus. *Cancer Prevention Research.* 2014, **7**(1), pp.114-127.
129. Ross-Innes, C.S., Becq, J., Warren, A., Cheetham, R.K., Northen, H., O'Donovan, M., Malhotra, S., di Pietro, M., Ivakhno, S., He, M., Weaver, J.M.J., Lynch, A.G., Kingsbury, Z., Ross, M., Humphray, S., Bentley, D. and Fitzgerald, R.C. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet.* 2015, **47**(9), pp.1038-1046.
130. Killcoyne, S., Gregson, E., Wedge, D.C., Woodcock, D.J., Eldridge, M.D., de la Rue, R., Miremadi, A., Abbas, S., Blasko, A., Kosmidou, C., Januszewicz, W., Jenkins, A.V., Gerstung, M. and Fitzgerald, R.C. Genomic copy number predicts esophageal cancer years before transformation. *Nat Med.* 2020, **26**(11), pp.1726-1732.

131. Douville, C., Moinova, H.R., Thota, P.N., Shaheen, N.J., Iyer, P.G., Canto, M.I., Wang, J.S., Dumot, J.A., Faulx, A., Kinzler, K.W., Papadopoulos, N., Vogelstein, B., Markowitz, S.D., Bettegowda, C., Willis, J.E. and Chak, A. Massively Parallel Sequencing of Esophageal Brushings Enables an Aneuploidy-Based Classification of Patients With Barrett's Esophagus. *Gastroenterology.* 2021, **160**(6), pp.2043-2054.e2042.

132. Weaver, J.M.J., Ross-Innes, C.S., Shannon, N., Lynch, A.G., Forshew, T., Barbera, M., Murtaza, M., Ong, C.-A.J., Lao-Sirieix, P., Dunning, M.J., Smith, L., Smith, M.L., Anderson, C.L., Carvalho, B., O'Donovan, M., Underwood, T.J., May, A.P., Grehan, N., Hardwick, R., Davies, J., Oloumi, A., Aparicio, S., Caldas, C., Eldridge, M.D., Edwards, P.A.W., Rosenfeld, N., Tavaré, S., Fitzgerald, R.C., Hayes, S.J., Yeng, A., Lydon, A.-M., Dharmaprasad, S., Greer, S., Preston, S., Oakes, S., Save, V., Paterson-Brown, S., Tucker, O., Alderson, D., Taniere, P., Kelly, J., Byrne, J., Sharland, D., Holling, N., Boulter, L., Noble, F., Stacey, B., Crichton, C., Barr, H., Shepherd, N., Almond, L.M., Old, O., Lagergren, J., Gossage, J., Davies, A., Mason, R., Chang, F., Zylstra, J., Sanders, G., Wheatley, T., Berrisford, R., Bracey, T., Harden, C., Bunting, D., Roques, T., Nobes, J., Loo, S., Lewis, M., Cheong, E., Priest, O., Parsons, S.L., Soomro, I., Kaye, P., Saunders, J., Pang, V., Welch, N.T., Catton, J.A., Duffy, J.P., Ragunath, K., Lovat, L., Haidry, R., Miah, H., Kerr, S., Eneh, V., Butawan, R., Igali, L., Ford, H., Gilligan, D., Safranek, P., Hindmarsh, A., Sudjendran, V., Metz, A., Carroll, N., Scott, M., Cluroe, A., Miremadi, A., Mahler-Araujo, B., Knight, O., Nutzinger, B., Peters, C., Abdullahi, Z., Debriram-Beecham, I., Malhotra, S., Crawte, J., MacRae, S., Noorani, A., Elliott, R.F., Li, X., Bower, L., Achilleos, A., Bornschein, J., Zeki, S., Chettouh, H., Secrier, M., de Silva, N., Gregson, E., Yang, T.-P., O'Neil, J.R. and the, O.C. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat Genet.* 2014, **46**(8), pp.837-843.

133. Stachler, M.D., Taylor-Weiner, A., Peng, S., McKenna, A., Agoston, A.T., Odze, R.D., Davison, J.M., Nason, K.S., Loda, M., Leshchiner, I., Stewart, C., Stojanov, P., Seepo, S., Lawrence, M.S., Ferrer-Torres, D., Lin, J., Chang, A.C., Gabriel, S.B., Lander, E.S., Beer, D.G., Getz, G., Carter, S.L. and Bass, A.J. Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet.* 2015, **47**(9), pp.1047-1055.

134. Luebeck, J., Ng, A.W.T., Galipeau, P.C., Li, X., Sanchez, C.A., Katz-Summercorn, A.C., Kim, H., Jammula, S., He, Y., Lippman, S.M., Verhaak, R.G.W., Maley, C.C., Alexandrov, L.B., Reid, B.J., Fitzgerald, R.C., Paulson, T.G., Chang, H.Y., Wu, S., Bafna, V. and Mischel, P.S. Extrachromosomal DNA in the cancerous transformation of Barrett's oesophagus. *Nature.* 2023, **616**(7958), pp.798-805.

135. Schröder, J., Chegwidden, L., Maj, C., Gehlen, J., Speller, J., Böhmer, A.C., Borisov, O., Hess, T., Kreuser, N., Venerito, M., Alakus, H., May, A., Gerges, C., Schmidt, T., Thieme, R., Heider, D., Hillmer, A.M., Reingruber, J., Lyros, O., Dietrich, A., Hoffmeister, A., Mehdorn, M., Lordick, F., Stocker, G., Hohaus, M., Reim, D., Kandler, J., Müller, M., Ebigbo, A., Fuchs, C., Bruns, C.J., Hölscher, A.H., Lang, H., Grimminger, P.P., Dakkak, D., Vashist, Y., May, S., Görg, S., Franke, A., Ellinghaus, D., Galavotti, S., Veits, L., Weismüller, J., Dommermuth, J., Benner, U., Rösch, T., Messmann, H., Schumacher, B., Neuhaus, H., Schmidt, C., Wissinowski, T.T., Nöthen, M.M., Dong, J., Ong, J.S., Buas, M.F., Thrift, A.P., Vaughan, T.L., Tomlinson, I., Whiteman, D.C., Fitzgerald, R.C., Jankowski, J., Vieth, M., Mayr, A., Gharahkhani, P., MacGregor, S., Gockel, I., Palles, C. and Schumacher, J. GWAS meta-analysis of 16 790 patients with Barrett's oesophagus and oesophageal adenocarcinoma identifies 16 novel genetic risk loci and provides insights into disease aetiology beyond the single marker level. *Gut.* 2023, **72**(4), pp.612-623.

136. Jones, C.M. *The Cellular Response to Microenvironmental*

Stress in Barrett's Oesophagus. thesis, University of Leeds, 2021.
137. Newell, F., Patel, K., Gartside, M., Krause, L., Brosda, S., Aoude, L.G., Loffler, K.A., Bonazzi, V.F., Patch, A.-M., Kazakoff, S.H., Holmes, O., Xu, Q., Wood, S., Leonard, C., Lampe, G., Lord, R.V., Whiteman, D.C., Pearson, J.V., Nones, K., Waddell, N. and Barbour, A.P. Complex structural rearrangements are present in high-grade dysplastic Barrett's oesophagus samples. *BMC Medical Genomics.* 2019, **12**(1), p.31.
138. Secrier, M., Li, X., de Silva, N., Eldridge, M.D., Contino, G., Bornschein, J., MacRae, S., Grehan, N., O'Donovan, M., Miremadi, A., Yang, T.-P., Bower, L., Chettouh, H., Crawte, J., Galeano-Dalmau, N., Grabowska, A., Saunders, J., Underwood, T., Waddell, N., Barbour, A.P., Nutzinger, B., Achilleos, A., Edwards, P.A.W., Lynch, A.G., Tavaré, S., Fitzgerald, R.C., Noorani, A., Elliott, R.F., Weaver, J., Ross-Innes, C., Smith, L., Abdullahi, Z., de la Rue, R., Cluroe, A., Malhotra, S., Hardwick, R., Ford, H., Smith, M.L., Davies, J., Turkington, R., Hayes, S.J., Ang, Y., Preston, S.R., Oakes, S., Bagwan, I., Save, V., Skipworth, R.J.E., Hupp, T.R., O'Neill, J.R., Tucker, O., Taniere, P., Noble, F., Owsley, J., Lovat, L., Haidry, R., Eneh, V., Crichton, C., Barr, H., Shepherd, N., Old, O., Lagergren, J., Gossage, J., Davies, A., Chang, F., Zylstra, J., Sanders, G., Berrisford, R., Harden, C., Bunting, D., Lewis, M., Cheong, E., Kumar, B., Parsons, S.L., Soomro, I., Kaye, P., Collier, P., Igali, L., Welch, I., Scott, M., Sothi, S., Suortamo, S., Lishman, S., Beardsmore, D., Francies, H.E., Garnett, M.J., Pearson, J.V., Nones, K., Patch, A.-M., Grimmond, S.M., the Oesophageal Cancer, C. and Molecular Stratification, C. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat Genet.* 2016, **48**(10), pp.1131-1141.
139. Dulak, A.M., Stojanov, P., Peng, S., Lawrence, M.S., Fox, C., Stewart, C., Bandla, S., Imamura, Y., Schumacher, S.E., Shefler, E., McKenna, A., Carter, S.L., Cibulskis, K., Sivachenko, A., Saksena, G., Voet, D., Ramos, A.H., Auclair, D., Thompson, K., Sougnez, C., Onofrio, R.C., Guiducci, C., Beroukhim, R., Zhou, Z., Lin, L., Lin, J., Reddy, R., Chang, A., Landrenau, R., Pennathur, A., Ogino, S., Luketich, J.D., Golub, T.R., Gabriel, S.B., Lander, E.S., Beer, D.G., Godfrey, T.E., Getz, G. and Bass, A.J. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet.* 2013, **45**(5), pp.478-486.
140. Frankell, A.M., Jammula, S., Li, X., Contino, G., Killcoyne, S., Abbas, S., Perner, J., Bower, L., Devonshire, G., Ococks, E., Grehan, N., Mok, J., O'Donovan, M., MacRae, S., Eldridge, M.D., Tavaré, S., Fitzgerald, R.C., Noorani, A., Edwards, P.A.W., Grehan, N., Nutzinger, B., Hughes, C., Fidziukiewicz, E., MacRae, S., Northrop, A., Contino, G., Li, X., de la Rue, R., Katz-Summercorn, A., Abbas, S., Loureda, D., O'Donovan, M., Miremadi, A., Malhotra, S., Tripathi, M., Tavaré, S., Lynch, A.G., Eldridge, M., Secrier, M., Devonshire, G., Perner, J., Jammula, S., Davies, J., Crichton, C., Carroll, N., Safranek, P., Hindmarsh, A., Sujendran, V., Hayes, S.J., Ang, Y., Sharrocks, A., Preston, S.R., Oakes, S., Bagwan, I., Save, V., Skipworth, R.J.E., Hupp, T.R., O'Neill, J.R., Tucker, O., Beggs, A., Taniere, P., Puig, S., Underwood, T.J., Walker, R.C., Grace, B.L., Barr, H., Shepherd, N., Old, O., Lagergren, J., Gossage, J., Davies, A., Chang, F., Zylstra, J., Mahadeva, U., Goh, V., Ciccarelli, F.D., Sanders, G., Berrisford, R., Harden, C., Lewis, M., Cheong, E., Kumar, B., Parsons, S.L., Soomro, I., Kaye, P., Saunders, J., Lovat, L., Haidry, R., Igali, L., Scott, M., Sothi, S., Suortamo, S., Lishman, S., Hanna, G.B., Moorthy, K., Peters, C.J., Grabowska, A., Turkington, R., McManus, D., Coleman, H., Khoo, D., Fickling, W., Fitzgerald, R.C., the Oesophageal Cancer, C.

and Molecular Stratification, C. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet.* 2019, **51**(3), pp.506-516.

141. Mourikis, T.P., Benedetti, L., Foxall, E., Temelkovski, D., Nulsen, J., Perner, J., Cereda, M., Lagergren, J., Howell, M., Yau, C., Fitzgerald, R.C., Scaffidi, P., Noorani, A., Edwards, P.A.W., Elliott, R.F., Grehan, N., Nutzinger, B., Hughes, C., Fidziukiewicz, E., Bornschein, J., MacRae, S., Crawte, J., Northrop, A., Contino, G., Li, X., de la Rue, R., Katz-Summercorn, A., Abbas, S., Loureda, D., O'Donovan, M., Miremadi, A., Malhotra, S., Tripathi, M., Tavaré, S., Lynch, A.G., Eldridge, M., Secrier, M., Bower, L., Devonshire, G., Jammula, S., Davies, J., Crichton, C., Carroll, N., Safranek, P., Hindmarsh, A., Sujendran, V., Hayes, S.J., Ang, Y., Sharrocks, A., Preston, S.R., Oakes, S., Bagwan, I., Save, V., Skipworth, R.J.E., Hupp, T.R., Robert O'Neill, J., Tucker, O., Beggs, A., Taniere, P., Puig, S., Underwood, T.J., Walker, R.C., Grace, B.L., Barr, H., Shepherd, N., Old, O., Gossage, J., Davies, A., Chang, F., Zylstra, J., Mahadeva, U., Goh, V., Sanders, G., Berrisford, R., Harden, C., Lewis, M., Cheong, E., Kumar, B., Parsons, S.L., Soomro, I., Kaye, P., Saunders, J., Lovat, L., Haidry, R., Igali, L., Scott, M., Sothi, S., Suortamo, S., Lishman, S., Hanna, G.B., Peters, C.J., Moorthy, K., Grabowska, A., Turkington, R., McManus, D., Khoo, D., Fickling, W., Ciccarelli, F.D., The Oesophageal Cancer, C. andMolecular Stratification, C. Patient-specific cancer genes contribute to recurrently perturbed pathways and establish therapeutic vulnerabilities in esophageal adenocarcinoma. *Nat Commun.* 2019, **10**(1), p.3101.

142. M. Naeini, M., Newell, F., Aoude, L.G., Bonazzi, V.F., Patel, K., Lampe, G., Koufariotis, L.T., Lakis, V., Addala, V., Kondrashova, O., Johnston, R.L., Sharma, S., Brosda, S., Holmes, O., Leonard, C., Wood, S., Xu, Q., Thomas, J., Walpole, E., Tao Mai, G., Ackland, S.P., Martin, J., Burge, M., Finch, R., Karapetis, C.S., Shannon, J., Nott, L., Bohmer, R., Wilson, K., Barnes, E., Zalcberg, J.R., Mark Smithers, B., Simes, J., Price, T., Gebski, V., Nones, K., Watson, D.I., Pearson, J.V., Barbour, A.P. and Waddell, N. Multi-omic features of oesophageal adenocarcinoma in patients treated with preoperative neoadjuvant therapy. *Nat Commun.* 2023, **14**(1), p.3155.

143. Kim, J., Bowlby, R., Mungall, A.J., Robertson, A.G., Odze, R.D., Cherniack, A.D., Shih, J., Pedamallu, C.S., Cibulskis, C., Dunford, A., Meier, S.R., Kim, J., Raphael, B.J., Wu, H.-T., Wong, A.M., Willis, J.E., Bass, A.J., Derks, S., Garman, K., McCall, S.J., Wiznerowicz, M., Pantazi, A., Parfenov, M., Thorsson, V., Shmulevich, I., Dhankani, V., Miller, M., Sakai, R., Wang, K., Schultz, N., Shen, R., Arora, A., Weinhold, N., Sánchez-Vega, F., Kelsen, D.P., Zhang, J., Felau, I., Demchok, J., Rabkin, C.S., Camargo, M.C., Zenklusen, J.C., Bowen, J., Leraas, K., Lichtenberg, T.M., Curtis, C., Seoane, J.A., Ojesina, A.I., Beer, D.G., Gulley, M.L., Pennathur, A., Luketich, J.D., Zhou, Z., Weisenberger, D.J., Akbani, R., Lee, J.-S., Liu, W., Mills, G.B., Zhang, W., Reid, B.J., Hinoue, T., Laird, P.W., Shen, H., Piazuelo, M.B., Schneider, B.G., McLellan, M., Taylor-Weiner, A., Cibulskis, C., Lawrence, M., Cibulskis, K., Stewart, C., Getz, G., Lander, E., Gabriel, S.B., Ding, L., McLellan, M.D., Miller, C.A., Appelbaum, E.L., Cordes, M.G., Fronick, C.C., Fulton, L.A., Mardis, E.R., Wilson, R.K., Schmidt, H.K., Fulton, R.S., Ally, A., Balasundaram, M., Bowlby, R., Carlsen, R., Chuah, E., Dhalla, N., Holt, R.A., Jones, S.J.M., Kasaian, K., Brooks, D., Li, H.I., Ma, Y., Marra, M.A., Mayo, M., Moore, R.A., Mungall, A.J., Mungall, K.L., Robertson, A.G., Schein, J.E., Sipahimalani, P., Tam, A., Thiessen, N., Wong, T., Cherniack, A.D., Shih, J., Pedamallu, C.S., Beroukhim, R., Bullman, S., Cibulskis, C., Murray, B.A., Saksena, G., Schumacher, S.E., Gabriel, S., Meyerson, M., Hadjipanayis, A., Kucherlapati, R., Pantazi, A., Parfenov, M., Ren, X., Park, P.J., Lee, S., Kucherlapati, M., Yang, L.,

Baylin, S.B., Hoadley, K.A., Weisenberger, D.J., Bootwalla, M.S., Lai, P.H., Van Den Berg, D.J., Berrios, M., Holbrook, A., Akbani, R., Hwang, J.-E., Jang, H.-J., Liu, W., Weinstein, J.N., Lee, J.-S., Lu, Y., Sohn, B.H., Mills, G., Seth, S., Protopopov, A., Bristow, C.A., Mahadeshwar, H.S., Tang, J., Song, X., Zhang, J., Laird, P.W., Hinoue, T., Shen, H., Cho, J., Defrietas, T., Frazer, S., Gehlenborg, N., Heiman, D.I., Lawrence, M.S., Lin, P., Meier, S.R., Noble, M.S., Voet, D., Zhang, H., Kim, J., Polak, P., Saksena, G., Chin, L., Getz, G., Wong, A.M., Raphael, B.J., Wu, H.-T., Lee, S., Park, P.J., Yang, L., Thorsson, V., Bernard, B., Iype, L., Miller, M., Reynolds, S.M., Shmulevich, I., Dhankani, V., Abeshouse, A., Arora, A., Armenia, J., Kundra, R., Ladanyi, M., Lehmann, K.-V., Gao, J., Sander, C., Schultz, N., Sánchez-Vega, F., Shen, R., Weinhold, N., Chakravarty, D., Zhang, H., Radenbaugh, A., Hegde, A., Akbani, R., Liu, W., Weinstein, J.N., Chin, L., Bristow, C.A., Lu, Y., Penny, R., Crain, D., Gardner, J., Curley, E., Mallery, D., Morris, S., Paulauskis, J., Shelton, T., Shelton, C., Bowen, J., Frick, J., Gastier-Foster, J.M., Gerken, M., Leraas, K.M., Lichtenberg, T.M., Ramirez, N.C., Wise, L., Zmuda, E., Tarvin, K., Saller, C., Park, Y.S., Button, M., Carvalho, A.L., Reis, R.M., Matsushita, M.M., Lucchesi, F., de Oliveira, A.T., Le, X., Paklina, O., Setdikova, G., Lee, J.-H., Bennett, J., Iacocca, M., Huelsenbeck-Dill, L., Potapova, O., Voronina, O., Liu, O., Fulidou, V., The Cancer Genome Atlas Research, N., Analysis Working Group: Asan, U., Agency, B.C.C., Brigham, Women's, H., Broad, I., Brown, U., Case Western Reserve, U., Dana-Farber Cancer, I., Duke, U., Greater Poland Cancer, C., Harvard Medical, S., Institute for Systems, B., Leuven, K.U., Mayo, C., Memorial Sloan Kettering Cancer, C., National Cancer, I., Nationwide Children's, H., Stanford, U., University of, A., University of, M., University of North, C., University of, P., University of, R., University of Southern, C., University of Texas, M.D.A.C.C., University of, W., Van Andel Research, I., Vanderbilt, U., Washington, U., Genome Sequencing Center: Broad, I., Washington University in St, L., Genome Characterization Centers, B.C.C.A., The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, U., University of Southern California Epigenome, C., Genome Data Analysis Centers: Broad, I., Brown, U., University of California Santa, C., Biospecimen Core Resource: International Genomics, C., The Research Institute at Nationwide Children's, H., Tissue Source Sites: Analytic Biologic, S., Asan Medical, C., Asterand, B., Barretos Cancer, H., BioreclamationIvt, Botkin Municipal, C., Chonnam National University Medical, S., Christiana Care Health, S. and Cureline. Integrated genomic characterization of oesophageal carcinoma. *Nature.* 2017, **541**(7636), pp.169-175.

144. Garman, K.S., Orlando, R.C. and Chen, X. Review: Experimental models for Barrett's esophagus and esophageal adenocarcinoma. *Am J Physiol Gastrointest Liver Physiol.* 2012, **302**(11), pp.G1231-1243.

145. Anand, A., Fang, H.-Y., Mohammad-Shahi, D., Ingermann, J., Baumeister, T., Strangmann, J., Schmid, R.M., Wang, T.C. and Quante, M. Elimination of NF-κB signaling in Vimentin+ stromal cells attenuates tumorigenesis in a mouse model of Barrett's Esophagus. *Carcinogenesis.* 2020, **42**(3), pp.405-413.

146. Münch, N.S., Fang, H.Y., Ingermann, J., Maurer, H.C., Anand, A., Kellner, V., Sahm, V., Wiethaler, M., Baumeister, T., Wein, F., Einwächter, H., Bolze, F., Klingenspor, M., Haller, D., Kavanagh, M., Lysaght, J., Friedman, R., Dannenberg, A.J., Pollak, M., Holt, P.R., Muthupalani, S., Fox, J.G., Whary, M.T., Lee, Y., Ren, T.Y., Elliot, R., Fitzgerald, R., Steiger, K., Schmid, R.M., Wang, T.C. and Quante, M. High-Fat Diet Accelerates Carcinogenesis in a Mouse Model of Barrett's Esophagus via Interleukin 8 and Alterations to the Gut Microbiome. *Gastroenterology.* 2019, **157**(2), pp.492-506.e492.

147. Ballout, F., Lu, H., Chen, L., Sriramajayam, K., Que, J., Meng, Z., Wang, T.C., Giordano, S., Zaika, A., McDonald, O., Peng, D. and El-Rifai, W. APE1 redox function is required for activation of Yes-associated protein 1 under reflux conditions in Barrett's-associated esophageal adenocarcinomas. *J Exp Clin Cancer Res.* 2022, **41**(1), p.264.
148. Kong, J., Sai, H., Crissey, M.A., Jhala, N., Falk, G.W., Ginsberg, G.G., Abrams, J.A., Nakagawa, H., Wang, K., Rustgi, A.K., Wang, T.C. and Lynch, J.P. Immature myeloid progenitors promote disease progression in a mouse model of Barrett's-like metaplasia. *Oncotarget.* 2015, **6**(32), pp.32980-33005.
149. Sahm, V., Maurer, C., Baumeister, T., Anand, A., Strangmann, J., Schmid, R.M., Wang, T.C. and Quante, M. Telomere shortening accelerates tumor initiation in the L2-IL1B mouse model of Barrett esophagus and emerges as a possible biomarker. *Oncotarget.* 2022, **13**, pp.347-359.
150. Kunze, B., Middelhoff, M., Maurer, H.C., Agibalova, T., Anand, A., Bührer, A.M., Fang, H.Y., Baumeister, T., Steiger, K., Strangmann, J., Schmid, R.M., Wang, T.C. and Quante, M. Notch signaling drives development of Barrett's metaplasia from Dclk1-positive epithelial tuft cells in the murine gastric mucosa. *Sci Rep.* 2021, **11**(1), p.4509.
151. Lee, Y., Urbanska, A.M., Hayakawa, Y., Wang, H., Au, A.S., Luna, A.M., Chang, W., Jin, G., Bhagat, G., Abrams, J.A., Friedman, R.A., Varro, A., Wang, K.K., Boyce, M., Rustgi, A.K., Sepulveda, A.R., Quante, M. and Wang, T.C. Gastrin stimulates a cholecystokinin-2-receptor-expressing cardia progenitor cell and promotes progression of Barrett's-like esophagus. *Oncotarget.* 2017, **8**(1), pp.203-214.
152. Ibrahim, H.M., AbdElbary, A.M., Mohamed, S.Y., Elwan, A., Abdelhamid, M.I. and Ibrahim, A. Prognostic Value of Cyclin D1 and CD44 Expression in Gastric Adenocarcinoma. *J Gastrointest Cancer.* 2019, **50**(3), pp.370-379.
153. Wang, J.S., Varro, A., Lightdale, C.J., Lertkowit, N., Slack, K.N., Fingerhood, M.L., Tsai, W.Y., Wang, T.C. and Abrams, J.A. Elevated serum gastrin is associated with a history of advanced neoplasia in Barrett's esophagus. *Am J Gastroenterol.* 2010, **105**(5), pp.1039-1045.
154. Muthupalani, S., Annamalai, D., Feng, Y., Ganesan, S.M., Ge, Z., Whary, M.T., Nakagawa, H., Rustgi, A.K., Wang, T.C. and Fox, J.G. IL-1β transgenic mouse model of inflammation driven esophageal and oral squamous cell carcinoma. *Scientific Reports.* 2023, **13**(1), p.12732.
155. Gotovac, J.R., Kader, T., Milne, J.V., Fujihara, K.M., Lara-Gonzalez, L.E., Gorringe, K.L., Kalimuthu, S.N., Jayawardana, M.W., Duong, C.P., Phillips, W.A. and Clemons, N.J. Loss of SMAD4 Is Sufficient to Promote Tumorigenesis in a Model of Dysplastic Barrett's Esophagus. *Cell Mol Gastroenterol Hepatol.* 2021, **12**(2), pp.689-713.
156. Blum, A.E., Venkitachalam, S., Ravillah, D., Chelluboyina, A.K., Kieber-Emmons, A.M., Ravi, L., Kresak, A., Chandar, A.K., Markowitz, S.D., Canto, M.I., Wang, J.S., Shaheen, N.J., Guo, Y., Shyr, Y., Willis, J.E., Chak, A., Varadan, V. and Guda, K. Systems Biology Analyses Show Hyperactivation of Transforming Growth Factor-β and JNK Signaling Pathways in Esophageal Cancer. *Gastroenterology.* 2019, **156**(6), pp.1761-1774.
157. Venkitachalam, S., Babu, D., Ravillah, D., Katabathula, R.M., Joseph, P., Singh, S., Udhayakumar, B., Miao, Y., Martinez-Uribe, O., Hogue, J.A., Kresak, A.M., Dawson, D., LaFramboise, T., Willis, J.E., Chak, A., Garman, K.S., Blum, A.E., Varadan, V. and Guda, K. The Ephrin B2 Receptor Tyrosine Kinase Is a Regulator of Proto-oncogene MYC and Molecular Programs Central to Barrett's Neoplasia. *Gastroenterology.* 2022, **163**(5), pp.1228-1241.

158. Liu, Y., Wu, W., Cai, C., Zhang, H., Shen, H. and Han, Y. Patient-derived xenograft models in cancer therapy: technologies and applications. *Signal Transduct Target Ther.* 2023, **8**(1), p.160.
159. Okada, S., Vaeteewoottacharn, K. and Kariya, R. Application of Highly Immunocompromised Mice for the Establishment of Patient-Derived Xenograft (PDX) Models. *Cells.* 2019, **8**(8).
160. Sun, D., Wang, X., Gai, Z., Song, X., Jia, X. and Tian, H. Bile acids but not acidic acids induce Barrett's esophagus. *Int J Clin Exp Pathol.* 2015, **8**(2), pp.1384-1392.
161. Wen, H., Liu, T., Liu, H., Teng, J.H. and Li, S.B. An improved surgical procedure to establish a gastroesophageal reflux model with a high incidence of Barrett's esophagus in rats. *Exp Ther Med.* 2018, **16**(5), pp.3863-3868.
162. Matsui, D., Omstead, A.N., Kosovec, J.E., Komatsu, Y., Lloyd, E.J., Raphael, H., Kelly, R.J., Zaidi, A.H. and Jobe, B.A. High yield reproducible rat model recapitulating human Barrett's carcinogenesis. *World J Gastroenterol.* 2017, **23**(33), pp.6077-6087.
163. Agoston, A.T., Pham, T.H., Odze, R.D., Wang, D.H., Das, K.M., Spechler, S.J. and Souza, R.F. Columnar-Lined Esophagus Develops via Wound Repair in a Surgical Model of Reflux Esophagitis. *Cell Mol Gastroenterol Hepatol.* 2018, **6**(4), pp.389-404.
164. Gillen, P., Keeling, P., Byrne, P.J., West, A.B. and Hennessy, T.P. Experimental columnar metaplasia in the canine oesophagus. *Br J Surg.* 1988, **75**(2), pp.113-115.
165. Li, H., Walsh, T.N., O'Dowd, G., Gillen, P., Byrne, P.J. and Hennessy, T.P. Mechanisms of columnar metaplasia and squamous regeneration in experimental Barrett's esophagus. *Surgery.* 1994, **115**(2), pp.176-181.
166. Sultan, F. and Ganaie, B.A. Comparative oncology: Integrating human and veterinary medicine. *Open Vet J.* 2018, **8**(1), pp.25-34.
167. Lee, S., Park, S., Kim, M., Hwang, S. and Kim, H.Y. Transhiatal esophagogastric anastomosis and postoperative monitoring of thoracic esophageal leiomyosarcoma in a dog. *Can Vet J.* 2020, **61**(4), pp.401-406.
168. Garman, K.S., Kruger, L., Thomas, S., Swiderska-Syn, M., Moser, B.K., Diehl, A.M. and McCall, S.J. Ductal metaplasia in oesophageal submucosal glands is associated with inflammation and oesophageal adenocarcinoma. *Histopathology.* 2015, **67**(6), pp.771-782.
169. Abdulnour-Nakhoul, S., Nakhoul, N.L., Wheeler, S.A., Haque, S., Wang, P., Brown, K., Orlando, G. and Orlando, R.C. Characterization of esophageal submucosal glands in pig tissue and cultures. *Dig Dis Sci.* 2007, **52**(11), pp.3054-3065.
170. Garman, K.S., Purkayastha, B.P.D., Hogue, J.A., Fecteau, R., Guda, K. and Chak, A. Genetic Defect in Submucosal Gland-Associated Caveolin-3: A New Paradigm in Esophageal Adenocarcinoma Risk. *Gastroenterology.* 2023, **165**(6), pp.1561-1564.e1563.
171. Zhou, F.Y., Ruiz-Puig, C., Owen, R.P., White, M.J., Rittscher, J. and Lu, X. Motion sensing superpixels (MOSES) is a systematic computational framework to quantify and discover cellular motion phenotypes. *eLife.* 2019, **8**, p.e40162.
172. Kalabis, J., Wong, G.S., Vega, M.E., Natsuizaka, M., Robertson, E.S., Herlyn, M., Nakagawa, H. and Rustgi, A.K. Isolation and characterization of mouse and human esophageal epithelial cells in 3D organotypic culture. *Nat Protoc.* 2012, **7**(2), pp.235-246.
173. Kosoff, R.E., Gardiner, K.L., Merlo, L.M., Pavlov, K., Rustgi, A.K. and Maley, C.C. Development and characterization of an organotypic model of Barrett's esophagus. *J Cell Physiol.* 2012, **227**(6), pp.2654-2659.

174. Roudebush, C., Catala-Valentin, A., Andl, T., Le Bras, G.F. and Andl, C.D. Activin A-mediated epithelial de-differentiation contributes to injury repair in an in vitro gastrointestinal reflux model. *Cytokine.* 2019, **123**, p.154782.
175. Sato, T., Stange, D.E., Ferrante, M., Vries, R.G., Van Es, J.H., Van den Brink, S., Van Houdt, W.J., Pronk, A., Van Gorp, J., Siersema, P.D. and Clevers, H. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology.* 2011, **141**(5), pp.1762-1772.
176. Derouet, M.F., Allen, J., Wilson, G.W., Ng, C., Radulovich, N., Kalimuthu, S., Tsao, M.S., Darling, G.E. and Yeung, J.C. Towards personalized induction therapy for esophageal adenocarcinoma: organoids derived from endoscopic biopsy recapitulate the pre-treatment tumor. *Sci Rep.* 2020, **10**(1), p.14514.
177. Kim, H., Jang, B., Zhang, C., Caldwell, B., Park, D.J., Kong, S.H., Lee, H.J., Yang, H.K., Goldenring, J.R. and Choi, E. Targeting Stem Cells and Dysplastic Features With Dual MEK/ERK and STAT3 Suppression in Gastric Carcinogenesis. *Gastroenterology.* 2024, **166**(1), pp.117-131.
178. Whelan, K.A., Muir, A.B. and Nakagawa, H. Esophageal 3D Culture Systems as Modeling Tools in Esophageal Epithelial Pathobiology and Personalized Medicine. *Cell Mol Gastroenterol Hepatol.* 2018, **5**(4), pp.461-478.
179. Liu, X., Cheng, Y., Abraham, J.M., Wang, Z., Wang, Z., Ke, X., Yan, R., Shin, E.J., Ngamruengphong, S., Khashab, M.A., Zhang, G., McNamara, G., Ewald, A.J., Lin, D., Liu, Z. and Meltzer, S.J. Modeling Wnt signaling by CRISPR-Cas9 genome editing recapitulates neoplasia in human Barrett epithelial organoids. *Cancer Lett.* 2018, **436**, pp.109-118.
180. Sontheimer-Phelps, A., Chou, D.B., Tovaglieri, A., Ferrante, T.C., Duckworth, T., Fadel, C., Frismantas, V., Sutherland, A.D., Jalili-Firoozinezhad, S., Kasendra, M., Stas, E., Weaver, J.C., Richmond, C.A., Levy, O., Prantil-Baun, R., Breault, D.T. and Ingber, D.E. Human Colon-on-a-Chip Enables Continuous In Vitro Analysis of Colon Mucus Layer Accumulation and Physiology. *Cell Mol Gastroenterol Hepatol.* 2020, **9**(3), pp.507-526.
181. Pimenta, J., Ribeiro, R., Almeida, R., Costa, P.F., da Silva, M.A. and Pereira, B. Organ-on-Chip Approaches for Intestinal 3D In Vitro Modeling. *Cell Mol Gastroenterol Hepatol.* 2022, **13**(2), pp.351-367.
182. Owen, R.P., White, M.J., Severson, D.T., Braden, B., Bailey, A., Goldin, R., Wang, L.M., Ruiz-Puig, C., Maynard, N.D., Green, A., Piazza, P., Buck, D., Middleton, M.R., Ponting, C.P., Schuster-Böckler, B. and Lu, X. Single cell RNA-seq reveals profound transcriptional similarity between Barrett's oesophagus and oesophageal submucosal glands. *Nat Commun.* 2018, **9**(1), p.4261.
183. Korbut, E., Janmaat, V.T., Wierdak, M., Hankus, J., Wójcik, D., Surmiak, M., Magierowska, K., Brzozowski, T., Peppelenbosch, M.P. and Magierowski, M. Molecular Profile of Barrett's Esophagus and Gastroesophageal Reflux Disease in the Development of Translational Physiological and Pharmacological Studies. *Int J Mol Sci.* 2020, **21**(17).
184. Zhang, Q., Agoston, A.T., Pham, T.H., Zhang, W., Zhang, X., Huo, X., Peng, S., Bajpai, M., Das, K., Odze, R.D., Spechler, S.J. and Souza, R.F. Acidic Bile Salts Induce Epithelial to Mesenchymal Transition via VEGF Signaling in Non-Neoplastic Barrett's Cells. *Gastroenterology.* 2019, **156**(1), pp.130-144.e110.
185. Matsuzaki, J., Suzuki, H., Tsugawa, H., Watanabe, M., Hossain, S., Arai, E., Saito, Y., Sekine, S., Akaike, T., Kanai, Y., Mukaisho, K., Auwerx, J. and Hibi, T. Bile acids increase levels of microRNAs 221 and 222, leading to degradation of CDX2 during esophageal carcinogenesis. *Gastroenterology.* 2013, **145**(6), pp.1300-1311.

186. Ghatak, S., Reveiller, M., Toia, L., Ivanov, A., Godfrey, T.E. and Peters, J.H. Bile acid at low pH reduces squamous differentiation and activates EGFR signaling in esophageal squamous cells in 3-D culture. *J Gastrointest Surg.* 2013, **17**(10), pp.1723-1731.
187. Zhou, Z., Lu, H., Zhu, S., Gomaa, A., Chen, Z., Yan, J., Washington, K., El-Rifai, W., Dang, C. and Peng, D. Activation of EGFR-DNA-PKcs pathway by IGFBP2 protects esophageal adenocarcinoma cells from acidic bile salts-induced DNA damage. *J Exp Clin Cancer Res.* 2019, **38**(1), p.13.
188. Stachler, M.D., Camarda, N.D., Deitrick, C., Kim, A., Agoston, A.T., Odze, R.D., Hornick, J.L., Nag, A., Thorner, A.R., Ducar, M., Noffsinger, A., Lash, R.H., Redston, M., Carter, S.L., Davison, J.M. and Bass, A.J. Detection of Mutations in Barrett's Esophagus Before Progression to High-Grade Dysplasia or Adenocarcinoma. *Gastroenterology.* 2018, **155**(1), pp.156-167.
189. Stachler, M.D., Taylor-Weiner, A., Peng, S., McKenna, A., Agoston, A.T., Odze, R.D., Davison, J.M., Nason, K.S., Loda, M., Leshchiner, I., Stewart, C., Stojanov, P., Seepo, S., Lawrence, M.S., Ferrer-Torres, D., Lin, J., Chang, A.C., Gabriel, S.B., Lander, E.S., Beer, D.G., Getz, G., Carter, S.L. and Bass, A.J. Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet.* 2015, **47**(9), pp.1047-1055.
190. Visrodia, K., Singh, S., Krishnamoorthi, R., Ahlquist, D.A., Wang, K.K., Iyer, P.G. and Katzka, D.A. Magnitude of Missed Esophageal Adenocarcinoma After Barrett's Esophagus Diagnosis: A Systematic Review and Meta-analysis. *Gastroenterology.* 2016, **150**(3), pp.599-607.e597; quiz e514-595.
191. Gier, R.A., Hueros, R.A.R., Rong, J., DeMarshall, M., Karakasheva, T.A., Muir, A.B., Falk, G.W., Zhang, N.R. and Shaffer, S.M. Clonal cell states link Barrett's esophagus and esophageal adenocarcinoma. 2023, p.2023.2001.2026.525564.
192. Schmidt, M., Hackett, R.J., Baker, A.M., McDonald, S.A.C., Quante, M. and Graham, T.A. Evolutionary dynamics in Barrett oesophagus: implications for surveillance, risk stratification and therapy. *Nat Rev Gastroenterol Hepatol.* 2022, **19**(2), pp.95-111.
193. Sjostrom, S.L., Bai, Y., Huang, M., Liu, Z., Nielsen, J., Joensson, H.N. and Svahn, H.A.J.L.o.a.C. High-throughput screening for industrial enzyme production hosts by droplet microfluidics. 2014, **14**(4), pp.806-813.
194. Scanlon, T.C., Dostal, S.M., Griswold, K.E.J.B. and bioengineering. A high-throughput screen for antibiotic drug discovery. 2014, **111**(2), pp.232-243.
195. Sauzade, M. and Brouzes, E.J.L.o.a.C. Deterministic trapping, encapsulation and retrieval of single-cells. 2017, **17**(13), pp.2186-2192.
196. Hosokawa, M., Hoshino, Y., Nishikawa, Y., Hirose, T., Yoon, D.H., Mori, T., Sekiguchi, T., Shoji, S. and Takeyama, H. Droplet-based microfluidics for high-throughput screening of a metagenomic library for isolation of microbial enzymes. *Biosensors and Bioelectronics.* 2015, **67**, pp.379-385.
197. Niu, X., Gielen, F., Edel, J.B. and Demello, A.J.J.N.c. A microdroplet dilutor for high-throughput screening. 2011, **3**(6), pp.437-442.
198. Zeng, S., Liu, X., Xie, H., Lin, B.J.M.T. and Applications. Basic technologies for droplet microfluidics. 2011, pp.69-90.
199. Köster, S., Angile, F.E., Duan, H., Agresti, J.J., Wintner, A., Schmitz, C., Rowat, A.C., Merten, C.A., Pisignano, D. and Griffiths, A.D.J.L.o.a.C. Drop-based microfluidic devices for encapsulation of single cells. 2008, **8**(7), pp.1110-1115.
200. Chabert, M. and Viovy, J.-L.J.P.o.t.N.A.o.S. Microfluidic high-throughput encapsulation and hydrodynamic self-sorting of single cells. 2008, **105**(9), pp.3191-3196.

201. Granieri, L., Baret, J.-C., Griffiths, A.D., Merten, C.A.J.C. and biology. High-throughput screening of enzymes by retroviral display using droplet-based microfluidics. 2010, **17**(3), pp.229-235.
202. Baret, J.-C., Miller, O.J., Taly, V., Ryckelynck, M., El-Harrak, A., Frenz, L., Rick, C., Samuels, M.L., Hutchison, J.B. and Agresti, J.J.J.L.o.a.C. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. 2009, **9**(13), pp.1850-1858.
203. Collins, D.J., Alan, T., Helmerson, K. and Neild, A.J.L.o.a.C. Surface acoustic waves for on-demand production of picoliter droplets and particle encapsulation. 2013, **13**(16), pp.3225-3231.
204. Du, W.-B., Sun, M., Gu, S.-Q., Zhu, Y. and Fang, Q.J.A.c. Automated microfluidic screening assay platform based on DropLab. 2010, **82**(23), pp.9941-9947.
205. Sun, M. and Fang, Q.J.L.o.a.C. High-throughput sample introduction for droplet-based screening with an on-chip integrated sampling probe and slotted-vial array. 2010, **10**(21), pp.2864-2868.
206. Huang, M., Bai, Y., Sjostrom, S.L., Hallström, B.M., Liu, Z., Petranovic, D., Uhlén, M., Joensson, H.N., Andersson-Svahn, H. and Nielsen, J.J.P.o.t.N.A.o.S. Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. 2015, **112**(34), pp.E4689-E4696.
207. Demirci, U. and Montesano, G.J.L.o.a.C. Single cell epitaxy by acoustic picolitre droplets. 2007, **7**(9), pp.1139-1145.
208. He, M., Edgar, J.S., Jeffries, G.D., Lorenz, R.M., Shelby, J.P. and Chiu, D.T.J.A.c. Selective encapsulation of single cells and subcellular organelles into picoliter-and femtoliter-volume droplets. 2005, **77**(6), pp.1539-1544.
209. Webster, J.A., Wuethrich, A., Shanmugasundaram, K.B., Richards, R.S., Zelek, W.M., Shah, A.K., Gordon, L.G., Kendall, B.J., Hartel, G. and Morgan, B.P.J.C. Development of EndoScreen Chip, a microfluidic pre-endoscopy triage test for esophageal adenocarcinoma. 2021, **13**(12), p.2865.
210. Shimshoni, E., Merry, G., Milot, Z., Oh, C., Horvath, V., Gould, R., Caruso, J., Chen-Tanyolac, C., Gascard, P. and Sangwan, V.J.G.H.A. Epithelial-stromal interactions in barrett's esophagus modeled in human organ chips. 2023, **2**(5), p.676.
211. Ohnaga, T., Shimada, Y., Takata, K., Obata, T., Okumura, T., Nagata, T., Kishi, H., Muraguchi, A. and Tsukada, K. Capture of esophageal and breast cancer cells with polymeric microfluidic devices for CTC isolation. *Mol Clin Oncol.* 2016, **4**(4), pp.599-602.
212. Sweedler, J.V., Arriaga, E.A.J.A. and chemistry, b. *Single cell analysis*. Springer. 2007, 387, pp.1-2.
213. Li, M., Xu, J., Romero-Gonzalez, M., Banwart, S.A. and Huang, W.E.J.C.o.i.b. Single cell Raman spectroscopy for cell sorting and imaging. 2012, **23**(1), pp.56-63.
214. Kobayashi, H., Lei, C., Wu, Y., Mao, A., Jiang, Y., Guo, B., Ozeki, Y. and Goda, K.J.S.r. Label-free detection of cellular drug responses by high-throughput bright-field imaging and machine learning. 2017, **7**(1), p.12454.
215. Schie, I.W., Rüger, J., Mondol, A.S., Ramoji, A., Neugebauer, U., Krafft, C. and Popp, J.r.J.A.c. High-throughput screening Raman spectroscopy platform for label-free cellomics. 2018, **90**(3), pp.2023-2030.
216. Charon, L. and Lilge, L. Single-Cell Analysis in Microfluidic Devices. In: Li, D. ed. *Encyclopedia of Microfluidics and Nanofluidics*. New York, NY: Springer New York, 2015, pp.3017-3027.
217. Smekal, A.J.N. Zur quantentheorie der dispersion. 1923, **11**(43), pp.873-875.

218. Raman, C.V. and Krishnan, K.S.J.N. A new type of secondary radiation. 1928, **121**(3048), pp.501-502.
219. Mallidis, C., Sanchez, V., Wistuba, J., Wuebbeling, F., Burger, M., Fallnich, C. and Schlatt, S.J.H.r.u. Raman microspectroscopy: shining a new light on reproductive medicine. 2014, **20**(3), pp.403-414.
220. Popp, J. The many facets of Raman spectroscopy. In: *Asia Communications and Photonics Conference*: Optica Publishing Group, 2012, p.ATh3E. 3.
221. Webster, S. *Raman Microscopy and Optical Spectroscopy of Conjugated Polymers at High Pressure*. thesis, University of Leeds (Department of Physics), 1994.
222. Baldwin, K.J. *Photophysics of Conjugated Polymers and Oligomers*. thesis, University of Leeds (Department of Physics and Astronomy), 1997.
223. Krafft, C., Dietzek, B., Schmitt, M. and Popp, J.J.J.o.b.o. Raman and coherent anti-Stokes Raman scattering microspectroscopy for biomedical applications. 2012, **17**(4), pp.040801-040801.
224. Matthäus, C., Bird, B., Miljković, M., Chernenko, T., Romeo, M. and Diem, M.J.M.i.c.b. Infrared and Raman microscopy in cell biology. 2008, **89**, pp.275-308.
225. Palonpon, A.F., Sodeoka, M. and Fujita, K. Molecular imaging of live cells by Raman microscopy. *Current Opinion in Chemical Biology.* 2013, **17**(4), pp.708-715.
226. Kong, K., Kendall, C., Stone, N. and Notingher, I.J.A.d.d.r. Raman spectroscopy for medical diagnostics—From in-vitro biofluid assays to in-vivo cancer detection. 2015, **89**, pp.121-134.
227. Smith, R., Wright, K.L. and Ashton, L.J.A. Raman spectroscopy: an evolving technique for live cell studies. 2016, **141**(12), pp.3590-3600.
228. Chrimes, A.F., Khoshmanesh, K., Stoddart, P.R., Mitchell, A. and Kalantar-Zadeh, K.J.C.S.R. Microfluidics and Raman microscopy: current applications and future challenges. 2013, **42**(13), pp.5880-5906.
229. Zheng, Y., Nguyen, J., Wei, Y. and Sun, Y.J.L.o.a.C. Recent advances in microfluidic techniques for single-cell biophysical characterization. 2013, **13**(13), pp.2464-2483.
230. Heath, J.R., Ribas, A. and Mischel, P.S.J.N.r.D.d. Single-cell analysis tools for drug discovery and development. 2016, **15**(3), pp.204-216.
231. Zhang, Q., Zhang, P., Gou, H., Mou, C., Huang, W.E., Yang, M., Xu, J. and Ma, B.J.A. Towards high-throughput microfluidic Raman-activated cell sorting. 2015, **140**(18), pp.6163-6174.
232. Bawa, R., Chang, E.H., Audette, G.F., Diwan, A. and Faiz, S.A. *Advances in Medical Biochemistry, Genomics, Physiology, and Pathology.* Jenny Stanford Publishing, 2021.
233. Mayo, D.W., Miller, F.A. and Hannah, R.W. *Course notes on the interpretation of infrared and Raman spectra.* John Wiley & Sons, 2004.
234. Banwell, C.N. and McCash, E.M. *Fundamentals of molecular spectroscopy.* Indian Edition, 2017.
235. Wen, Z.Q.J.J.o.p.s. Raman spectroscopy of protein pharmaceuticals. 2007, **96**(11), pp.2861-2878.
236. Krishna, C.M., Sockalingum, G., Kurien, J., Rao, L., Venteo, L., Pluot, M., Manfait, M. and Kartha, V.J.A.s. Micro-Raman spectroscopy for optical pathology of oral squamous cell carcinoma. 2004, **58**(9), pp.1128-1135.
237. Abramczyk, H., Brozek-Pluska, B., Surmacki, J., Jablonska, J. and Kordek, R. The label-free Raman imaging of human breast cancer. *Journal of Molecular Liquids - J MOL LIQ.* 2011, **164**, pp.123-131.
238. Movasaghi, Z., Rehman, S. and Rehman, I.U.J.A.S.R. Raman spectroscopy of biological tissues. 2007, **42**(5), pp.493-541.

239. Gautam, R., Vanga, S., Ariese, F., Umapathy, S.J.E.T. and Instrumentation. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. 2015, **2**, pp.1-38.
240. Diem, M., Miljković, M., Bird, B., Chernenko, T., Schubert, J., Marcsisin, E., Mazur, A., Kingston, E., Zuser, E. and Papamarkakis, K.J.J.o.S. Applications of infrared and Raman microspectroscopy of cells and tissue in medical diagnostics: present status and future promises. 2012, **27**(5-6), pp.463-496.
241. El-Mashtoly, S.F., Yosef, H.K., Petersen, D., Mavarani, L., Maghnouj, A., Hahn, S., Kötting, C. and Gerwert, K.J.A.c. Label-free Raman spectroscopic imaging monitors the integral physiologically relevant drug responses in cancer cells. 2015, **87**(14), pp.7297-7304.
242. Kurouski, D., Van Duyne, R.P. and Lednev, I.K.J.A. Exploring the structure and formation mechanism of amyloid fibrils by Raman spectroscopy: a review. 2015, **140**(15), pp.4967-4980.
243. Maiti, N.C., Apetri, M.M., Zagorski, M.G., Carey, P.R. and Anderson, V.E.J.J.o.t.A.C.S. Raman spectroscopic characterization of secondary structure in natively unfolded proteins: α-synuclein. 2004, **126**(8), pp.2399-2408.
244. Rygula, A., Majzner, K., Marzec, K.M., Kaczor, A., Pilarczyk, M. and Baranska, M. Raman spectroscopy of proteins: a review. 2013, **44**(8), pp.1061-1076.
245. David, C.J.H.S.W. Raman Spectroscopy for proteins. 2012, pp.1-53.
246. Talari, A.C.S., Movasaghi, Z., Rehman, S. and Rehman, I.u. Raman Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews.* 2015, **50**(1), pp.46-111.
247. Swain, R.J., Jell, G. and Stevens, M.M.J.J.o.c.b. Non-invasive analysis of cell cycle dynamics in single living cells with Raman micro-spectroscopy. 2008, **104**(4), pp.1427-1438.
248. Gala de Pablo, J., Armistead, F.J., Peyman, S.A., Bonthron, D., Lones, M., Smith, S. and Evans, S.D. Biochemical fingerprint of colorectal cancer cell lines using label-free live single-cell Raman spectroscopy. 2018, **49**(8), pp.1323-1332.
249. Oshima, Y., Shinzawa, H., Takenaka, T., Furihata, C. and Sato, H. Discrimination Analysis of Human Lung Cancer Cells Associated with Histological Type and Malignancy Using Raman Spectroscopy. *Journal of biomedical optics.* 2010, **15**, p.017009.
250. Pully, V., Lenferink, A.T. and Otto, C.J.J.o.R.s. Time-lapse Raman imaging of single live lymphocytes. 2011, **42**(2), pp.167-173.
251. Harvey, T.J., Faria, E.C., Henderson, A., Gazi, E., Ward, A.D., Clarke, N.W., Brown, M.D., Snook, R.D. and Gardner, P.J.J.o.B.O. Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers. 2008, **13**(6), pp.064004-064004-064004-064012.
252. Casabella, S., Scully, P., Goddard, N. and Gardner, P.J.A. Automated analysis of single cells using Laser Tweezers Raman Spectroscopy. 2016, **141**(2), pp.689-696.
253. Haka, A.S., Volynskaya, Z., Gardecki, J.A., Nazemi, J., Shenk, R., Wang, N., Dasari, R.R., Fitzmaurice, M. and Feld, M.S.J.J.o.b.o. Diagnosing breast cancer using Raman spectroscopy: prospective analysis. 2009, **14**(5), pp.054023-054023-054028.
254. Downes, A. and Elfick, A.J.S. Raman spectroscopy and related techniques in biomedicine. 2010, **10**(3), pp.1871-1889.
255. Kamemoto, L.E., Misra, A.K., Sharma, S.K., Goodman, M.T., Luk, H., Dykes, A.C. and Acosta, T.J.A.s. Near-infrared micro-Raman spectroscopy for in vitro detection of cervical cancer. 2010, **64**(3), pp.255-261.
256. Carey, P. *Biochemical applications of Raman and resonance Raman spectroscopes.* Elsevier, 2012.

257. Pallaoro, A., Hoonejani, M.R., Braun, G.B., Meinhart, C.D. and Moskovits, M.J.A.N. Rapid identification by surface-enhanced Raman spectroscopy of cancer cells at low concentrations flowing in a microfluidic channel. 2015, **9**(4), pp.4328-4336.
258. Yan, B., Li, B., Wen, Z., Luo, X., Xue, L. and Li, L.J.B.c. Label-free blood serum detection by using surface-enhanced Raman spectroscopy and support vector machine for the preoperative diagnosis of parotid gland tumors. 2015, **15**, pp.1-9.
259. Klein, K., Gigler, A.M., Aschenbrenner, T., Monetti, R., Bunk, W., Jamitzky, F., Morfill, G., Stark, R.W. and Schlegel, J.J.B.j. Label-free live-cell imaging with confocal Raman microscopy. 2012, **102**(2), pp.360-368.
260. Smith, Z.J., Lee, C., Rojalin, T., Carney, R.P., Hazari, S., Knudson, A., Lam, K., Saari, H., Ibañez, E.L. and Viitala, T.J.J.o.e.v. Single exosome study reveals subpopulations distributed among cell lines with variability related to membrane content. 2015, **4**(1), p.28533.
261. Talari, A., Evans, C., Holen, I., Coleman, R. and Rehman, I.U.J.J.o.R.S. Raman spectroscopic analysis differentiates between breast cancer cell lines. 2015, **46**(5), pp.421-427.
262. Abramczyk, H., Brozek-Pluska, B., Surmacki, J., Jablonska-Gajewicz, J. and Kordek, R. Raman 'optical biopsy' of human breast cancer. *Progress in Biophysics and Molecular Biology.* 2012, **108**, pp.74–81.
263. Neugebauer, U., Bocklitz, T., Clement, J.H., Krafft, C. and Popp, J. Towards detection and identification of circulating tumour cells using Raman spectroscopy. *Analyst.* 2010, **135**(12), pp.3178-3182.
264. Larraona-Puy, M., Ghita, A., Zoladek, A., Perkins, W., Varma, S., Leach, I.H., Koloydenko, A.A., Williams, H. and Notingher, I.J.J.o.b.o. Development of Raman microspectroscopy for automated detection and imaging of basal cell carcinoma. 2009, **14**(5), pp.054031-054031-054010.
265. Canetta, E., Riches, A., Borger, E., Herrington, S., Dholakia, K. and Adya, A.K.J.A.b. Discrimination of bladder cancer cells from normal urothelial cells with high specificity and sensitivity: combined application of atomic force microscopy and modulated Raman spectroscopy. 2014, **10**(5), pp.2043-2055.
266. Maitra, I., Morais, C.L.M., Lima, K.M.G., Ashton, K.M., Date, R.S. and Martin, F.L. Raman spectral discrimination in human liquid biopsies of oesophageal transformation to adenocarcinoma. *J Biophotonics.* 2020, **13**(3), p.e201960132.
267. Maitra, I., Morais, C.L.M., Lima, K.M.G., Ashton, K.M., Bury, D., Date, R.S. and Martin, F.L. Establishing spectrochemical changes in the natural history of oesophageal adenocarcinoma from tissue Raman mapping analysis. *Anal Bioanal Chem.* 2020, **412**(17), pp.4077-4087.
268. Hao, J., Chen, C., Jin, H., Chen, N., Zhou, J., Zhu, Y., Chung, K. and Pu, Q. The efficacy of Raman spectroscopy in the diagnosis of esophageal cancer: a systematic review and meta-analysis. *Transl Cancer Res.* 2020, **9**(8), pp.4750-4761.
269. ur Rehman, I., Movasaghi, Z. and Rehman, S. *Vibrational Spectroscopy for Tissue Analysis.* Taylor & Francis, 2012.
270. Huang, C., Wang, Q., Yao, H.L., Wang, G. and Li, Y.Q. Determination of red cell by Raman spectroscopy based on microfluidic chip and optical tweezers. *Fenxi Huaxue/ Chinese Journal of Analytical Chemistry.* 2007, **35**, pp.1410-1414.
271. Lau, A.Y., Lee, L.P. and Chan, J.W.J.L.o.a.C. An integrated optofluidic platform for Raman-activated cell sorting. 2008, **8**(7), pp.1116-1120.
272. Dochow, S., Becker, M., Spittel, R., Beleites, C., Stanca, S., Latka, I., Schuster, K., Kobelke, J., Unger, S. and Henkel, T.J.L.o.a.C. Raman-on-chip device and detection

fibres with fibre Bragg grating for analysis of solutions and particles. 2013, **13**(6), pp.1109-1113.
273. Krafft, C., Beleites, C., Schie, I., Clement, J. and Popp, J. *Raman-based identification of circulating tumor cells for cancer diagnosis.* 2016.
274. Cao, C., Zhou, D., Chen, T., Streets, A.M. and Huang, Y.J.A.c. Label-free digital quantification of lipid droplets in single cells by stimulated Raman microscopy on a microfluidic platform. 2016, **88**(9), pp.4931-4939.
275. Ideguchi, T., Nakamura, T., Takizawa, S., Tamamitsu, M., Lee, S., Hiramatsu, K., Ramaiah-Badarla, V., Park, J.-w., Kasai, Y. and Hayakawa, T.J.O.L. Microfluidic single-particle chemical analyzer with dual-comb coherent Raman spectroscopy. 2018, **43**(16), pp.4057-4060.
276. Kosinski, J.A.J.I.j.o.h.s.e. and systems. New piezoelectric substrates for SAW devices. 2000, **10**(04), pp.1017-1068.
277. Picot, J., Guerin, C.L., Le Van Kim, C. and Boulanger, C.M.J.C. Flow cytometry: retrospective, fundamentals and recent instrumentation. 2012, **64**, pp.109-130.
278. Kaern, M., Elston, T.C., Blake, W.J. and Collins, J.J.J.N.R.G. Stochasticity in gene expression: from theories to phenotypes. 2005, **6**(6), pp.451-464.
279. Tsimring, L.S.J.R.o.P.i.P. Noise in biology. 2014, **77**(2), p.026601.
280. Zheng, Y., Nguyen, J., Wei, Y. and Sun, Y. Recent advances in microfluidic techniques for single-cell biophysical characterization. *Lab on a Chip.* 2013, **13**(13), pp.2464-2483.
281. Shelby, J.P., White, J., Ganesan, K., Rathod, P.K. and Chiu, D.T.J.P.o.t.N.A.o.S. A microfluidic model for single-cell capillary obstruction by Plasmodium falciparum-infected erythrocytes. 2003, **100**(25), pp.14618-14622.
282. Zheng, Y., Shojaei-Baghini, E., Azad, A., Wang, C. and Sun, Y.J.L.o.a.C. High-throughput biophysical measurement of human red blood cells. 2012, **12**(14), pp.2560-2567.
283. Faustino, V., Pinho, D., Yaginuma, T., Calhelha, R.C., Ferreira, I.C. and Lima, R.J.B.J. Extensional flow-based microfluidic device: deformability assessment of red blood cells in contact with tumor cells. 2014, **8**, pp.42-47.
284. Yaginuma, T., Oliveira, M.S., Lima, R., Ishikawa, T. and Yamaguchi, T.J.B. Human red blood cell behavior under homogeneous extensional flow in a hyperbolic-shaped microchannel. 2013, **7**(5).
285. Zheng, Y., Nguyen, J., Wang, C. and Sun, Y.J.L.o.a.C. Electrical measurement of red blood cell deformability on a microfluidic device. 2013, **13**(16), pp.3275-3283.
286. Lee, S.S., Yim, Y., Ahn, K.H. and Lee, S.J.J.B.m. Extensional flow-based assessment of red blood cell deformability using hyperbolic converging microchannel. 2009, **11**, pp.1021-1027.
287. Di Carlo, D., Irimia, D., Tompkins, R.G. and Toner, M.J.P.o.t.N.A.o.S. Continuous inertial focusing, ordering, and separation of particles in microchannels. 2007, **104**(48), pp.18892-18897.
288. Cha, S., Shin, T., Lee, S.S., Shim, W., Lee, G., Lee, S.J., Kim, Y. and Kim, J.M.J.A.c. Cell stretching measurement utilizing viscoelastic particle focusing. 2012, **84**(23), pp.10471-10477.
289. Dudani, J.S., Gossett, D.R., Henry, T. and Di Carlo, D.J.L.o.a.C. Pinched-flow hydrodynamic stretching of single-cells. 2013, **13**(18), pp.3728-3734.
290. Darling, E.M. and Di Carlo, D.J.A.r.o.b.e. High-throughput assessment of cellular mechanical properties. 2015, **17**(1), pp.35-62.
291. Tse, H.T., Gossett, D.R., Moon, Y.S., Masaeli, M., Sohsman, M., Ying, Y., Mislick, K., Adams, R.P., Rao, J. and Di Carlo, D.J.S.t.m. Quantitative diagnosis of malignant

pleural effusions by single-cell mechanophenotyping. 2013, **5**(212), pp.212ra163-212ra163.
292. Otto, O., Rosendahl, P., Golfier, S., Mietke, A., Herbig, M., Jacobi, A., Töpfner, N., Herold, C., Klaue, D. and Girardo, S. Real-time deformability cytometry as a label-free indicator of cell function. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*: IEEE, 2015, pp.1861-1864.
293. Mietke, A., Otto, O., Girardo, S., Rosendahl, P., Taubenberger, A., Golfier, S., Ulbricht, E., Aland, S., Guck, J. and Fischer-Friedrich, E.J.B.j. Extracting cell stiffness from real-time deformability cytometry: theory and experiment. 2015, **109**(10), pp.2023-2036.
294. Guillou, L., Dahl, J.B., Lin, J.-M.G., Barakat, A.I., Husson, J., Muller, S.J. and Kumar, S.J.B.j. Measuring cell viscoelastic properties using a microfluidic extensional flow device. 2016, **111**(9), pp.2039-2050.
295. Guck, J., Schinkinger, S., Lincoln, B., Wottawah, F., Ebert, S., Romeyke, M., Lenz, D., Erickson, H.M., Ananthakrishnan, R. and Mitchell, D.J.B.j. Optical deformability as an inherent cell marker for testing malignant transformation and metastatic competence. 2005, **88**(5), pp.3689-3698.
296. Guo, Q., Park, S. and Ma, H.J.L.o.a.C. Microfluidic micropipette aspiration for measuring the deformability of single cells. 2012, **12**(15), pp.2687-2695.
297. Wyss, H.M.J.B.J. Cell mechanics: Combining speed with precision. 2015, **109**(10), pp.1997-1998.
298. Ahmad, I.L. and Ahmad, M.R. Technological advancements in characterizing single cell's stiffness properties-A review. In: *2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES)*: IEEE, 2014, pp.187-192.
299. Purcell, E.M. Life at low Reynolds number. In: *Physics and our world: reissue of the proceedings of a symposium in honor of Victor F Weisskopf*: World Scientific, 2014, pp.47-67.
300. Beebe, D.J., Mensing, G.A. and Walker, G.M.J.A.r.o.b.e. Physics and applications of microfluidics in biology. 2002, **4**(1), pp.261-286.
301. Avila, K., Moxey, D., De Lozar, A., Avila, M., Barkley, D. and Hof, B.J.S. The onset of turbulence in pipe flow. 2011, **333**(6039), pp.192-196.
302. Asmolov, E.S.J.J.o.f.m. The inertial lift on a spherical particle in a plane Poiseuille flow at large channel Reynolds number. 1999, **381**, pp.63-87.
303. Abraham, J., Sparrow, E., Minkowycz, W.J.I.J.o.H. and Transfer, M. Internal-flow Nusselt numbers for the low-Reynolds-number end of the laminar-to-turbulent transition regime. 2011, **54**(1-3), pp.584-588.
304. Baskurt, O.K., Gelmont, D., Meiselman, H.J.J.A.j.o.r. and medicine, c.c. Red blood cell deformability in sepsis. 1998, **157**(2), pp.421-427.
305. McMillan, D.E., Utterback, N.G. and Puma, J.L.J.D. Reduced erythrocyte deformability in diabetes. 1978, **27**(9), pp.895-901.
306. Guo, Q., Reiling, S.J., Rohrbach, P. and Ma, H.J.L.o.a.C. Microfluidic biomechanical assay for red blood cells parasitized by Plasmodium falciparum. 2012, **12**(6), pp.1143-1150.
307. Stuart, J. and Nash, G.J.B.r. Red cell deformability and haematological disorders. 1990, **4**(3), pp.141-147.
308. Suresh, S. Biomechanics and biophysics of cancer cells. *Acta Biomater.* 2007, **3**(4), pp.413-438.

309. Cross, S.E., Jin, Y.-S., Rao, J. and Gimzewski, J.K. Nanomechanical analysis of cells from cancer patients. In: *Nano-enabled medical applications*. Jenny Stanford Publishing, 2020, pp.547-566.
310. Armistead, F.J., Gala De Pablo, J., Gadêlha, H., Peyman, S.A. and Evans, S.D. Physical Biomarkers of Disease Progression: On-Chip Monitoring of Changes in Mechanobiology of Colorectal Cancer Cells. *Scientific Reports.* 2020, **10**(1), p.3254.
311. Che, J., Yu, V., Dhar, M., Renier, C., Matsumoto, M., Heirich, K., Garon, E.B., Goldman, J., Rao, J. and Sledge, G.W.J.O. Classification of large circulating tumor cells isolated with ultra-high throughput microfluidic Vortex technology. 2016, **7**(11), p.12748.
312. Nguyen, A.V., Nyberg, K.D., Scott, M.B., Welsh, A.M., Nguyen, A.H., Wu, N., Hohlbauch, S.V., Geisse, N.A., Gibb, E.A. and Robertson, A.G.J.I.b. Stiffness of pancreatic cancer cells is associated with increased invasive potential. 2016, **8**(12), pp.1232-1245.
313. Khan, Z. and Vanapalli, S.J.B. Probing the mechanical properties of brain cancer cells using a microfluidic cell squeezer device. 2013, **7**(1).
314. Ahmmed, S.M., Bithi, S.S., Pore, A.A., Mubtasim, N., Schuster, C., Gollahon, L.S. and Vanapalli, S.A.J.A.b. Multi-sample deformability cytometry of cancer cells. 2018, **2**(3).
315. Xavier, M., Rosendahl, P., Herbig, M., Kräter, M., Spencer, D., Bornhäuser, M., Oreffo, R.O., Morgan, H., Guck, J. and Otto, O.J.I.B. Mechanical phenotyping of primary human skeletal stem cells in heterogeneous populations by real-time deformability cytometry. 2016, **8**(5), pp.616-623.
316. Fuhrmann, A., Staunton, J.R., Nandakumar, V., Banyai, N., Davies, P.C.W. and Ros, R. AFM stiffness nanotomography of normal, metaplastic and dysplastic human esophageal cells. *Physical Biology.* 2011, **8**(1), p.015007.
317. Chance, B., Mueller, P., De Vault, D. and Powers, L. Biological membranes. *Physics Today.* 1980, **33**(10), pp.32-38.
318. Nematbakhsh, Y. and Lim, C.T.J.A.M.S. Cell biomechanics and its applications in human disease diagnosis. 2015, **31**, pp.268-273.
319. Wegner, A. and Engel, J.J.B.c. Kinetics of the cooperative association of actin to actin filament. 1975, **3**(3), pp.215-225.
320. Wen, Q. and Janmey, P.A. Polymer physics of the cytoskeleton. *Curr Opin Solid State Mater Sci.* 2011, **15**(5), pp.177-182.
321. Clark, A.G., Dierkes, K. and Paluch, E.K. Monitoring actin cortex thickness in live cells. *Biophys J.* 2013, **105**(3), pp.570-580.
322. Petry, S. and Vale, R.D.J.N.c.b. Microtubule nucleation at the centrosome and beyond. 2015, **17**(9), pp.1089-1093.
323. Qin, Z., Buehler, M.J. and Kreplak, L.J.J.o.b. A multi-scale approach to understand the mechanobiology of intermediate filaments. 2010, **43**(1), pp.15-22.
324. Janmey, P.A., Euteneuer, U., Traub, P. and Schliwa, M.J.T.J.o.c.b. Viscoelastic properties of vimentin compared with other filamentous biopolymer networks. 1991, **113**(1), pp.155-160.
325. Carlo, D.D., Wu, L.Y. and Lee, L.P. Dynamic single cell culture array. *Lab on a Chip.* 2006, **6**(11), pp.1445-1449.
326. Tan, W.H. and Takeuchi, S. A trap-and-release integrated microfluidic system for dynamic microarray applications. *Proc Natl Acad Sci U S A.* 2007, **104**(4), pp.1146-1151.
327. Bao, P., Paterson, D.A., Harrison, P.L., Miller, K., Peyman, S., Jones, J.C., Sandoe, J., Evans, S.D., Bushby, R.J. and Gleeson, H.F.J.L.o.a.C. Lipid coated liquid crystal

droplets for the on-chip detection of antimicrobial peptides. 2019, **19**(6), pp.1082-1089.
328. Koch, M., Suhr, C., Roth, B. and Meinhardt-Wollweber, M. Iterative morphological and mollifier-based baseline correction for Raman spectra. 2017, **48**(2), pp.336-342.
329. Liu, Y., Wang, Z., Zhou, Z. and Xiong, T. Analysis and comparison of machine learning methods for blood identification using single-cell laser tweezer Raman spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy.* 2022, **277**, p.121274.
330. Kongklad, G., Chitaree, R., Taechalertpaisarn, T., Panvisavas, N. and Nuntawong, N. Discriminant Analysis PCA-LDA Assisted Surface-Enhanced Raman Spectroscopy for Direct Identification of Malaria-Infected Red Blood Cells. 2022, **5**(3), p.49.
331. Géron, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* " O'Reilly Media, Inc.", 2022.
332. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp.1251-1258.
333. He, K., Zhang, X., Ren, S. and Sun, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp.770-778.
334. Lippeveld, M., Knill, C., Ladlow, E., Fuller, A., Michaelis, L.J., Saeys, Y., Filby, A. and Peralta, D.J.C.P.A. Classification of human white blood cells using machine learning for stain-free imaging flow cytometry. 2020, **97**(3), pp.308-319.
335. Soldati, G., Del Ben, F., Brisotto, G., Biscontin, E., Bulfoni, M., Piruska, A., Steffan, A., Turetta, M. and Della Mea, V.J.A.j.o.t.r. Microfluidic droplets content classification and analysis through convolutional neural networks in a liquid biopsy workflow. 2018, **10**(12), p.4004.
336. Kuhar, N., Sil, S., Verma, T. and Umapathy, S. Challenges in application of Raman spectroscopy to biology and materials. *RSC Advances.* 2018, **8**(46), pp.25888-25908.
337. Miura, T., Takeuchi, H. and Harada, I. Characterization of individual tryptophan side chains in proteins using Raman spectroscopy and hydrogen-deuterium exchange kinetics. *Biochemistry.* 1988, **27**(1), pp.88-94.
338. Stone, N., Kendall, C., Smith, J., Crow, P. and Barr, H. Raman spectroscopy for identification of epithelial cancers. *Faraday Discuss.* 2004, **126**, pp.141-157; discussion 169-183.
339. Shipp, D.W., Sinjab, F. and Notingher, I. Raman spectroscopy: techniques and applications in the life sciences. *Advances in Optics and Photonics.* 2017, **9**(2), pp.315-428.
340. Stone, N., Kendall, C., Shepherd, N., Crow, P. and Barr, H. Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers. 2002, **33**(7), pp.564-573.
341. Contorno, S., Darienzo, R.E. and Tannenbaum, R. Evaluation of aromatic amino acids as potential biomarkers in breast cancer by Raman spectroscopy analysis. *Scientific Reports.* 2021, **11**(1), p.1698.
342. Puccetti, P., Fallarino, F., Italiano, A., Soubeyran, I., MacGrogan, G., Debled, M., Velasco, V., Bodet, D., Eimer, S., Veldhoen, M., Prendergast, G.C., Platten, M., Bessede, A. and Guillemin, G.J. Accumulation of an Endogenous Tryptophan-Derived Metabolite in Colorectal and Breast Cancers. *PLoS One.* 2015, **10**(4), p.e0122046.
343. Paul, M.K. and Mukhopadhyay, A.K. Tyrosine kinase - Role and significance in Cancer. *Int J Med Sci.* 2004, **1**(2), pp.101-115.

344. Chan, J.W., Taylor, D.S., Zwerdling, T., Lane, S.M., Ihara, K. and Huser, T. Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells. *Biophys J.* 2006, **90**(2), pp.648-656.
345. Bergholt, M.S., Zheng, W., Lin, K., Ho, K.Y., Teh, M., Yeoh, K.G., So, J.B.Y. and Huang, Z. In Vivo Diagnosis of Esophageal Cancer Using Image-Guided Raman Endoscopy and Biomolecular Modeling. *Technology in Cancer Research & Treatment.* 2011, **10**(2), pp.103-112.
346. Xu, Z., Gao, Y., Hao, Y., Li, E., Wang, Y., Zhang, J., Wang, W., Gao, Z. and Wang, Q. Application of a microfluidic chip-based 3D co-culture to test drug sensitivity for individualized treatment of lung cancer. *Biomaterials.* 2013, **34**(16), pp.4109-4117.
347. Cheng, W.T., Liu, M.T., Liu, H.N. and Lin, S.Y. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma. *Microsc Res Tech.* 2005, **68**(2), pp.75-79.
348. Kaminaka, S., Ito, T., Yamazaki, H., Kohda, E. and Hamaguchi, H.-O. Near-infrared multichannel Raman spectroscopy toward real-timein vivo cancer diagnosis. *Journal of Raman Spectroscopy - J RAMAN SPECTROSC.* 2002, **33**, pp.498-502.
349. Malini, R., Venkatakrishna, K., Kurien, J., M. Pai, K., Rao, L., Kartha, V. and Krishna, C.M.J.B.O.R.o.B. Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: a Raman spectroscopy study. 2006, **81**(3), pp.179-193.
350. Szlasa, W., Zendran, I., Zalesińska, A., Tarek, M. and Kulbacka, J. Lipid composition of the cancer cell membrane. *J Bioenerg Biomembr.* 2020, **52**(5), pp.321-342.
351. Overholtzer, M.H., Yakowec, P.S. and Cameron, V. The effect of amino acid substitutions in the conserved aromatic region of subunit II of cytochrome c oxidase in Saccharomyces cerevisiae. *J Biol Chem.* 1996, **271**(13), pp.7719-7724.
352. Eley, C.G., Moore, G.R., Williams, R.J., Neupert, W., Boon, P.J., Brinkhof, H.H., Nivard, R.J. and Tesser, G.I. Structural role of the tyrosine residues of cytochrome c. *Biochem J.* 1982, **205**(1), pp.153-165.
353. Kendall, C., Stone, N., Shepherd, N., Geboes, K., Warren, B., Bennett, R. and Barr, H. Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus. *J Pathol.* 2003, **200**(5), pp.602-609.
354. Stone, N., Kendall, C., Crow, P. and Barr, H. Stone N, Kendall C, Smith J, Crow P, Barr HRaman spectroscopy for identification of epithelial cancers. Faraday Discuss 126: 141-157. *Faraday discussions.* 2004, **126**, pp.141-157; discussion 169.
355. Abramczyk, H., Surmacki, J., Brozek-Pluska, B. and Kopec, M. Revision of Commonly Accepted Warburg Mechanism of Cancer Development: Redox-Sensitive Mitochondrial Cytochromes in Breast and Brain Cancers by Raman Imaging. *Cancers.* 2021, **13**, p.2599.
356. McDermott, E.W., Barron, E.T., Smyth, P.P. and O'Higgins, N.J. Premorphological metabolic changes in human breast carcinogenesis. *Br J Surg.* 1990, **77**(10), pp.1179-1182.
357. Majeed, M.F., Hasan, A.F. and Khudair, Y.Y. Classification of oral cavity cancer using linear discriminant analysis (LDA) and principal component analysis (PCA). *AIP Conference Proceedings.* 2025, **3255**(1).
358. Adebiyi, M.O., Arowolo, M.O., Mshelia, M.D. and Olugbara, O.O. A Linear Discriminant Analysis and Classification Model for Breast Cancer Diagnosis. 2022, **12**(22), p.11455.
359. Hofmann, A.F.J.A.o.i.m. The continuing importance of bile acids in liver and intestinal disease. 1999, **159**(22), pp.2647-2658.

360. Cheng, K. and Raufman, J.P. Bile acid-induced proliferation of a human colon cancer cell line is mediated by transactivation of epidermal growth factor receptors. *Biochem Pharmacol.* 2005, **70**(7), pp.1035-1047.
361. Bhat, A.A., Lu, H., Soutto, M., Capobianco, A., Rai, P., Zaika, A. and El-Rifai, W. Exposure of Barrett's and esophageal adenocarcinoma cells to bile acids activates EGFR-STAT3 signaling axis via induction of APE1. *Oncogene.* 2018, **37**(46), pp.6011-6024.
362. Bhardwaj, V., Horvat, A., Korolkova, O., Washington, M.K., El-Rifai, W., Dikalov, S.I. and Zaika, A.I. Prevention of DNA damage in Barrett's esophageal cells exposed to acidic bile salts. *Carcinogenesis.* 2016, **37**(12), pp.1161-1169.
363. Souza, R.F., Shewmake, K., Terada, L.S. and Spechler, S.J. Acid exposure activates the mitogen-activated protein kinase pathways in Barrett's esophagus. *Gastroenterology.* 2002, **122**(2), pp.299-307.
364. Keswani, R.N., Chumsangsri, A., Mustafi, R., Delgado, J., Cohen, E.E. and Bissonnette, M. Sorafenib inhibits MAPK-mediated proliferation in a Barrett's esophageal adenocarcinoma cell line. *Dis Esophagus.* 2008, **21**(6), pp.514-521.
365. Sarosi, G.A., Jr., Jaiswal, K., Herndon, E., Lopez-Guzman, C., Spechler, S.J. and Souza, R.F. Acid increases MAPK-mediated proliferation in Barrett's esophageal adenocarcinoma cells via intracellular acidification through a Cl-/HCO3- exchanger. *Am J Physiol Gastrointest Liver Physiol.* 2005, **289**(6), pp.G991-997.
366. Huang, J., Liu, H., Sun, T., Fang, J.Y., Wang, J. and Xiong, H. Omeprazole prevents CDX2 and SOX9 expression by inhibiting hedgehog signaling in Barrett's esophagus cells. *Clin Sci (Lond).* 2019, **133**(3), pp.483-495.
367. Si, J., Fu, X., Behar, J., Wands, J., Beer, D.G., Souza, R.F., Spechler, S.J., Lambeth, D. and Cao, W. NADPH oxidase NOX5-S mediates acid-induced cyclooxygenase-2 expression via activation of NF-kappaB in Barrett's esophageal adenocarcinoma cells. *J Biol Chem.* 2007, **282**(22), pp.16244-16255.
368. Jenkins, G.J., Harries, K., Doak, S.H., Wilmes, A., Griffiths, A.P., Baxter, J.N. and Parry, J.M. The bile acid deoxycholic acid (DCA) at neutral pH activates NF-kappaB and induces IL-8 expression in oesophageal cells in vitro. *Carcinogenesis.* 2004, **25**(3), pp.317-323.
369. Hong, J., Li, D., Wands, J., Souza, R. and Cao, W. Role of NADPH oxidase NOX5-S, NF-κB, and DNMT1 in acid-induced p16 hypermethylation in Barrett's cells. *Am J Physiol Cell Physiol.* 2013, **305**(10), pp.C1069-1079.
370. Vaezi, M.F. and Richter, J.E.J.G. Role of acid and duodenogastroesophageal reflux in gastroesophageal reflux disease. 1996, **111**(5), pp.1192-1199.
371. Rokkas, T. and Sladen, G.J.A.J.o.G. Ambulatory esophageal pH recording in gastroesophageal reflux: relevance to the development of esophagitis. 1988, **83**(6).
372. Neumann, C. and Cooper, B.J.G. Oesophageal pH monitoring in Barrett's oesophagus. 2003, **52**(1), pp.153-153.
373. Wiener, G., Morgan, T., Copper, J., Castell, D., Sinclair, J., Richter, J.J.D.d. and sciences. Ambulatory 24-hour esophageal pH monitoring: reproducibility and variability of pH parameters. 1988, **33**, pp.1127-1133.
374. Nehra, D., Howell, P., Williams, C., Pye, J. and Beynon, J.J.G. Toxic bile acids in gastro-oesophageal reflux disease: influence of gastric acidity. 1999, **44**(5), p.598.
375. Kauer, W.K., Peters, J.H., DeMeester, T.R., Feussner, H., Ireland, A.P., Stein, H.J. and Siewert, R.J.J.S. Composition and concentration of bile acid reflux into the esophagus of patients with gastroesophageal reflux disease. 1997, **122**(5), pp.874-881.

376. Theisen, J., Nehra, D., Citron, D., Johansson, J., Hagen, J.A., Crookes, P.F., DeMeester, S.R., Bremner, C.G., DeMeester, T.R. and Peters, J.H.J.J.o.G.S. Suppression of gastric acid secretion in patients with gastroesophageal reflux disease results in gastric bacterial overgrowth and deconjugation of bile acids. 2000, **4**(1), pp.50-54.
377. Krafft, C., Neudert, L., Simat, T. and Salzer, R. Near infrared Raman spectra of human brain lipids. *Spectrochim Acta A Mol Biomol Spectrosc.* 2005, **61**(7), pp.1529-1535.
378. Huo, X., Zhang, H.Y., Zhang, X.I., Lynch, J.P., Strauch, E.D., Wang, J.Y., Melton, S.D., Genta, R.M., Wang, D.H., Spechler, S.J. and Souza, R.F. Acid and bile salt-induced CDX2 expression differs in esophageal squamous cells from patients with and without Barrett's esophagus. *Gastroenterology.* 2010, **139**(1), pp.194-203.e191.
379. Li, D., Hong, J. and Cao, W. Silencer-of-Death Domain Mediates Acid-Induced Decrease in Cell Apoptosis in Barrett's Associated Esophageal Adenocarcinoma Cells. *J Pharmacol Exp Ther.* 2017, **360**(1), pp.14-22.
380. Huo, X., Zhang, X., Yu, C., Cheng, E., Zhang, Q., Dunbar, K.B., Pham, T.H., Lynch, J.P., Wang, D.H., Bresalier, R.S., Spechler, S.J. and Souza, R.F. Aspirin prevents NF-κB activation and CDX2 expression stimulated by acid and bile salts in oesophageal squamous cells of patients with Barrett's oesophagus. *Gut.* 2018, **67**(4), pp.606-615.
381. Zaika, E., Wei, J., Yin, D., Andl, C., Moll, U., El-Rifai, W. and Zaika, A.I. p73 protein regulates DNA damage repair. *Faseb j.* 2011, **25**(12), pp.4406-4414.
382. Zaika, E., Bhardwaj, V., Wei, J., Washington, M.K., Souza, R., El-Rifai, W. and Zaika, A. Proinflammatory cytokines and bile acids upregulate ΔNp73 protein, an inhibitor of p53 and p73 tumor suppressors. *PLoS One.* 2013, **8**(5), p.e64306.
383. Das, K.M., Kong, Y., Bajpai, M., Kulkarni, D., Geng, X., Mishra, P., Banerjee, D. and Hirshfield, K. Transformation of benign Barrett's epithelium by repeated acid and bile exposure over 65 weeks: a novel in vitro model. *Int J Cancer.* 2011, **128**(2), pp.274-282.
384. Galenkamp, K.M.O., Sosicka, P., Jung, M., Recouvreux, M.V., Zhang, Y., Moldenhauer, M.R., Brandi, G., Freeze, H.H. and Commisso, C. Golgi Acidification by NHE7 Regulates Cytosolic pH Homeostasis in Pancreatic Cancer Cells. *Cancer Discov.* 2020, **10**(6), pp.822-835.
385. Mahapatra, K.K., Mishra, S.R., Behera, B.P., Patil, S., Gewirtz, D.A. and Bhutia, S.K. The lysosome as an imperative regulator of autophagy and cell death. *Cell Mol Life Sci.* 2021, **78**(23), pp.7435-7449.
386. Lingwood, D. and Simons, K.J.s. Lipid rafts as a membrane-organizing principle. 2010, **327**(5961), pp.46-50.
387. Hancock, J.F.J.N.R.M.C.B. Lipid rafts: contentious only from simplistic standpoints. 2006, **7**(6), pp.456-462.
388. Gowrishankar, K., Ghosh, S., Saha, S., Rumamol, C., Mayor, S. and Rao, M.J.C. Active remodeling of cortical actin regulates spatiotemporal organization of cell surface molecules. 2012, **149**(6), pp.1353-1367.
389. Charras, G.T. A short history of blebbing. *J Microsc.* 2008, **231**(3), pp.466-478.
390. Zhou, Y., Maxwell, K.N., Sezgin, E., Lu, M., Liang, H., Hancock, J.F., Dial, E.J., Lichtenberger, L.M. and Levental, I. Bile acids modulate signaling by functional perturbation of plasma membrane domains. *J Biol Chem.* 2013, **288**(50), pp.35660-35670.
391. Udensi, U.K. and Tchounwou, P.B. Potassium Homeostasis, Oxidative Stress, and Human Disease. *Int J Clin Exp Physiol.* 2017, **4**(3), pp.111-122.

392. Zhang, Z.X., Gan, I., Pavlosky, A., Huang, X., Fuhrmann, B. and Jevnikar, A.M. Intracellular pH Regulates TRAIL-Induced Apoptosis and Necroptosis in Endothelial Cells. *J Immunol Res.* 2017, **2017**, p.1503960.
393. Yu, S.P., Yeh, C.-H., Sensi, S.L., Gwag, B.J., Canzoniero, L.M., Farhangrazi, Z.S., Ying, H.S., Tian, M., Dugan, L.L. and Choi, D.W.J.S. Mediation of neuronal apoptosis by enhancement of outward potassium current. 1997, **278**(5335), pp.114-117.
394. Hughes, F.M., Bortner, C.D., Purdy, G.D. and Cidlowski, J.A.J.J.o.B.C. Intracellular K+ suppresses the activation of apoptosis in lymphocytes. 1997, **272**(48), pp.30567-30576.
395. Cain, K., Langlais, C., Sun, X.-M., Brown, D.G. and Cohen, G.M.J.J.o.B.C. Physiological concentrations of K+ inhibit cytochrome c-dependent formation of the apoptosome. 2001, **276**(45), pp.41985-41990.
396. Burg, E., Remillard, C. and Yuan, J.-J.J.T.J.o.m.b. K+ channels in apoptosis. 2006, **209**, pp.3-20.
397. Michl, J., Monterisi, S., White, B., Blaszczak, W., Hulikova, A., Abdullayeva, G., Bridges, E., Yin, Z., Bodmer, W.F. and Swietach, P. Acid-adapted cancer cells alkalinize their cytoplasm by degrading the acid-loading membrane transporter anion exchanger 2, SLC4A2. *Cell Reports.* 2023, **42**(6).
398. Ritter, M., Bresgen, N. and Kerschbaum, H.H. From Pinocytosis to Methuosis-Fluid Consumption as a Risk Factor for Cell Death. *Front Cell Dev Biol.* 2021, **9**, p.651982.
399. Shubin, A.V., Demidyuk, I.V., Komissarov, A.A., Rafieva, L.M. and Kostrov, S.V. Cytoplasmic vacuolization in cell death and survival. *Oncotarget.* 2016, **7**(34), pp.55863-55889.
400. Damaghi, M., Tafreshi, N.K., Lloyd, M.C., Sprung, R., Estrella, V., Wojtkowiak, J.W., Morse, D.L., Koomen, J.M., Bui, M.M., Gatenby, R.A. and Gillies, R.J. Chronic acidosis in the tumour microenvironment selects for overexpression of LAMP2 in the plasma membrane. *Nat Commun.* 2015, **6**, p.8752.
401. Ramaekers, F.C., Bosman, F.T.J.T.J.o.P.A.J.o.t.P.S.o.G.B. and Ireland. The cytoskeleton and disease. 2004, **204**(4), pp.351-354.
402. Davidson, P.M., Denais, C., Bakshi, M.C. and Lammerding, J. Nuclear Deformability Constitutes a Rate-Limiting Step During Cell Migration in 3-D Environments. *Cellular and Molecular Bioengineering.* 2014, **7**(3), pp.293-306.
403. Integrated genomic characterization of oesophageal carcinoma. *Nature.* 2017, **541**(7636), pp.169-175.
404. Xie, L., Sun, Z., Brown, N.J., Glinskii, O.V., Meininger, G.A. and Glinsky, V.V. Changes in dynamics of tumor/endothelial cell adhesive interactions depending on endothelial cell growth state and elastic properties. *PLoS One.* 2022, **17**(6), p.e0269552.
405. Korbut, E., Krukowska, K. and Magierowski, M. Barrett's Metaplasia Progression towards Esophageal Adenocarcinoma: An Attempt to Select a Panel of Molecular Sensors and to Reflect Clinical Alterations by Experimental Models. *Int J Mol Sci.* 2022, **23**(6).
406. Limberger, B., Ludwig, A., Lyros, O., Nowotny, R., Gockel, I. and Thieme, R. 427. CHARACTERIZATION OF THE CYTOSKELETON-ASSOCIATED PROTEIN 4 (CKAP4) IN A BARRETT'S ESOPHAGUS CELL CULTURE MODEL. *Diseases of the Esophagus.* 2022, **35**(Supplement_2).
407. Spechler, S.J. and Souza, R.F. Barrett's esophagus. *N Engl J Med.* 2014, **371**(9), pp.836-845.