DEVELOPMENT OF RAMAN SPECTROSCOPIC POINT OF CARE DIAGNOSTICS FOR THE DETECTION OF KEY TARGETS OF DISEASE

By

EMMA LOUISE BUCHAN

A thesis submitted to
the University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Advanced Nano Materials and Applications Group
School of Chemical Engineering
College of Engineering and Physical Sciences
University of Birmingham
December 2023

ABSTRACT

Inflammatory bowel disease (IBD) and cardiovascular disease (CVD) represent two distinct yet urgent healthcare concerns. IBD, characterised by chronic inflammation of the gastrointestinal tract, poses a significant burden on the affected individuals. Both conditions require early and accurate diagnostic methods to optimise patient care and treatment outcomes. Conversely, CVD continues to be a leading global cause of morbidity and mortality, necessitating innovative approaches to diagnosis and monitoring. Here, we explore the transformative potential of Raman spectroscopy (RS) as a rapid and non-invasive diagnostic tool in the assessment of saliva and blood plasma. The multidimensional approach combining RS with multivariate analysis unveils distinct molecular profiles thus enabling early and accurate disease diagnosis. Classification accuracies of >88 % are demonstrated between three sample types applied to healthy, CVD and follow-up blood plasma. Similarly, in IBD, classification yields accuracies >90 % over four sample types applied to saliva and tissue. This further extends into biomarker discovery, unveiling novel CVD- and IBD-indicative molecular signatures. Feature extraction, inherently connected to classification results, indicates an increase in circulating cytokines, a hallmark of inflammation in both CVD and IBD. Extracellular vesicles (EVs), as biomarkers serve as key carriers of disease-specific biomolecules, therefore, rendering them invaluable in diagnostics and personalised treatment. Conventional EV isolation methods suffer from several limitations such as, low throughput, sample loss, limited scalability and EV damage thus this has led to the exploration of alternative EV isolation techniques. The developed methodologies harness the power of RS to redefine the paradigm of disease diagnosis and risk stratification and as a means of increasing diagnostic accuracy, bridging the translational gap, and facilitating the clinical integration of RS-based diagnostics *via* a novel lab-on-a-chip (LoC) platform for the isolation and separation of disease-indicative EVs from biofluids. The developed platform marks a significant advance in utilising the diagnostic potential of EVs, offering a streamlined and efficient approach to their isolation and analysis.

ACKNOWLEDGEMENTS

I would like to extend my sincere appreciation to the individuals and institutions whose support and collaboration have been integral to the completion of this thesis. This work was conducted in the dynamic and collaborative environment of the Advanced Nano-materials Structures and Applications (ANMSA) group, led by Professor Pola Goldberg Oppenheimer and as part of the EPSRC Sci-Phy-4-Health Centre for Doctoral Training studentship. I would like to begin by expressing my gratitude to everyone who has been part of the Sci-Phy centre. Their commitment to fostering academic excellence and supporting cutting-edge research initiatives have been instrumental in the development and realisation of this thesis.

I am profoundly grateful to my primary supervisor, Prof. Pola Goldberg Oppenheimer whose unwavering support and guidance has been invaluable throughout this research journey. Offering insightful feedback, encouragement during challenging times, and a wealth of expertise that has shaped the trajectory of my research have made Prof. Goldberg Oppenheimer an incredible source of support and exceptional mentor.

I extend my sincere appreciation to the members of the ANMSA group who have been instrumental in shaping and enriching the research presented in this thesis. The collaborative spirit, intellectual discussions, and collective commitment to advancing knowledge within our research group have been central to the outcome of this work. Each member's unique contributions and shared dedication to excellence have made this journey both rewarding and enlightening.

I have also had the benefit of guidance from both Dr Mark Thomas and Dr Marietta Iacucci who have allowed me to discuss my project and facilitated the execution of the study. I am especially grateful to their research groups for their direction and assistance without which this project would not have gained the same success.

Over the years, there have been several research fellows, post-doctoral researchers, faculty, and staff that have helped me to accomplish my PhD. To this end, I am especially grateful to Dr Jonathan James Rickard and Dr Hin On Martin Chu who have provided exceptional support and played a cricual role in navigating the challenges of this project. Their willingness to lend a helping hand, field numerous questions, and share their expertise has been invaluable.

Finally, I would like to thank my friends and family whose support has been the cornerstone of my academic journey. To my parents, whose constant encouragement, sacrifices, and belief in my abilities have been my guiding light. I am deeply grateful for the values and strength you instilled in me and whose unwavering support has been instrumental in my every achievement. To my sisters, Samantha and Laura, thank you for being my biggest cheerleaders and source of inspiration. Whether through kind words of encouragement or shared laughter (often at my expense!), your support has been a cherished gift and made even the biggest challenges more achievable. To my Grandma, whose boundless wisdom and belief in me have provided me with the courage to face challenges head-on. To my best friend, Fiona, for being a pillar of strength and an endless bundle of fun. Your enthusiasm and countless late-night discussions have been a constant source of motivation and comfort. I would like to

make a special mention to Georgia who has been an incredible friend on this journey. Form being flatmates during the late-night chats to being workmates navigating the challenges of PhD life, your incredible support and shared joy have turned accomplishments into shared victories.

My deepest appreciation and thanks to you all for being a vital part of this significant milestone!

CONTENTS

LIST OF PUBLICATIONS

The following six papers and two patent applications form the basis of this thesis:

1. **Buchan E.,** Kelleher L., Clancy M., Rickard J.J.S., and Goldberg Oppenheimer P. [2021], 'Spectroscopic molecular-fingerprint profiling of saliva.' *Analytica Chimica Acta,* **1185,** 339074.
2. **Buchan E.,** Hardy M., Carvalho Gomes P., Kelleher L., Chu H.O.M., Goldberg Oppenheimer P. [2022], 'Emerging Raman spectroscopy and saliva-based diagnostics: from challenges to applications.' *Applied Spectroscopy Reviews*, 130351
3. **Buchan E.,** Majumder S., Lorenzo Parigi T., Nardone O.M., Abdawn Z., Hejmadi R., Bossa C., Panton M., Shivaji U., Iacucci M., Goldberg Oppenheimer P. [2023], 'Raman spectroscopy aided diagnosis of inflammatory bowel disease *via* assessing saliva and endoscopic tissue biopsy.'
4. **Buchan E.,** Rickard J.J.S., Goldberg Oppenheimer P. [2023], 'Raman spectroscopic molecular fingerprinting of biomarkers for inflammatory bowel disease.' *Clinical and Translational Medicine, DOI: 10.1002/ctm2.1345*
5. **Buchan E.** and Goldberg Oppenheimer P. [2023], 'Review – Vibrational spectroscopy aided diagnosis, prognosis and treatment of inflammatory bowel disease.' *Clinical and Translational Discovery, DOI: 10.1002/ctd2.249*
6. **Buchan E.,** Rickard, J.J.S, Thomas M., Goldberg Oppenheimer P., [2023], 'Advanced Integrated Multipurpose Spectroscopic Lab-on-a-chip for Timely Detection of Extracellular Vesicles as Key-Markers of Diseases.' Submitted.
7. **Buchan E.,** Rickard J.J.S., and Goldberg Oppenheimer P. [2023], 'Lab-on-a-chip.' UK Patent Application No. GB2304236.9
8. **Buchan E.,** Rickard J.J.S., and Goldberg Oppenheimer P. [2023], 'Lab-on-a-chip method of diagnosis.' UK Patent Application No. GB2304237.7

In addition, the following papers were published;

1. Hardy M., Kelleher L., Gomes P., **Buchan E**., Chu H.O.M., Goldberg Oppenheimer P. [2021], 'Methods in Raman spectroscopy for saliva studies – a review. *Applied Spectroscopy Reviews,* **7**, 177 – 232
2. Chu H.O.M., **Buchan E**., Smith D., Goldberg Oppenheimer P. [2024], 'Development and application of an optimised Bayesian shrinkage prior for spectroscopic biomedical diagnostics. *Computer Methods and Programs in Biomedicine,* DOI 10.1016/j.cmpb.2024.108014.
3. **Buchan E**., Thomas M., Goldberg Oppenheimer P. 'Molecular profiling for diagnosis, monitoring and biomarker classification of cardiovascular disease. Submitted.
4. Fernandez-Rhodes M., **Buchan E**., Gagnon S., Qian J., Gethings L., Lees R., Peacock B., Capel A., Martin N., Goldberg Oppenheimer P., Lewis M., Davies O. [2024], 'Extracellular vesicles may provide an alternative detoxification pathway during skeletal muscle myoblast ageing.' Submitted.

The following conference abstracts were also accepted for presentation;

1. **Buchan E,** Majumder S., Lorenzo Parigi T., Nardone O.M., Abdawn Z., Hejmadi R., Bossa C., Panton M., Shivaji U., Iacucci M., Goldberg Oppenheimer P. 'Raman spectroscopy analysis of saliva combined with an artificial neural network algorithm could discriminate between ulcerative colitis and Crohn's disease.' ECCO, Denmark (Poster).
2. **Buchan E.,** Lorenzo Parigi T., Nardone O.M., Abdawn Z., Hejmadi R., Shivaji U., Iacucci M., Goldberg Oppenheimer P. 'Raman spectroscopy aided diagnosis of inflammatory bowel disease *via* assessing saliva and endoscopic tissue biopsy.' Digestive Disease Week, USA (Oral).
3. **Buchan E.,** Rickard J.J., Goldberg Oppenheimer P. 'Lab-on-a-chip for the isolation and separation of vesicles as key targets of cardiovascular disease.' Medical Devices: Materials, Mechanics and Manufacturing, Greece (Oral).
4. Fernandez-Rhodes M., **Buchan E**., Martin N., Goldberg Oppenheimer P., Lewis M., Davies O. 'Extracellular vesicles provide an insight into age-related changes in a primary skeletal muscle model.' ISCT, France (Poster).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

One of the biggest public health challenges faced worldwide is the growth of the global ageing population. This is in part due to advances in medical care and a decline in birth rates. Associated with this particular healthcare challenge is the increase in incidence of chronic health conditions, such as cardiovascular disease (CVD) and inflammatory bowel disease (IBD) [1], [2]. The period of life spent in good health is not keeping pace with the increase in life expectancy and on average an adult will spend ten years of their life unwell. This in turn is contributing to a higher demand on the health care system. The steadily rising elderly population will require more from the health and social services, for a longer period of time as patients with multiple physiological age-related diseases require long term management with repeated visits from health care professionals. To date, coupled with an inadequate funding, this pressure on the healthcare system has been negatively impacting the quality of care given to patients.

Clearly there is an unmet need for interventions which can increase the period of time individuals spend in good health and also ensure that they reach an old age in good biological health. At present, there is no easy way of determining the biological age of individuals or assessing if someone is ageing well in the community without the application of a costly and time-consuming assay that involves epigenetic assessment. Therefore, the development of a quick, disposable, simple and low-cost bedside device to measure potential biomarkers of ageing and associated disease, from readily available biofluids, such as blood-plasma or saliva would be of a vital importance and a major step towards the determination of individual's ageing and diseased state [3], [4]. It would also allow for the accurate measurement of the efficacy of specific clinical therapies and interventions in diseases, indicating how an individual is clinically

progressing. Current diagnostic methods for diagnosis of inflammatory bowel disease (IBD) face several challenges in terms of invasiveness, cost and accessibility, delays to diagnosis, limitations with biomarkers, and risk of complications. The aim of this work is to therefore develop a hybrid methodology for real-time, quantitative diagnosis of IBD and CVD from easily accessible biofluids, including saliva.

One of the more patient oriented methods aimed at overcoming some of these challenges is the point of care diagnostics (PoCT). The development of a PoCT would reduce the number of costly and often frequent hospital visits required for diagnosis and monitoring of disease, through the introduction of analytical testing in primary care settings and indeed, at the home environment. PoCT devices would be convenient and efficient for patients and reduce the time spent waiting for a definitive diagnosis thus, not only improve diagnosis but also save many lives.

1.1 Thesis Overview

This thesis presents a comprehensive investigation into the development of a lab-on-a-chip platform for the detection of extracellular vesicles as key targets of ageing and the associated diseases of IBD and CVD. To achieve this, the study is divided into eight main chapters, each focussing on a specific aspect of the research.

Following **Chapter 1** as a general introduction to the thesis, **Chapter 2** serves as an introduction to inflammatory bowel disease pathophysiology and diagnostics, exploring clinically adopted methodologies. Further to this, the application of saliva in healthcare diagnostics is investigated, with the healthcare advantages associated with saliva highlighted. This is followed by an introduction to Raman spectroscopy, which provides the fundamental framework for our experimental study. Additional applications of Raman spectroscopy in saliva and IBD diagnostics are demonstrated, highlighting recent progress, current challenges and limitations. The primary results are described in **Chapters 3-7**, with 3-5 comprising self-contained articles published during the respective studies. Collectively, these chapters address the following key aims:

- **Chapter 3**: Develop proof-of-concept Raman spectroscopy profiling and classification methodology, towards establishing a baseline of healthy saliva for ongoing and future spectroscopic studies, to include diagnostic and forensic applications.
- **Chapter 4**: Demonstrate the potential of Raman spectroscopy to detect salivary changes consistent with IBD and compare those with spectra obtained from endoscopic tissue biopsy to establish a useful non-invasive diagnostic adjunct.

- **Chapter 5**: Evaluate the potential of Raman spectroscopy in the identification of IBD biomarkers *via* the characteristic barcoding of biochemical fingerprinting for current emerging diagnostic and prognostic studies.
- **Chapter 6**: Develop proof-of-concept hybrid Raman-AI tool to detect biochemical changes indicative of cardiovascular disease from blood plasma and to further classify a panel of potential biomarkers of disease for early disease diagnosis.
- **Chapter 7**: Design and develop a unique optofluidic lab-on-a-chip *via* novel nanophotonic-sensing platforms for detection and analysis of extracellular vesicles, indicative of ageing and the associated diseases.

Although the structure includes some areas of overlap, each chapter explores differing disease avenues to build a highly informative Raman spectral library of disease, central to the presented healthcare challenge. The developed tools and methodologies are combined and expanded in **Chapter 7** in an attempt to bridge the gap between benchtop laboratory research and point of care assessment, in the form of a lab-on-a-chip system for the rapid diagnosis of CVD and IBD, with the platform resulting in two patent applications. As healthcare processes tend to be ‘re-active’ as opposed to pro-active’ in their approach to diagnosis, this leads to a delay in diagnosis, treatment and poor prognosis. Addressing these challenges and ensuring the timely intervention and diagnosis of ageing associated disease requires, just the revolution in approach that this work embodied combining an inherently interdisciplinary methodology.

1.2 References

[1] C. Bielecki *et al.*, “Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells,” *J Biomed Opt*, vol. 17, no. 7, p. 0760301, 2012, doi: 10.1117/1.jbo.17.7.076030.

[2] T. Gaziano, K. Reddy, and P. F, “Cardiovascular Disease,” in *Disease Control Priorities in Developing Countries*, 2nd ed., D. Jamison, J. Breman, and A. Measham, Eds., Oxford University Press, 2006, pp. 1–62.

[3] Dr. G. Calado *et al.*, “Raman Spectral Study of Saliva: A New Tool for Detection of Malignant and Premalignant Oral Lesions,” *Oral Surg Oral Med Oral Pathol Oral Radiol*, vol. 128, no. 1, pp. e90–e90, 2019, doi: 10.1016/j.oooo.2019.02.230.

[4] C. Bielecki *et al.*, “Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells,” *J Biomed Opt*, vol. 17, no. 7, p. 0760301, 2012, doi: 10.1117/1.jbo.17.7.076030.

CHAPTER 2

INFLAMMATORY BOWEL DISEASE, SALIVA AND RAMAN SPECTROSCOPY IN HEALTHCARE

This introductory thesis chapter is predominantly based on two previously published works.

1. **Buchan E.,** Hardy M., Carvalho Gomes P., Kelleher L., Chu H.O.M., Goldberg Oppenheimer P. [2022], 'Emerging Raman spectroscopy and saliva-based diagnostics: from challenges to applications.' *Applied Spectroscopy Reviews*, 130351

Author contributions – **E Buchan** conceptualised the review with **M. Hardy** and **P. Goldberg Oppenheimer**. **E Buchan** performed meta-analysis with **M. Hardy. E Buchan** wrote the manuscript with contributions to illicit drugs section from M. Hardy, dental applications from **H.O.M. Chu,** protein analysis from **P. Carvahlo Gomes** and forensics from **L. Kelleher. E Buchan** revised and edited the manuscript with **P. Goldberg Oppenheimer** and **M.Hardy.**

2. **Buchan E.** and Goldberg Oppenheimer P. [2023], 'Review – Vibrational spectroscopy aided diagnosis, prognosis and treatment of inflammatory bowel disease.' *Clinical and Translational Discovery, DOI: 10.1002/ctd2.249*

Author contributions – **E Buchan** conceptualised the review with **P Goldberg Oppenheimer**. **E Buchan** performed the literature review and wrote the manuscript. **E Buchan** with **P Goldberg Oppenheimer** edited and reviewed the manuscript.

Inflammatory bowel disease (IBD) is a group of chronic inflammatory intestinal conditions which affects an estimated 10 million people worldwide [1], [2]. The incidence and prevalence of IBD varies considerably both between and within geographic regions with North America and Northern Europe accounting for the highest incidence at 15.4 people per 100,000 and 23.1 per 100,000 [3]. Over time, the prevalence of IBD is increasing primarily due to the chronic condition, lacking a definitive cure. Further factors contributing to this trend include young age of onset, relatively low mortality rates and the potential for exponential growth due to increasing incidence rates as well as an aging population [4]. Despite extensive research, the precise aetiology of IBD remains unclear, with factors such as variations in the immune system response, genetic factors, and environmental influences implicated in its development [5], [6], [7], [8]. Clinically, IBD is categorised into two major types, ulcerative colitis (UC) and Crohn's disease (CD), distinguished by their symptomatic presentations, location of inflammation, and histopathological characteristics. UC typically causes persistent inflammation and superficial ulceration limited to the colon, whereas CD manifests as a transmural disease with a tendency for granuloma formation and can arise in any part of the GI tract [9], [10]. Currently, there is no known cure for the disease. Initial signs of IBD often appear upon exposure to medications or an infection, which irritates the intestines with common symptoms including abdominal pain, diarrhoea (oftentimes with blood), faecal incontinence, rectal bleeding, weight loss and malnutrition [11], [12], [13], [14]. Complications associated with IBD can lead to life-threatening conditions such as, the development of colon cancer, primary sclerosing cholangitis, coronary artery disease, and endothelial dysfunction.

2.1 Pathophysiology and Pathobiology of IBD

The intestinal mucosa consists of several key components; epithelial cells which form a monolayer closely connected by tight junctions thus preventing bacteria or antigen entry into the circulation, goblet and Paneth cells which limit the spread of luminal microorganisms *via* the production of mucus and secretion of a-defensins with intrinsic antimicrobial activity, respectively, as well as stroma and immune cells [1], [2]. The intestine is a highly dynamic tissue organised as a series of protrusions and invaginations, known as villi and crypts of Lieberkühn respectively. This epithelial layer serves a dual purpose of facilitating nutrient absorption whilst also acting as a physical barrier against luminal contents. It also interacts with both the gut microbiota and the immune system, transmitting and receiving signals from both [3]. Beneath the epithelium lies the lamina propria which contains stromal cells including fibroblasts, myofibroblasts, and perivascular pericytes, which contribute to fibrosis and wound healing. Plasma cells, also found in the lamina propria, release immunoglobulin A (IgA), which inhibits the infiltration of pathogenic microorganisms, thus maintaining a balance between the host and commensal microbiota [4], [5] (**Figure 2.1)**. Both the epithelium and non-immune intestinal components play crucial roles in intestinal homeostasis and IBD pathophysiology, often through interactions with components of the immune system [6], [7].

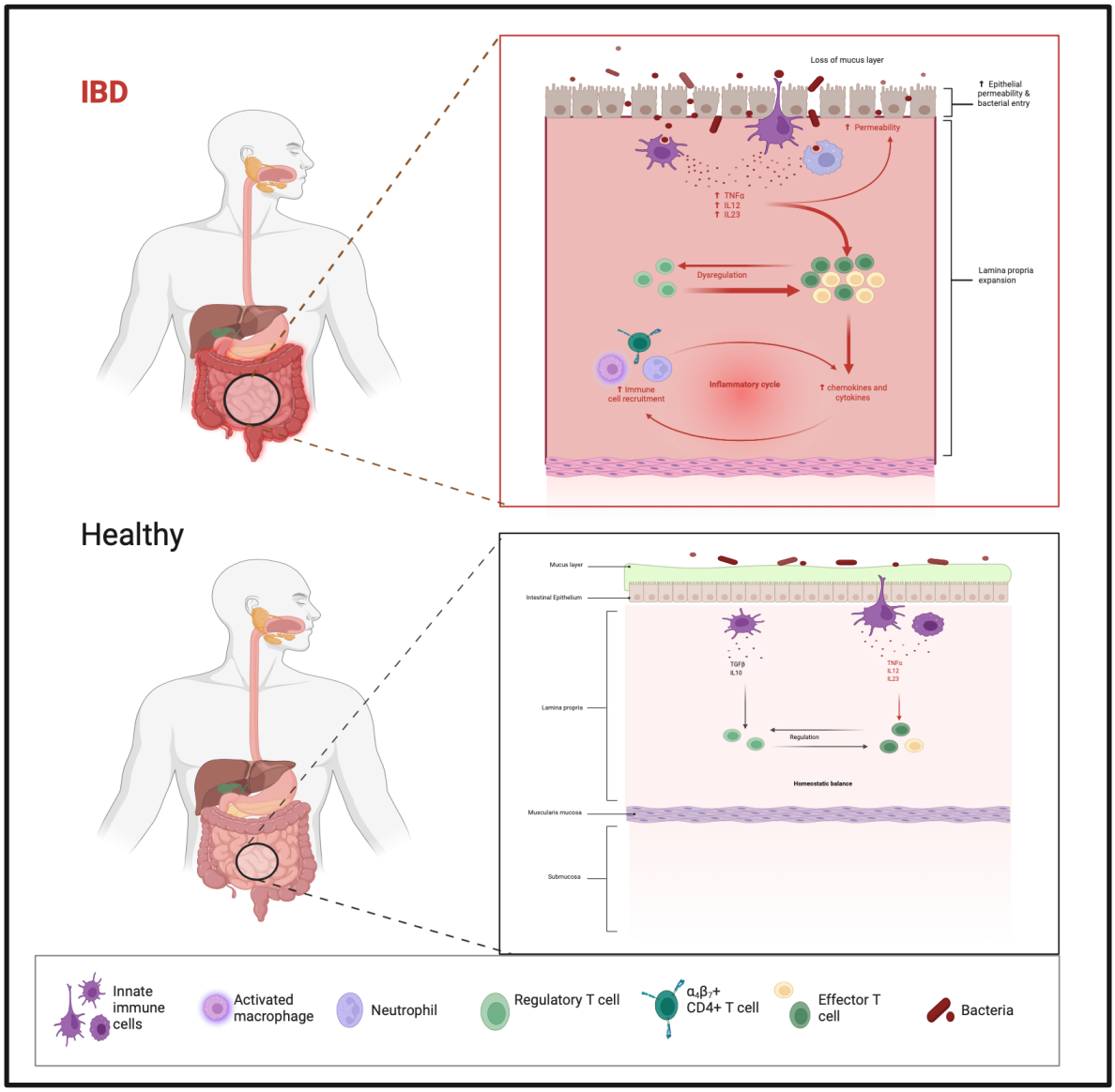

Figure 2.1. Schematic illustrating the mechanism of inflammation in IBD and the intestinal structure in both healthy and IBD patients. The immune system triggers an abnormal response which results in tissue damage in the GI tract (Figure created in Biorendor.com, adapted from IBD inflammation template).

The immune system plays a central role in host defence against pathogens and consists of innate and adaptive immunity. Innate immunity, comprising myeloid cells amongst others, elicits a rapid and non-specific response to microbial products through

pattern recognition receptors (PRRs) like toll-like receptors (TLRs) and Nod-like receptors (NLRs) [8], [9]. Innate immune cells also generate cytokines and chemokines, thus triggering the complement cascade and phagocytosis, as well as stimulating adaptive immunity *via* antigen presentation. Key innate immune cells include neutrophils, monocytes, macrophages, and dendritic cells [10], [11].

Although the exact aetiology of IBD remains elusive, it appears to present in individuals with genetic predispositions due to an aberrant immune response to the intestinal flora. Dysregulation in both innate and adaptive immune responses have been linked to such an abnormal immune response in IBD patients [12], [13]. Characteristic to IBD are breaches in the epithelial barrier and increased intestinal permeability at various areas within the intestine, along with persistent mucosal damage. While the precise cause of this damage is often unidentified, potential triggers could include infectious agents, chemical compounds, or metabolic changes potentially linked to dietary-induced dysbiosis [14], [15], [16]. Subsequently, the disease is believed to be perpetuated by an inadequate resolution of the initial inflammatory response as a result of intestinal injury. There also remains a degree of uncertainty regarding whether alterations in the epithelial barrier occur before or after the onset of inflammation within the lamina propria.

2.1.1 Crohn's Disease

CD may present at any age; however, it typically begins between 15 and 30 years of age. There is a familial predisposition to CD, with first-degree relatives seeing a 13-18% increase in incidence, suggesting a genetic component. A key pathological

feature of CD is its ability to extend through all layers of the bowel wall. Examination of tissue samples from individuals with CD reveals several microscopic abnormalities, including lymphoid hyperplasia, widespread infiltration of granulomas, and discrete noncaseating granulomas in the submucosa and lamina propria [17], [18]. Early mucosal lesions in CD often manifest as aphthous ulcers, small superficial ulcerations. Over time, these lesions can enlarge and merge, forming longitudinal and transverse ulcers with a distinctive cobblestone appearance due to deep mucosal ulceration and submucosal thickening. Another characteristic feature of CD is the presence of patchy lesion distribution with sharply delineated areas of disease surrounded by normal mucosa [19], [20].

In CD, immune cells such as CD4, CD8, B-cells, CD14 monocytes and natural killer cells infiltrate the gut of CD patients, with the T-cell response dominated by T-helper 1 (Th1) cells [21], [22], [23], [24]. Activation of immune cells leads to the production of various inflammatory mediators such as cytokines, chemokines, and growth factors. These mediators contribute to the inflammation and tissue damage, ultimately resulting in the clinical symptoms of disease [22], [23], [24], [25].

2.1.2 Ulcerative Colitis

UC typically presents in individuals aged between 20 and 50 years of age as either acute attacks or as a chronic condition with minimal pain. While approximately 70% of patients experience complete symptom remission between intermittent episodes, around 15% endure continuous symptoms without. Symptoms typical of UC presentation include rectal bleeding with diarrhoea, with periods of remission and

episodes of relapse [26]. UC primarily affects the mucosa lining the colon, extending from the rectum to the cecum. Unlike CD, UC is confined to inflammation of the mucosal layer and does not extend through the entire thickness of the intestinal wall. Sharp localised abdominal pain and fistula formation are, however, rare in UC. Instead, extensive superficial mucosal ulceration develops alongside chronic inflammation and the concurrent release of a diverse array of inflammatory mediators [27], [28], [29].

Histopathological examination of UC typically reveals neutrophil-mediated epithelial injury, which may present as neutrophils infiltrating crypt epithelium (cryptitis), infiltration of surface epithelium with or without mucosal ulceration or as a collection of neutrophils within crypt lumens (crypt abscesses) [30]. These crypt abscesses can spread laterally, compromising mucosal integrity, and leading to the formation of pseudo polyps observed during endoscopy and radiography. Following mucosal damage, the healing process involves the formation of highly vascular granulation tissue [13]. Given that UC shares many histological features with various other inflammatory disorders of the colon, including infections, diagnosis relies on the identification of specific clinical and pathological criteria and the exclusion of alternative causes. In terms of the immune response, UC may be driven by a Th2 response involving interleukin-4 (IL-4) and IL-13 or mediated by specialised cells such as natural killer T cells (NKT) [31], [32], [33].

2.2 Diagnostic Methodologies

IBD typically presents as a sudden flare-up symptom with the majority of individuals diagnosed under the age of 35. At present there is no single test to diagnose IBD as

the symptoms of the disease significantly overlap with other conditions, including infections and other digestive system disorders [31]. Thus, a typical IBD diagnosis comprises a combination of a physical exam, colonoscopy, upper endoscopy with biopsies, stool sample analysis, imaging (computerised tomography (CT), magnetic resonance imaging (MRI) or X-rays) and blood work such as ASCA and ANCA antibodies [34], [35]. Current diagnostics of IBD are associated with certain challenges such as, the 'gold standard' colonoscopy and endoscopy being invasive, time-consuming, not timely enough and may risk infection in patients. Moreover, distinct morphological or structural abnormalities may not be apparent in early pathologies [36]. Blood and stool tests often lack sensitivity and specificity and may fail to detect mild or early-stage disease [34], [37], [38]. Furthermore, imaging such as MRI and CT entail high costs, placing a burden on healthcare systems with access to such specialised tools and the required expertise often limited, particularly in remote geographical locations, causing delays in diagnosis and treatment.

Early and accurate detection of IBD is crucial in providing appropriate interventions. It can further facilitate monitoring the progression of disease as well as the therapeutic responses. However, timely diagnosis of IBD still remains an unmet need. With an overlap in presentation, symptoms and disease progression, discrimination of UC from CD is essential when selecting the most appropriate therapeutic or surgical regimen to improve patient prognosis. The determination of the correct therapeutic regime is often evaluated based on the severity of active inflammation which is often further complicated by the lack of recognised gold standards for accurate diagnosis [39]. Clinically, IBD diagnosis is based on a combination of symptom presentation,

histological and endoscopic evaluation of the mucosa [40]. This diagnostic process is both time consuming and highly invasive. Moreover, endoscopic evaluation of chronic colitis varies between CD and UC with 89% of patients with endoscopists often facing a disease with a non-differentiating appearance and histopathologists one of inflammation consistent with both forms of the disease [40], [41]. This further highlights the need for non-invasive, rapid, and accurate tools to reduce the burden of disease diagnosis as well as successfully differentiate UC from CD. The use of salivary biomarkers is currently being investigated for the diagnosis and monitoring of various diseases including IBDs and cancers due to its non-invasive collection and its constituents containing a diverse array of biomolecules reflective of systemic health and diseased states.

2.4 Saliva

In recent years, biofluid analysis has become a significant area of research, driven by the recognition that many biomarkers can be identified and thus serve as valuable indicators of disease [42], [43], [44], [45]. Saliva is central to the preservation and maintenance of oral health; however, it often receives little attention until the overall quality or quantity has declined. Beyond its oral functions, saliva contains a diverse array of molecules, including proteins, enzymes, hormones, and genetic material, making it a rich source of biological information [46], [47]. Recent advancements in analytical techniques have unlocked the potential of saliva as a non-invasive tool for disease diagnosis and monitoring of health status. Pfaffe et al. acknowledged the potential of saliva early on, noting that while technologies with sufficient analytical sensitivity exist to utilise saliva for diagnosis, they have yet to be widely integrated into

clinical practice beyond simple home-testing kits [48]. More comprehensive research into identifying, validating, and establishing associations between biomarkers and disease is required to progress this field of study.

2.4.1 Saliva Function

Saliva plays a fundamental role in maintaining oral health and supporting various physiological functions throughout the body. One of its primary functions is lubricating the oral cavity, which is necessary for speech, mastication, and swallowing, facilitating the passage of food through the digestive tract without inflicting damage to the mucosa [49], [50]. Saliva also acts as a natural cleanser, with the oral cavity almost constantly flushed with saliva, thus, washing away food particles, debris, and bacteria from the teeth and gums, thereby aiding in the prevention of dental decay and gum disease [51], [52], [53]. Furthermore, enzymes such as amylase found in saliva initiate the digestion of dietary starches, and lipase, which begins the breakdown of fats, contributing to the digestive process. Additionally, saliva contains three buffer systems (bicarbonate, phosphate, and protein) helps to maintain the pH balance (6.0-7.5) within the mouth [46], [49], [54], [55], [56]. Depending upon the taste, consistency, and concentration of a substance placed in the oral cavity, the flow of saliva will increase accordingly. When the volume of saliva is approximately 1.1 mL the swallowing reflex is triggered. This process of salivary stimulation, along with the dilution of tastes and subsequent act of swallowing, continues until the flavour concentration reaches a threshold where it no longer stimulates salivary flow [55], [56].

Beyond its oral functions, saliva contains antimicrobial compounds, such as lysozyme and lactoferrin, and antibodies that play a crucial role in protecting against oral infections and promoting wound healing [46], [54]. Saliva also serves as a reservoir for various substances, including hormones, electrolytes, and antibodies, which can be sampled and analysed for diagnostic purposes, offering insights into systemic health conditions [57], [58]. The overall multifaceted functions of saliva highlight its importance in oral health, digestion, immunity, and diagnostic medicine.

2.4.2 Saliva Composition

Saliva, primarily composed of water (~99%) and typically exhibiting a frothy texture, is primarily produced by three pairs of major salivary glands: the parotid, submandibular and sublingual glands, along with numerous minor salivary glands scattered throughout the oral cavity [47], [59] (**Figure 2.2)**. The parotid glands, located near the ears, produce a watery type of saliva rich in enzymes, such as amylase, whereas the submandibular glands situated beneath the lower jaw, which accounts for up to 70% of salivary flow, produces a mixture of serous and mucous saliva containing both enzymes and mucins [50], [60]. The sublingual glands, found beneath the tongue primarily secrete mucous saliva which aids in lubrication and protection of the oral tissues [54].

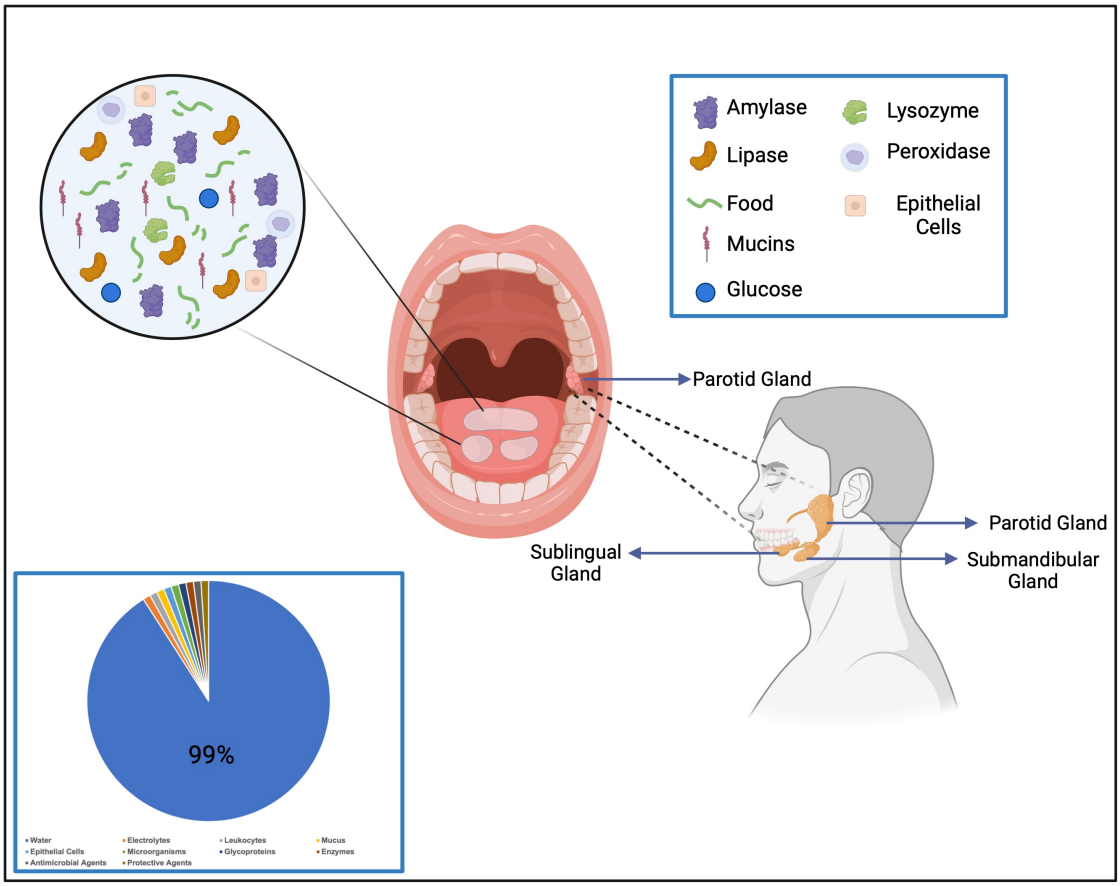

Figure 2.2. Schematic indicating the location of the primary salivary glands; parotid, submandibular and sublingual and the composition of saliva. (Figure made in Biorendor.com)

Clusters of cells called acini are the basic secretory units of salivary glands. These cells secrete a fluid consisting of water, electrolytes, mucus and enzymes with which flows out of the acinus and enters collecting ducts [55], [56], [61]. The composition of such secretions is altered within the ducts with the sodium predominantly reabsorbed, potassium secreted, and significant quantities of the bicarbonate ions also secreted. The smaller collecting ducts within the salivary glands converge into larger ducts, ultimately forming a single larger duct that empties into the oral cavity [56].

Collectively, proline-rich peptides (PRPs), the glycoprotein α-amylase, and mucins constitute nearly 80% of salivary proteins. Mucins, another crucial component, contribute to saliva's viscosity, facilitating lubrication of intraoral structures, aiding swallowing, and forming a barrier against microbial invasion [47], [49], [50]. Saliva also contains extracellular vesicles (EVs), lipids, and various non-salivary constituents such as epithelial cells, microorganisms, and food debris [51], [62], [63]. Alongside lysozyme and lactoferrin, salivary peroxidase and histatins are also known to have antimicrobial properties. Other organic components such as, glucose, urea, cortisol, sex hormones and blood group substances are contained in saliva and have been used as screening/diagnostic tools [64], [65].

It is important to note salivary composition can be influenced by factors such as hormonal fluctuations, psychological states, level of physical activity, oral hygiene, and whether saliva production is stimulated or unstimulated [54], [63], [65], [66]. Stimulated and unstimulated saliva, for example, differ in composition due to the differing gland contributions. The submandibular glands contribute 60% of the saliva with the parotid accounting for a further 20% in unstimulated saliva production [67]. However, the parotid gland typically becomes the dominant contributor to saliva production *via* stimulation of saliva through chewing. Aside from salivary collection method, saliva composition naturally varies throughout the day as a result of circadian rhythm and yearly seasons [68]. Typically, in healthy individuals salivary flow rate peaks in the late afternoon, however sodium and chloride levels peak in the morning.

2.4.3 Saliva *versus* Blood

The recognition of saliva's utility as a diagnostic biofluid has created significant interest in leveraging it for disease diagnosis and monitoring in healthcare [69], [70], [71]. Only recently has the full potential of saliva in medical applications been recognised, particularly in its ability to reflect a broad spectrum of health and disease states and thus offering unique advantages over blood. Blood, a fluid connective tissue is composed of 55% plasma and 45% formed elements that include white blood cells (lymphocytes, monocytes, neutrophils, eosinophils, and basophils), red blood cells, and platelets [84]. The plasma component of blood is primarily composed of water, but also contains proteins such as, albumin, clotting factors, antibodies, enzymes, and hormones, as well as, sugars (glucose), and fat particles [86]. Analogous to blood, saliva contains a diverse array of constituents as previously outlined, with many of these components entering saliva from the bloodstream *via* transcellular (passive intracellular diffusion and active transport) or paracellular routes (extracellular ultrafiltration) [72], [73], [74]. Consequently, most compounds found in the blood are also present in saliva, thus making saliva functionally equivalent to serum in reflecting the body's physiological state, including emotional, hormonal, nutritional, and metabolic variants [72], [73], [75]. Paffe et al. for example, have assembled an extensive list of biomolecules typically detected in both blood and saliva, along with their clinically significant ranges [76]. Such a compilation underscores markers for various conditions such as cardiovascular diseases and breast cancers, among others [73], [74], [77], [78], [79], [80], [81]. While concerns have arisen regarding relatively lower concentrations of diagnostic components in saliva compared to blood, advances in sensitive techniques, such as nanotechnology and molecular diagnostics have

mitigated this constraint. As a result, an increasing number of proof-of-concept diagnostic assays have been developed using saliva to monitor various diseases, including immune responses to viral infections such as hepatitis A, B and C [82], [83], [84], [85].

Furthermore, the use of saliva as a diagnostic biofluid offers numerous advantages. Saliva collection is non-invasive and painless compared to the invasive procedures required for blood collection, thus it is more acceptable and feasible, especially in populations where blood draws may be challenging or undesirable [43], [74], [86], [87]. It also diminishes privacy issues associated with collection of urine. In addition, saliva can be easily collected in various settings without the need for specialised equipment or trained personnel, allowing for convenient sampling in field or community-based studies. Its non-invasiveness and ease of collection also facilitate repeated sampling over time, enabling longitudinal studies and monitoring of disease progression or treatment response without undue burden or participants [88], [89]. Moreover, saliva is generally more stable than blood, particularly in terms of protein and nucleic acid stability, and can be stored at room temperature for short periods, reducing the need for stringent storage conditions.

Investigating saliva as a diagnostic tool effectively meets the demands for a cost-effective, non-invasive, and user-friendly screening approach. In the clinical setting, saliva offers numerous benefits concerning its collection, storage, transportation, and capacity for large-scale sampling with each of these processes conducted more economically then for example, serum or urine.

2.5 Current Analysis Methods

In many instances of disease, tissue biopsies are obtained for histopathological analysis, however the procedure is inherently invasive, uncomfortable for the patient and carries risks of complications. Moreover, the process of obtaining biopsies requires skilled personnel and specialised equipment, making it impractical for routine screening and monitoring of conditions. Histopathological analysis of tissue samples obtained from biopsies is time-consuming and reliant on expert pathologists for interpretation, limiting its scalability and accessibility, especially in resource-limited settings [31]. Other traditional approaches such as enzyme linked immunosorbent assay (ELISA) and polymerase chain reaction (PCR) which are commonly applied in biofluid analysis have several limitations [90], [91]. ELISA, while sensitive and widely used, demands laborious sample preparation, and relies on specific antibodies for detection, limiting its applicability to known biomarkers and increasing the risk of false positives or negatives [92]. Similarly, PCR, although highly sensitive and specific, necessitates skilled personnel, costly instrumentation, and time-consuming procedures for amplification and analysis, impeding its suitability for point-of-care diagnostics [93]. Moreover, mass spectrometry, while also offering high sensitivity and the ability to analyse complex mixtures, requires extensive sample preparation [94], [95]. Its application to saliva analysis often involves challenges such as sample variability and matrix effects which can affect the reliability and reproducibility of results.

2.6 Proposed Methods of Analysis with Extracellular Vesicles

Biomarkers obtained *via* liquid biopsies are generating considerable attention within the research and medical communities. This interest stems from their minimally invasive acquisition and their potential to facilitate precision medicine, which seeks to manage a variety of diseases based upon their distinct molecular signatures unique to each patient. Extracellular vesicles (EVs) represent a subset of liquid biopsy markers that can be secured from clinical samples, including saliva, urine, whole blood, and cerebrospinal fluid [96], [97], [98], [99].

EVs are a heterogeneous group of membrane-bound vesicles which vary in submicron size. They represent a potential source for biomarker discovery and have been increasingly recognised as strong vehicles of long-range intercellular communication in the body [100]. They can transfer proteins, nucleic acids, and lipids to other cells, thereby influencing the recipient cell function [101]. Initially underappreciated and thought to act as a mechanism of cell disposal, EVs are now considered abundant and stable sources of circulatory biomarkers [98], [102], [103]. They provide a snapshot of the cells from which they originate and as such, are depositaries of significant information. Recent research has looked at the use of EVs (in specific, exosomes) as therapeutic vectors functionalised to trigger an immune response in a host, and application to COVID-19 has been identified [104].

Typically, EV subpopulations are broadly divided into three main groups based on their biogenesis: a) exosomes, which constitute a homogenous population of spherical

vesicles of approximately 40-100 nm; b) microvesicles, which are large bilayered vesicles 100-1000 nm in diameter; and c) apoptotic bodies that comprise a highly heterogeneous group of bilayered vesicles 800-5000 nm in diameter [105], [106]. Due to the overlap in EV subtypes and difficulty in separating them, MISEV 2018 guidelines state that EVs can be divided into medium/large EVs (>200 nm) and small EVs (<200 nm) based on their physical properties. They have been found in the extracellular space as well as circulating through many different body fluids including saliva and blood [107], [108]. These biofluids contain non-vesicular macromolecular structures which could interfere with analytical results and thus, EVs isolation and enrichment is a required pre-analytical step for biomedical research [109]. In general, there are three main methodologies used for the isolation and purification of EVs including, ultracentrifugation/differential centrifugation, adsorption to microbeads (immunoaffinity) and size exclusion chromatography [110], [111], [112].

The study of EVs has gathered significant attention in recent years most notably due to their potential application as clinical biomarkers in the diagnosis and monitoring of disease [113]. The valuable cargo EVs can carry have made them potential biomarkers for diseases such as, liver hepatitis, hepatocyte injury in alcoholic and inflammatory liver disease, intrinsic kidney disease, atherosclerosis, stroke, asthma, Alzheimer's disease and diabetes [114], [115], [116], [117], [118], [119], [120], [121]. One notable advantage of EVs as biomarkers of interest, over the presence of soluble molecules including hormones and cytokines in the blood, is their inherent ability to protect the cargo from degradation thereby, maintaining its functionality and integrity [113]. Consequently, the release of EVs from any location and subsequent movement into

the bloodstream thus renders them easily accessible as liquid biopsies. Moreover, current literature has indicated that the characteristics, quantity or content of EVs can undergo alterations during disease progression [122], [123], [124]. Tumour cells for example, continuously release EVs and as such, tumour derived EVs are observed at significantly higher concentrations in plasma compared to those healthy individuals [122], [124]. This underlines the ability of EV molecular profiles to offer invaluable insights in the advancement of precision medicine.

Traditionally, microscopic methods such as atomic force microscopy (AFM), dynamic light scattering (DLS), flow cyto(fluoro)metry, and nanoparticle tracking analysis (NTA) have been used to characterise EVs and measure physical features such as vesicle morphology and size. ELISA may be used for the purpose of specificity, where only antibodies for external epitopes lead to a signal and surface plasmon resonance (SPR) is an easy way to obtain interaction parameters. However, given the importance of EVs in liquid biopsies better quantification of EVs is required. New sensing strategies relying on detection principles including fluorescence, colourimetry and electrochemistry-based technologies, allow for the detection of EVs either with the naked eye or without the requirement of purification steps, promising for the development of future point-of-care testing (PoCT) devices [125]. Due to its simplicity, fluorescence is the most popular approach for the detection of EVs. Fluorescent molecules are used as tags to label DNA probes and antibodies thus, enabling the detection of EVs via nucleic acid hybridisation and antibody/antigen interactions [126]. Moreover, colorimetric detection allows for direct visual observation of EVs. This simplifies the detection process and reduces the need for bulky instrumentation, fitting

for PoC testing [127]. However, there can be significant differences between output data between the different EV characterisation techniques, and arguably, there is no gold standard for routine benchtop EV analysis [112].

One of the main factors impacting the design and implementation of diagnostic EV methodologies is due to the inherent sample complexity and co-isolation of non-EVs. This complexity arises from several factors, including their diverse origin and composition. EVs carry a wide range of biomolecules, including proteins, nucleic acids, lipids, and metabolites, reflecting their cell of origin and physiological state [128], [129]. Consequently, EV samples are inherently complex, containing a mixture of vesicles with different cargo and functional properties [130], [131]. Furthermore, traditional isolation techniques for EVs, such as ultracentrifugation and size exclusion chromatography (**Table 2.1)**, often co-isolate non-EV components from biological fluids. These contaminants include lipoproteins, such as high-density lipoproteins and low density lipoproteins, protein aggregates, ribonucleoprotein complexes, and other membrane structures [132], [133], [134]. Co-isolation of non-EV components not only complicates downstream analyses but also introduces variability and artifacts in EV characterisation. EVs also share physical properties such as, size, density, and surface markers, with other extracellular particles and structures. For example, small EVs (exosomes) may overlap in size with lipoproteins, protein aggregates, and cellular debris thus making it challenging to selectively isolate pure populations of EVs [131], [134].

Table 2.1. Principle EV isolation methods applied in biomedical research.

Methods	Approximate Time (min)	Advantages	Disadvantages	Refs
Ultracentrifugation	140	- Isolation of all vesicular particles - Maintaining EVs structure - Absence of additional chemicals	- Time-consuming - Expensive equipment - Labour intensive	[135], [136], [137], [138]
Sucrose density gradient	250	- Divide EVs into different populations - Pure preparations	- Not preferable for samples with diverse EV populations - Complexity - Loss of sample	[139], [140], [141]
Ultrafiltration	130	- Easier to handle compared to ultracentrifugation - Simple procedure - Concurrent processing of many samples	- Retain contaminating proteins - Filter plugging - Small quantity of exosomal proteins	[142], [143], [144]
Exoquick-TC™	45 (overnight incubation)	Simple, fast, highly scalable	Low yield of EVs	[140], [145], [146], [147]
Immunoaffinity	240 (plus additional 2-8 hours for rinsing time)	Isolation of certain subpopulation of EVs	- Nonspecific binding - Recovery efficiency is less than 100%	[141], [148], [149], [150]
Size exclusion chromatography	1 mL/min (+ column wash)	- Purity - Enhances elimination of nonspecific impurities - Prevents Aggregation	- Limited sample volume and number of separated peaks - Highly specialised equipment	[151], [152]
Precipitation	65	- Simplicity - Preserves EV integrity	- Contamination - Polymer retention	[147]

At present the majority of salivary EV applications in healthcare are focussed on oral or dental disease such as oral squamous cell carcinoma, periodontitis and Sjögren's syndrome [153], [154], [155], [156]. Further applications include detection of

biomarkers of traumatic brain injury and Parkinson's disease and childhood asthma [157], [158], [159]. To date very few studies have investigated salivary EVs in IBD, with only Park *et al.* identifying changes in salivary microbial EVs before and after anti-TNF-α therapy in individuals with IBD [160]. A further study by Zhao *et al.* did, however, use blood plasma derived EVs to determine the levels of double stranded DNA (dsDNA) in CD patients and to determine to what extent the dsDNA may promote intestinal inflammation [161]. Moreover, blood plasma has found many differing applications in EV research from determination of clinical significance in glioblastoma patients, cancer screening, diabetic wound healing and infertility [162], [163], [164], [165], [166]. The diagnostic and prognostic potential of EVs in CVD has gained much interest in recent years with both EV- counts and EV- content associated with CVD. Cheow *et al*. determined that plasma EVs contain a novel diagnostic biomarker of EV-derived fibrinogen components that may potentially be further developed for clinical use [167]. A study by Oggero *et al.* used vein derived endothelial EVs to determine a mechanistic role for EVs in the development of EVs with significant and differential changes observed in their cargoes amongst patients post major cardiac event. The authors observed significantly higher levels of CD14+ and CD14+/CD41+ in EVs [168]. Gidlöf *et al*. determined that EVs isolated from plasma provided additional diagnostic value compared to just plasma alone. They identified proto-oncogene tyrosine-protein kinase SRC as a significant indicator of CVD in EVs. Further studies include determination of EV markers in relation to obesity and metabolic complications and how they lead to CVD manifestations and the identification of circulating EVs and their association with CVD risk markers [169], [170].

Raman spectroscopy offers an alternative method in the analytical study and detection of EVs. Recently, Gualerzi *et* al. described Raman spectroscopy as a tool to assess the purity of EV stem cell preparations. The authors indicate the ability of Raman to rapidly assess the composition and purity of EVs. The Raman spectral fingerprints obtained from the different sample preparations can discern both the cell source of the EV and which isolation procedure (ultracentrifugation, size exclusion) was applied [171]. A further study by Kwizera *et* al. describes the use of gold nanorods as a surface enhanced Raman scattering (SERS)-active platform in the detection of breast cancer cells [172]. Here, exosomes were captured on a miniaturised affinity-based device and target proteins on the EVs were detected using gold nanorods coated with QSY21 Raman reporters. The authors indicated that by using this technique, differences can be observed in the protein profiles of breast cancer cells compared to those normal breast cells. These studies analyse blood samples, but Raman spectroscopy also has the potential for similar EV analysis in saliva.

Notably, tip-enhanced Raman spectroscopy (TERS) might prove a useful technique for EV analysis, having so far been seemingly unexploited in the literature, whether for EVs specifically, or other salivary constituents [173]. TERS opens the possibility of characterising individual EVs by combining the chemical information provided by Raman scattering, the sensitivity of SERS and the spatial resolution of scanning probe microscopy. This technique uses a scanning tunnelling microscope (STM), or AFM set-up, with a plasmonically active tip, which when proximal to analyte molecules and illuminated with light of appropriate wavelength, can produce a SERS effect [174], [175]. EV damage however, caused by tip collision or plasmonic heating could be

problematic, as could inhomogeneity, should too few measurements be taken [176]. It is worth mentioning, that TERS has a long-standing and well-known problem of reproducible tip manufacture (and thus signal uniformity).

2.7 Raman Spectroscopy Principles and Instrumentation

The identification of quick, easy, and label-free methods to detect and evaluate the composition of EVs is vital to the development of new clinical tools to improve prognosis and reduce mortality. Simple measurements on particle counts and protein concentration provide brief overviews but fail to provide detailed information as to particle biochemical composition and cargos. A step towards early and non-invasive disease diagnosis is through the utilisation of EVs as biomarkers of disease. One such method useful in the detection and analysis of EVs is Raman spectroscopy (RS).

The first observation of Raman scattering was published in 1928 by C.V. Raman and K.S. Krishnan [177], [178]. Although industrial laboratories have used Raman spectroscopy for many years, since the early 1980s interest has grown. It is a versatile method suitable for the analysis of a wide range of samples across various fields, including physics, chemistry, materials science, biology and medicine [42], [179], [180], [181]. RS resolves many of the limitations associated with other spectroscopic techniques and is both a qualitative as well as quantitative technique. By measuring the frequency of scattered radiations, the technique can be described as qualitative while measuring the intensity of scattered radiations indicates quantitative analysis [182], [183].

RS is described as a scattering technique, based on the Raman effect where the frequency of a small portion of the scattered light is at a different frequency from the monochromatic incident light **(Fig.2.3)**. This scattered light is used to measure the vibrational modes of the sample and can provide both chemical and structural information, as well as the identification of substances through the characteristic "fingerprint."

In RS, the sample is illuminated with a monochromatic laser beam where it interacts with the molecules within the sample and thus scatters the incident light [184]. In the majority of scattering events, upon its interaction with the photon, the energy of the molecule remains unchanged, therefore the energy (or wavelength) of the scattered photon is equal to that of the incident photon. This is called elastic or Rayleigh scattering. In a much rarer event (approximately 1 in 10 million photons) Raman scattering occurs. This involves a transfer of energy between the scattered photon and the molecule and is termed inelastic scattering. If during the scattering event the molecule gains energy from the photon, i.e. it is excited to a higher vibrational energy level and the resulting scattered photon will lose energy and increase in wavelength then Stokes lines appear in the Raman spectrum [182], [185]. However, if the molecule relaxes to a lower energy level due to a loss in energy, then the photon will gain the equivalent energy with a decrease in wavelength and thus gives rise to anti-Stokes lines in the Raman spectrum as observed in **Fig 2.3**. When light is scattered by a molecule, the molecule is left in a higher energy state with the energy of the resulting photon transferred to the molecule when the oscillating electromagnetic field of a photon induces a polarisation of the molecular electron cloud. A complex called the

virtual state of the molecule forms between the molecule and the photon. The virtual state is unstable; therefore, this complex is short-lived with the resulting photo re-emitted as scattered light almost immediately [186], [187].

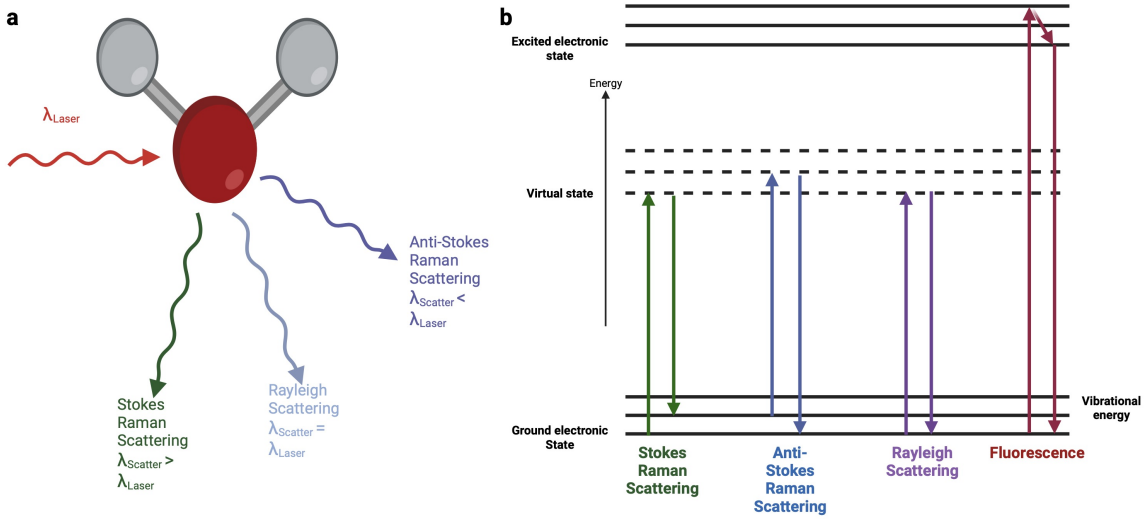

Figure 2.3. (a) Basic principle of Raman spectroscopy and **(b)** Jablonski diagram illustrating energy transitions for Rayleigh and Raman scattering. Rayleigh scattering is elastic with the incident photon having the same energy as the scattered photon. Raman is inelastic and in Stokes scattering the incident photon has a greater energy than the scattered photon while in anti-Stokes scattering, the incident photon has a lower energy than the scattered photon.

Stokes bands are more intense than anti-Stokes bands since Stokes shifted Raman bands involve transitions from lower to higher vibrational levels, thus they are measured in conventional Raman spectroscopy. In contrast, anti-Stokes bands are measured with fluorescing samples as fluorescence causes interference with Stokes bands [185]. The overall magnitude of Raman shift is not dependent on the wavelength of incident radiation, it is rather Raman scattering that is dependent on the wavelength of incident radiation. Essential to obtaining a Raman spectrum is a change in

polarizability. Here, water is an ideal solvent for sample preparation as Raman scattering due to the presence of water is low [182], [183].

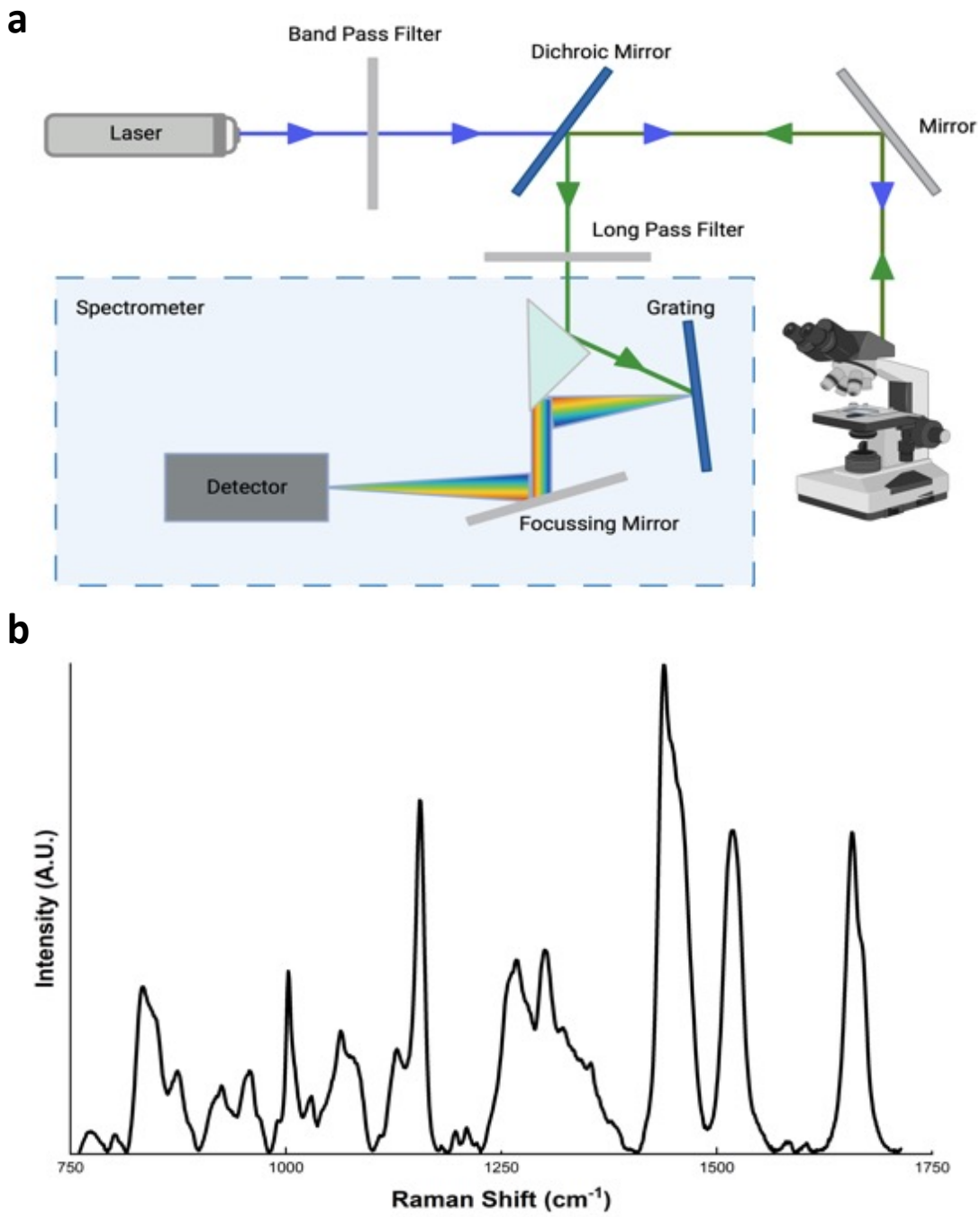

Figure 2.4. (a) Schematic representation of a Raman spectrometer with excitation light indicated by the blue line and scattered light shown in green. Excitation light travels from the laser, through the band pass filters, beam expander and dichroic mirror where it is then reflected off a mirror into the optical

system and directed onto the sample of interest. Light scattered by the sample is collected by the optics and directed onto a grating where it is dispersed via a focussing mirror and long pass filter. Finally, the dispersed light is focused onto the detector. **(b)** Example Raman spectra obtained from human biofluid indicating the intensity and Raman shifts of the scattered Raman light.

A Raman spectrum comprises a series of peaks, showing the intensity and wavelength position of the scattered Raman light (**Fig 2.4)**. Each peak on a Raman spectrum corresponds to a specific molecular bond vibration. Examples include individual bonds for example, C=C, N-O, C-H and C-C as well as groups of bonds such as lattice modes, benzene and tryptophan ring breathing modes etc. Raman spectra can be recorded over a spectral range from 10-4,000 cm^{-1}, Raman active normal modes of vibration of organic molecules occur in the range of 400-3,800 cm^{-1} [184], [187]. Typically, a Raman spectrum is a distinct chemical fingerprint for a particular molecule or material and can be used to quickly distinguish one molecule from another and identify the materials within the molecule. Based on the Raman spectrum, Raman spectral libraries are often used for identification of an unknown material – libraries containing thousands of spectra are rapidly searched to find a match with the spectrum of the analyte.

Raman spectrophotometers can be described as either dispersive or non-dispersive. Dispersive Raman spectrophotometers use a prism or grating, whereas non-dispersive Raman spectrophotometers use interferometers, for example the Michelson interferometer used in Fourier Transform Raman spectrophotometers. Until the 1960s, a 435.8 nm line of coiled low-pressure mercury lamp was used as the light source with traditional lasers later replacing the lamp (**Fig 2.4a)** [184]. At present there are a wide

range of available lasers that provide a more stable and intense beam of light. These include 488 and 514.5 nm Argon ion lasers, 530.9 and 647.1 Krypton ion laser, 632.8 nm Helium-Neon (He-Ne) and 785 and 830 nm Near Infrared (IR) diode lasers. Argon ion and Krypton, short wavelength sources, can produce significant fluorescence and photochemical decomposition leading to molecular breakdown. Diode lasers on the other hand, can be used at much higher power without causing any sample breakdown and they also reduce the levels of fluorescence in most cases [183].

Previously, thermoelectrically cooled photomultiplier tubes and photodiode array detectors were used in Raman instrumentation. Owing to advantages in technology, CCDs and charge-injection devices (CIDs) that have greater sensitivities have replaced these detectors. The scattered light, containing valuable spectral information is typically passed through a series of optical components including band pass filters or more frequently in dispersive instruments combined notch filters and high-quality grating monochromators are used. Such components ensure the light is directed towards the detector [182], [187]. The CCD detector for example, consists of an array of pixels organized in rows and columns. Each pixel can hold a specific amount of charge.

As the scattered light strikes the pixels, charges accumulate in each pixel based on the intensity of the incoming light. The position of the pixel corresponds to the spatial location of the incoming light on the detector. After the exposure time is complete, the accumulated charges in each pixel are read out sequentially. This process involves shifting the charges from pixel to pixel within each row and then transferring them row by row to an output register. The charges are then converted into electrical signals and

digitized. The digitized signals from each pixel correspond to the intensities of different Raman-shifted wavelengths. By analysing these intensities, a Raman spectrum is generated. Each pixel contributes to a specific Raman shift, and the overall spectrum provides information about the molecular vibrations and composition of the sample [185], [186], [187].

Nowadays, Raman spectrophotometers can broadly be classified into two classes: those that are lab based and in-situ spectrophotometers which include hand-held, portable devices. Both rely on the same basic principles and are differentiated only by their size, relative cost, and versatility.

2.7.1 Raman Spectroscopic Methodologies

Raman spectroscopic methodologies currently applied in literature range from used custom built Raman set-ups to lab-based commercial systems as outlined in **Table 2.2**. Most of the Raman set-ups identified in Raman-IBD studies, for example, exploited an excitation laser wavelength of 785 nm, except Addis *et al.,* who used 514 nm and Chemavaiska *et al*., a 532 nm laser [188], [189]. Due to the efficiency of Raman scattering, scaling inversely with excitation wavelength to the fourth power, there has been a movement towards shorter excitation wavelengths, which typically result in stronger signal, enhancing sensitivity and enabling detection of trace analytes.

However, shorter wavelengths are more likely to result in sample fluorescence, complicating sufficient Raman signal, detection, and analysis. Hence, the 785 nm excitation is the most frequently used in Raman-IBD studies since it reduces the

fluorescence. Furthermore, at this longer wavelength, the light is capable to penetrate deeper into thick or turbid samples thus, making it suitable for *in-vivo* or non-destructive analysis of biological tissue with the lower energy photons at near-infrared excitation less likely to cause photodamage to sensitive biological samples [190].

Across the Raman studies, a range of different laser powers and optics are employed, ranging from 10 mW by Smith *et al.* up to 90 mW by Morasso *et al.* [41], [191], [192]. The choice of laser power in RS requires establishing a balance between maximising the signal intensity and minimising the potential sample damage or interference. It depends upon the nature of the sample, the desired signal strength, and the potential for fluorescence. In the selected studies, typically those involving tissue samples, RS was employed using the lower powers and shorter exposure times than those involving plasma, urine, or faeces.

Table 2.2. Overview of publications applying Raman spectroscopy in diagnosis, treatment, or biomarker discovery of IBD.

Publication	Application	Sample Type	Patient Cohorts			Raman Laser	AI Classification	Performance		
			UC	CD	HC			Sensitivity (%)	Specificity (%)	Accuracy (%)
Pence et al., 2017 [193]	Diagnosis of IBD	Tissue	8	15	8	785 nm	SMLR	62-86.2	22.9-74.5	-
Smith et al., 2021 [191]	Mucosal healing and patient response to biologics	Tissue	426	328	-9	785 nm	SKiNET	96.3	95	95.6
		Tissue	6	8	9			96.2	88	91.6
Acri et al., 2020 [194]	Diagnosis of IBD in paediatric patients	Faecal	9	15	19	785 nm	ROC, Youden Index	-	-	-
Bielecki et al., 2012 [40]	Discrimination of IBD	Tissue	13	14	-	785 nm	PCA, K-means clustering	93.84	90.3	88.26
Kirchberger-Tolstik et al., 2020 [195]	Assessment of disease severity	Tissue	140	-	-	785 nm	1D-CNN	78	93	-
Ding et al., 2017 [196]	Determination of histological inflammatory status in vivo	Tissue	18	-	31	785 nm	Statistical analysis - Tukey honest significant difference	83.5	97.1	-
Addis et al., 2016 [188]	Assessment of mucosal healing	Tissue	60	-	-	514 nm	Linear regression analysis	-	-	-
Tefas et al., 2020 [197]	Diagnosis of UC	Plasma	28	-	35	785 nm	PCA-LDA	86	92	89
							PLS-DA	89	94	92

Morasso et al., 2020 [41]	Diagnosis of CD	Plasma	-	77	45	785 nm	PCA-LDA	80	85.7	83.6
Li et al., 2021 [198]	Diagnosis of CD	Urine	-	98	45	785 nm	PCA-LDA	-	-	69.9
Zhu et al., 2023 [199]	Perianal fistula in CD patients' identification for treatment strategies	Urine	-	110	-	785 nm	PCA-SVM	71.43	80	75.71
Bi et al. 2011 [200]	Discrimination of IBD	Tissue	12	9	-	785 nm	-	-	-	-
Veenstra et al., 2014 [201]	Diagnosis of UC	Tissue	4	-	-	785 nm	PCA	82 87	89 93	
Wu et al., 2022 [202]	Discriminating active CD from inactive CD	Urine	-	100	88	785 nm	PCA-SVM	-	-	63.6
Chernavaiska et al., 2016 [189]	Identification of disease severity	Tissue	6	7	7	532 nm	Multimodal Imaging	-	-	-
Buchan et al., 2023	Identification of potential biomarkers of IBD	Saliva	26	25	50	785 nm	SKiNET	-	-	98.2

For instance, Pence *et al.* set an exposure time of 250 ms whilst Addis *et al.* applied a 60 s exposure to faecal samples, with the former stating they were unable to increase integration times due to a substantial background noise and autofluorescence [188], [193]. Choosing an adequate laser power is known to improve the signal-to-noise (SNR), which is in turn, central to obtaining high-quality spectra. The higher the SNR, the greater the overall accuracy and precision of spectral measurements and data analysis. Laser power also influences the depth from which Raman signals can be obtained [192]. Higher power generally penetrates deeper into the samples, allowing for the analysis of subsurface layers of thicker samples, particularly important for applications with *in-vivo* tissue analysis, as observed by both Pence *et al.* and Ding *et al.,* who examined colon tissue *in-vivo* using portable, custom designed Raman systems [193], [196]. Higher laser power can, however, lead to sample damage particularly for sensitive, biological samples. This photodamage or degradation can occur, leading to sample alteration and the introduction of artificial carbon D and G spectral bands. Optimisation of Raman-setup, therefore, requires careful empirical testing and adjustments for each unique experimental setup and sample type.

2.7.2 Raman Spectral Pre-Processing

Generally, in Raman-IBD studies, a form of spectral pre-processing is applied including for instance, baseline subtraction, normalisation, or cosmic ray removal. Firstly, in many Raman spectrometers baseline shifts exist whereby the generated spectra contain both the desired signals as well as the more undesirable elements including background noise. Many methods of baseline subtraction can be applied such as, asymmetric least squares (AsLS), polynomial fitting, differencing, and filtering with the chosen method highly dependent upon the level of precision required, patterns and obtained computational times [203], [204]. Polynomial fitting methods for example, are a simple and effective way to remove baseline artefacts caused by factors such as instrumental drift, scattering, sample thickness or sample inhomogeneity. The method involves manually identifying the areas in the spectrum that contain undesirable or non-Raman like locations and estimating this background by way of a polynomial function. The degree of polynomial is dependent on the overall complexity of the baseline with the most commonly applied fits being linear (1st degree), quadratic (2nd degree) or cubic (3rd degree) polynomials. The selected fit is then applied to spectral areas with no features of interest and subsequently subtracted from the raw spectrum point-by-point leaving behind a baseline-corrected spectrum [204], [205]. Polynomial baseline subtraction enhances the clarity of spectral features and aids in the interpretation and analysis of spectral data. AsLS on the other hand, combines a smoothing filter with the asymmetric weighting deviations from the determined smooth trend in order to develop an effective baseline subtraction method [203], [206]. This method has the drawback associated with the smoothness constraint in that it only considers the second derivative. In most studies, baseline subtraction is performed

using a high order polynomial fit (4^{th} or 5^{th}) whilst, Kirchberger-Tolstik *et* al. employed an AsLS method and Ding *et* al. used a cubic spline interpolation [195], [196].

Subsequently, normalisation is routinely performed to account for intensity variabilities due to factors such as sample concentration variability, laser power fluctuations and differences between sample set-ups [207], [208]. Commonly employed normalisation methods in RS include, scaling - where the minimum intensity is set as 0 and the maximum as 1 and all values scaled proportionally, Z-score normalisation - where the data has a mean of 0 and a standard deviation of 1 in order to centre the data and vector normalisation with the spectrum being scaled by dividing it by its own vector length, ensuring that the normalised spectra have a final magnitude of 1 [207]. Normalisation using area under the curve has been used by Addis *et* al*.,* with five studies using vector normalisation [36], [40], [195], [198].

In addition to baseline subtraction and normalisation, smoothing is often used to overcome the inherent issue of noise in Raman spectra. Smoothing is used to remove high frequency components from the spectra to reduce noise and simplify identification of important spectral features [209]. The most used smoothing filter is the Savitsky-Golay filter, where a polynomial regression is applied to the data, *i.e.,* a polynomial curve is fitted to small segments or peaks within the Raman spectrum with the smoothed curve providing an estimate as to the underlying Raman signal [209], [210]. The use of a Savitsky-Golay filter for spectral smoothing was employed by three of the studies [188], [193], [196] with only Addis *et al.* indicating the filter smoothing width of 9 and polynomial order of 3 [188]. Due to the differences in spectral acquisition and

Raman setup, each study required the pre-processing techniques used to be applied in a specific order. The most applied procedure included baseline subtraction, followed by normalisation leading to data classification (**Fig. 2.3**).

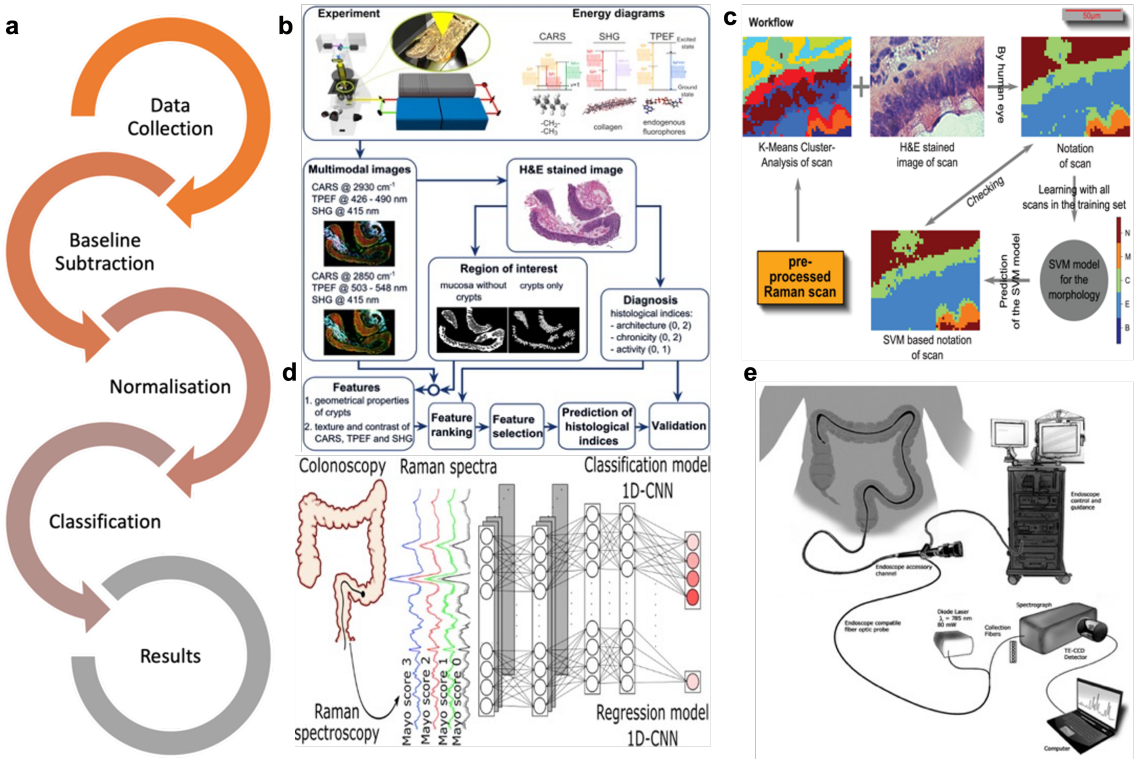

Figure 2.5. (a) Overview of a typical procedure for Raman data analysis and processing. **(b)** Workflow used in *real-time* endoscopic assessment of IBD disease activity including initial multimodal imaging of CARS, TPEF, SHG and H and E staining closely examined by pathologist and feature extraction of regions of interest to discriminate both morphology and intensity related to diagnostic criteria [189]. **(c)** Overall work path in the classification of IBD based on imaging of epithelium cells consisting of *k*-means clustering of morphological features, followed by SVM learning and subsequent prediction *via* a 10-fold cross validation with comparisons made to those identified by the pathologist as well as with H and E staining [40]. **(d)** Schematic

representation of the methodology used by Kirchberger-Tolstik *et* al. [195] comparing Raman spectra to Mayo endoscopic scores in UC [195]. **(e)** Illustration of an integrated Raman set-up and endoscope instrumentation for *in-vivo* analysis used by Pence *et* al. [193].

2.7.3 Classification *versus* Molecular Characterisation

Typical Raman healthcare studies identified in the field of Raman-IBD predominantly focus on using RS for either classification or molecular characterisation of spectral data. Whilst both are essential for the comprehensive understanding of IBD and its potential translation into the clinical fields, they each serve a different purpose and apply diverse methodologies. Raman characterisation focusses on identifying and understanding the molecular composition and the associated chemical properties of the biofluid or tissue under investigation. This approach classically involves detecting and quantifying specific molecules, chemical bonds, or structures within the sample. Characterisation, aims to provide detailed information as to the sample's biochemical makeup, identifying biomolecules including lipids, proteins, and nucleic acids whilst also, assessing changes to the molecular constituents such as, protein structure and confirmation or lipid profiles in samples where disease is present. Commonly, the methodology involves using RS to obtain spectra, which serve as chemical fingerprints with subsequent analysis of spectral peak intensities and Raman shifts used to identify specific molecular vibrations and chemical bonds thus, allowing for the determination as to the composition and structural characteristics of the biofluid. This approach was employed by Ding *et* al., Addis *et al.,* Chernavaskia *et* al., Bi *et* al. and Wu *et* al. [188], [189], [193], [196], [200], [202], where the authors applied statistical methods including,

the parametric (*t-test*). Shapiro Wilk and Kolmogorov-Smirnov statistical analysis was employed by Wu *et* al.*,* whilst Addis *et* al. applied the statistical Mann-Whitney U test and Ding *et* al. the ROC curves investigation. These outputs provided disease-related biochemical changes associated with IBD.

Conversely, RS classification focusses on using the spectral information obtained from the analysis to classify biofluid or tissue samples into different categories or states such as healthy or diseased. Such classification usually involves the development of models or algorithms capable of differentiating between the data classes. Machine learning techniques are most routinely used to train the classifiers and label the data sets accordingly. Of the studies identified, ten have [36], [191], [193], [194], [195], [197], [198], [199], [201], [211] applied a variation of multivariate analysis / machine learning to train their models using spectral data from a set of known biofluid samples, including primarily healthy and diseased classes, mucosal healing and inflammation or disease severity. Several further researchers have developed and applied further forms of supervised machine learning approaches including, the artificial neural networks (ANNs) by Smith *et* al., Buchan *et* al. and Kirchberger-Tolstik *et* al*.* [191], [195]. Other studies harnessed dimensionality reduction and feature extraction methods such as, principal component analysis-linear discriminant analysis (PCA-LDA) [36], [197], [198], partial least squares-discriminant analysis (PLS-DA) [197] or principal component analysis-support vector machine (PCA-SVM) [199], [202]. In each instance, the goal was to provide a diagnostic tool which could rapidly and accurately identify the presence or absence of IBD based on the obtained spectral signature.

Overall, both approaches complement one another and are vital to gain a comprehensive understanding as to the disease development and diagnosis using RS.

2.7.4 AI / Artificial Intelligence/Artificial Neural Network Multivariate Data Classification

The majority of Raman-IBD studies employ multivariate analysis to investigate spectral changes across the whole spectrum. Notably, Ding *et al*. and Acri *et al.* applied somewhat different approaches using the Tukey honest significance variance and receiver operator curve (ROC), respectively [194], [196]. Multivariate analysis of RS data enables powerful extraction of meaningful information from complex high-dimensional, spectral datasets with multiple variables [212]. It also enables the detection of subtle spectral shifts, which would be otherwise difficult to differentiate using traditional methods such as, manual peak selecting approaches. Frequently applied principal component analysis (PCA) can reduce the dimensionality of data while preserving the relevant information. Techniques such as, linear discriminant analysis or support vector machines employed by Tefas *et al.,* Li *et al.* and Morasso *et al.,* can classify samples into predefined groups such as, IBD and healthy controls, based on their spectral patterns, which is particularly useful in sample identification and disease diagnosis [36], [197], [198]. Approaches such as partial least squares (PLS) regression, used by Tefas *et al.,* are also capable of establishing quantitative relationships between spectral data and sample properties, valuable for the accurate and rapid quantification of biomarkers in a dataset [197]. Machine learning, a subfield of artificial intelligence (AI), is another approach applied in studies by Smith *et al* and Buchan *et al.*[191]. In these studies, a novel supervised machine learning algorithm

uses an artificial neural network inspired by the structure and functioning neural networks (ANNs), such as the brain. They consist of interconnected nodes, termed artificial neurons or nodes, organised into layers [213]. The power of ANNs lies in their ability to learn complex patterns and representations from data thus, presenting a valuable decision support tool to aid in medical diagnosis. It is based on self-organising maps (SOMs) with self-optimizing Kohonen index network (SKiNET) as a framework for multivariate analysis that simultaneously provides (i) dimensionality reduction, (ii) feature extraction and (iii) multiclass classification. SKiNET performs visual separation to identify the underlying chemical differences between classes, providing accurate classification for simultaneously rich-information and high-classification specificity even for low laser powers and short acquisition times, representative of the real-world point-of-care conditions.

Machine learning applied in these studies is essential for the clinical translation of RS based disease classification, enabling accurate, automated, and scalable disease diagnosis whilst enhancing objectivity and reproducibility. It can also be integrated into RS systems and methodologies as well as clinical workflows thus, improving time to results as well as patient care and long-term outcomes. It also plays a crucial role in realising the full potential of RS in clinical applications and beyond. Although the focus here is on IBDs, further afield similar approaches have been used in cancer diagnostics where for instance, Hernándes-Vidales *et* al. used PCA-SVM to distinguish between biomarkers of cancer with a high accuracy of 94%, highlighting its suitability for the investigation of various forms of cancer [214]. Similarly, Mehta *et* al. applied PCA-LDA to differentiate blood serum from control and meningioma patients, yielding 92%

classification accuracy [215]. This was further expanded to cervical cancer diagnosis, where Daniel *et* al. used Near-IR RS combined with ANN to determine biochemical changes associated with cancerous cells, successfully achieving 99% accuracy [216].

2.8 Applications of Raman Spectroscopy in Diagnostics

This section reviews both *ex-vivo* and *in-vivo* applications of RS, with an emphasis on studies analysing saliva and/or IBD. Although our interest pertains to the use of RS in *ex-vivo* evaluations current trends in the field lean towards the use of surface enhanced Raman spectroscopy (SERS) for many sensing applications. Spectroscopic 'fingerprints' of Raman data have been successfully shown to identify the disease states of the subject from which the biofluid has been collected and in contrast to the in-vitro bioassays, the availability of inexpensive, portable Raman instruments makes this technique particularly attractive for point-of-care, rapid sampling, analysis and screening of biofluids and tissue. It is important to note that although the ability of RS to rapidly and accurately identify target molecules renders it a valuable technique, analysing biological samples introduces an added layer of complexity. Since biological systems and samples are typically composed of a broad array of biochemicals including proteins, carbohydrates, lipids etc., spectra from biological specimens, such as tissue and biofluids, may appear highly similar, as vibrations from each of these molecules will be present in the Raman spectra [36], [41], [217], [218]. Analyses of these can subsequently, not be straightforward to untrained personnel. Therefore, the combination of Raman spectroscopy with subsequent computational algorithms enables it to rapidly and accurately decompose the signal or separate the data

according to statistical properties inherent in the dataset using various multivariate techniques, discussed in more details below.

RS, although a powerful technique in analysing various biofluids, is associated with some limitations. These include fluorescence interference, known to occasionally overwhelm the Raman signal and obscure spectral signatures, sample complexity - where untangling the Raman spectra of complex biological samples can be challenging, the cost of specialised equipment, and low sensitivity due to Raman scattering being an inherently weak process thus rendering it difficult to detect low-concentrations of biomolecules [219]. Nevertheless, RS is continuously optimised and remains a valuable tool for analysis of biofluids. Ongoing research aims to address these challenges *via* a combination of advancements in instrumentation, data analysis techniques and the development of enhanced Raman probes and substrates. One such development is known as surface enhanced Raman spectroscopy (SERS).

With the advent of SERS and the subsequent enhancement of the Raman signal of up to 10^{11} times, single molecule detection can be achieved [220], [221]. SERS typically uses nanoscale surfaces of gold or silver to induce a highly localised electromagnetic field by surface-confined, laser excitations [222], [223], [224]. A substantial enhancement in the Raman signal is observed when a molecule is absorbed or within 10nm of the surface. For instance, analysis of plasma *via* conventional Raman spectroscopy reflects different proteins yet, the smaller molecules remain almost undetectable, however, *via* SERS a considerably amplified signal is produced enabling detection of *minute* biomolecule levels [198], [225], [226], [227], [228], [229], [230].

SERS, therefore, exhibits several advantages including high sensitivity and selectivity with the ability to detect analytes down to attomolar level. It further has a relatively low sensitivity to water, rendering it suitable for *in-situ* and *in-vitro* applications of biological samples. Overall, SERS is well-suited to an extensive scope of applications ranging from surface chemistry, catalysis, food science and pharmaceuticals [225], [228], [230], [231], [232], [233], [234]. However, despite these advantages SERS has often been underutilised due to challenges in reproducibility, high costs associated with development of suitable substrates and the need for highly specialised equipment and expertise [222], [225].

2.8.1 Raman Spectroscopy and Saliva in Healthcare

A problem in healthcare research is the non-specificity of biomolecules, which is debatably compounded by a base of small studies that focus on specific diseases [235]. For instance, in the context of traumatic brain injury, certain biomarkers may also be indicative of polytrauma or neurodegenerative disease [236], *e.g.,* Tau, neurofilament light chain, myelin basic protein. For this reason, the full extent of the injury and/or co-morbidities should be noted. In general, this would appear not to be the case within current Raman-saliva studies, many of which are preliminary in nature. Moreover, there is the broad difficulty of patients' data security, permissions, and ethics [235]. Elsewhere, problems may arise in a lack of specificity to the locale of disease, *e.g.,* sialic acid (SA) - a cancer biomarker, is detected in many different types of cancers. Furthermore, healthcare related studies tend to be concerned with changes to Raman peak intensities, or more precisely, relative peak intensity ratio changes rather than the appearance of distinct spectral features. This can make analysis

challenging, even with supporting chemometric tools, especially if physiological changes are subtle. In turn, medical and clinical studies can be difficult to power correctly to reach statistical significance [235], [237], [238]. This is exacerbated by the difficulty of obtaining large sample numbers for the various studied diseases and much of the current biomarker research, at large, is still restricted to small cohorts [235].

With 19.3 million new cancer cases and almost 10.0 million deaths in 2020, cancer is one of the most prevalent and fatal diseases worldwide [239]. With an estimated 2.3 million cases (11.7%), female breast cancer has surpassed lung cancer as the most diagnosed cancer globally, followed by lung (11.4%), colorectal (10.0%), prostate (7.3%), and stomach (5.6%) cancers [239]. By 2040, the global cancer burden is projected to be 28.4 million cases, a rise of 47% from 2020. The diversity and prevalence of cancer highlights the need for a global escalation in cancer diagnostics, distribution of preventative cancer measurements as well as providing cancer care in those transitioning countries to control the disease [239]. Raman spectroscopy is emerging as a potentially useful clinical aide in cancer diagnostics. RS can identify and differentiate, in a label-free and non-invasive way, the subtle changes in tissue biochemistry associated with disease and cancer progression. The close association of salivary composition to serum metabolites, combined with the wide spectrum of molecules present in saliva provides valuable information for clinical diagnostic applications, of which the main Raman peak assignments relating to cancer diagnostics are illustrated in **Fig 2.6** with further spectral assignments made in **Table SA.3**. Several studies have attempted spectral analysis of biofluids using Raman spectroscopy for disease diagnosis and screening as it may represent a prospective

viable avenue for cancer diagnosis (**Fig SA.2 and SA.3)**. Saliva has been investigated as a diagnostic medium and its analytes (protein, RNA, DNA) as biomarkers in the diagnosis of several malignancies including lung cancer, breast cancer, prostate cancer, ovarian cancer and squamous cell carcinoma.

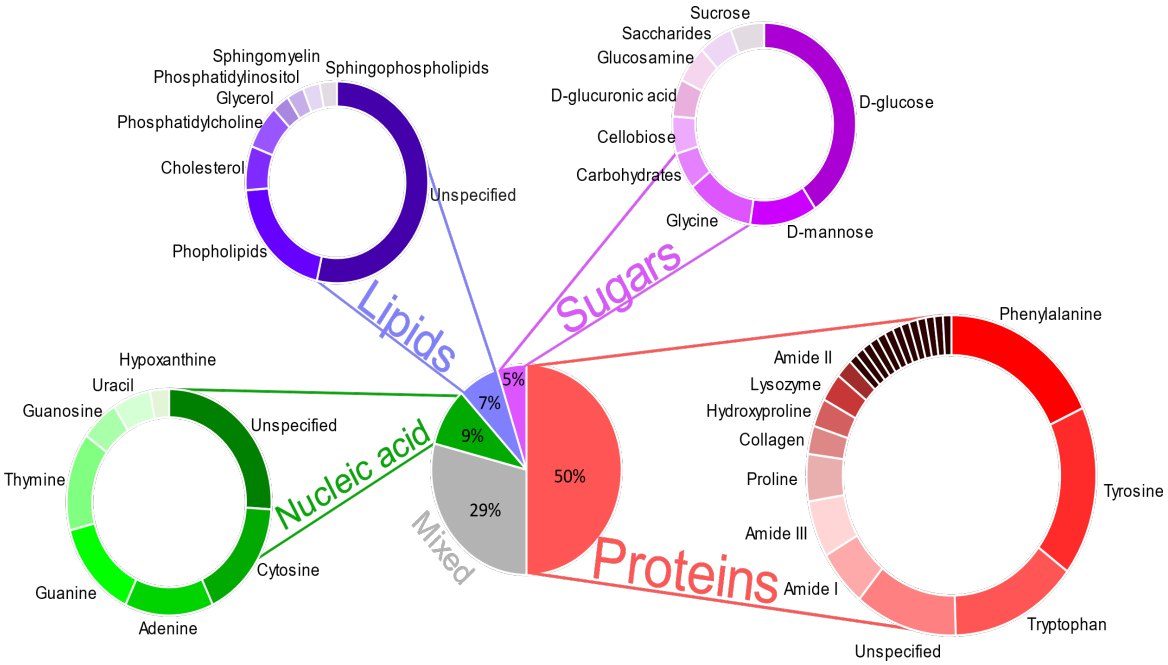

Figure 2.6. Raman peak assignment on cancerous saliva samples (*n* = 732). An overview pie chart of the most prominent peaks in the literature displayed by their biomolecular group. A subset of the Raman-saliva literature with a perceived focus on chemometrics has been analysed. Further details are included in Supporting Information, **Figs. SA.1-SA.2 and Table SA.3**.

Breast cancer is the most diagnosed cancer and the leading cause of death in women. Routine diagnostic procedures include histopathological examination of biopsy samples, fluorescence, optical bioluminescence, X-ray mammography, computed

tomography, magnetic resonance imaging and ultrasound, which are often time-consuming and costly. Issues include low sensitivity and resolution [240]. To overcome this, Wu *et* al. developed a non-invasive breast tumour detection method using saliva protein SERS combined with regularized multinomial regression (RMR). Their study indicated diagnostic accuracies of 92.7% (85/97), 95.8% (93/97) and 88.6% (86/97) were acquired, while successfully discriminating among the normal group, the benign breast tumour group and the malignant breast tumour group [241]. Likewise, Feng *et* al. also exploited the capability of saliva protein analysis, based on membrane purification and SERS for detecting benign and malignant breast tumours. The authors detected subtle but discernible changes in the mean SERS spectra in each of the three groups. Multiclass partial least squares discriminant analysis was used to analyse and classify the saliva protein SERS spectra from healthy subjects, benign breast tumour patients, and malignant breast tumour patients, yielding diagnostic sensitivities of 75.7%, 72.3% and 74.1% as well as specificities of 93.7%, 81.2% and 86.3%, respectively [242].

Furthermore, Hernández-Arteaga *et* al. indicate the importance of SA in the diagnosis of breast cancer [243]. The authors evaluated the feasibility of Raman as a method of quantification of SA in saliva using citrate-reduced silver nanoparticles as a SERS substrate. They determined the mean SA concentration was significantly higher among breast cancer patients (18.3 ± 9.4 mg/dL) than in the healthy control group (3.5 ± 1.0 mg/dL) with SERS test sensitivity of 94% and specificity of 98% [243]. In a further study, the authors once again measured SA levels using SERS with tailored citrate-

reduced silver nanoparticles, however exhibiting a lower sensitivity and specificity of 80% and 93%, respectively [243].

In addition, lung cancer remains the leading cause of cancer mortality, leading to more than 1 million deaths worldwide in 2020. Clinical studies indicate that approximately 80% of patients are diagnosed with advanced lung cancer and with a 5-year survival rate of only 20% [244]. Thus, effective screening and early detection have the potential to decrease lung cancer mortality. Wang *et* al. demonstrated early detection technology of lung cancer based on SERS. They analysed 19 lung cancer and 45 healthy patient samples and obtained an accuracy of 96.9% by Logistic Regression Analysis. The authors also detected two prominent peaks of interest, 758 cm^{-1} and 1244 cm^{-1}[245] . Similarly, Li *et* al. applied SERS for the detection of lung cancer using silver colloids. The peak variations between the healthy and diseased groups were investigated. The authors determined that lung cancer peaks were largely detected at a lower Raman intensity than those of their healthy counterparts. Principal component analysis (PCA) and linear discriminant analysis (LDA) were used to discriminate between the data sets resulting in an 80% accuracy, 78% sensitivity and 83% specificity [79]. Qian *et* al. also applied SERS in their new method of lung cancer detection. The authors combined a SERS system with a gold-modified nanochip. Support vector machine and random forest algorithms were used to analyse differences between healthy and diseased data sets. Sensitivity and specificity of Raman spectroscopy data was 100% [246].

As well as cancer detection, salivary Raman methods have found applications within many other healthcare studies. For example, Cao *et* al. developed a potential method for the non-invasive detection of acute myocardial infarction (MI). The authors observed significant differences between healthy and MI Raman bands, yielding a diagnostic sensitivity of 80.4% [247]. Multiple studies have focused on the diagnosis of Sjögren's syndrome, *e.g.,* Moisoiu *et* al. combined SERS of saliva with two-dimensional shear wave elastography (2D-SWE) of the parotid glands whereas Stefancu *et* al., applied SERS based liquid biopsy of saliva from patients with Sjögren's syndrome, with silver nanoparticles used as SERS substrate and the data analysed using PCA-LDA to achieve an overall higher classification accuracy of 94% [228], [248]. In addition, Eom *et* al. used SERS in the diagnosis of Tamiflu-resistant influenza virus [249]. The authors combined SERS-active substrates with urchin gold nanoparticles and oseltamivir hexylthiol, an excellent receptor for the oseltamivir-resistant pandemic H1N1 (pH1N1)/H275Y mutant virus. A further application was identified by Ember *et* al. in the detection of COVID-19. The authors developed a reagent-free method of detecting COVID-19 combining Raman spectroscopy, machine learning and droplet segmentation to achieve sensitivities of 79 and 84% in males and females and specificities of 75 and 64%, respectively [42]. Carlomagno *et* al. aimed to discriminate amyotrophic Lateral sclerosis (ALS) onset using Raman spectroscopy to provide a fast, sensitive procedure to efficiently diagnose and monitor the therapeutic and rehabilitative processes [250]. Saliva and Raman have also found applications in screening for Alzheimer's Disease (AD). Ralbovsky *et* al. combined Raman hyper-spectroscopy with machine learning to develop a novel method for the diagnosis of AD based on salivary analysis, indicating its remarkable potential for use

as a non-invasive, efficient, and accurate (99%) method [43]. Further applications have been identified as a non-invasive alternative procedure for early detection in childhood asthma biomarkers [251], analysis of thiocyanate in saliva using droplet SERS-microfluidic chip [229] and for the detection of *Pseudomonas aeruginosa* in saliva using SERS [252].

2.8.2 Raman and IBD Applications in Healthcare

Early and accurate detection of IBD is crucial in providing appropriate interventions. It can further facilitate monitoring the progression of disease as well as the therapeutic responses. However, timely diagnosis of IBD still remains an unmet need. With an overlap in presentation, symptoms and disease progression, discrimination of UC from CD is essential when selecting the most appropriate therapeutic or surgical regimen to improve patient prognosis. Herein, the predominant focus is on the potential of Raman spectroscopy in the diagnosis and differentiation of IBD as well as the disease indicative biomarkers. We overview the different Raman methodologies currently applied within clinical IBD research as well as their ability to act as monitoring and therapeutic tools in the future. Additional synopsis includes the use of RS in the identification of biomarkers and how these may contribute to the non-invasive and early detection of IBD. An in-depth summary of the primary methodologies applied in Raman-IBD analyses along with a comprehensive spectral library with corresponding peak assignments from the current Raman-IBD literature are further overviewed and generated. These lay a platform for introduction of new tools for advancing the understanding of the overall biochemical changes of IBDs and development of timely diagnostic and therapeutic approaches. Furthermore, it would enable improving the

clinical management of patients as in the case of an evolving and multifaceted pathology such as IBDs, RS may provide integrative and complementary clinically useful information, or be useful at different time points or in different settings.

2.8.3 IBD Diagnosis and Discrimination

Raman spectroscopic based approaches have been employed in a wide array of diagnostic applications, aiming to provide objective and *real-time* assessments of diseases [43], [79], [253], [254]. There has been a notable increase in interest surrounding the utilisation of Raman analysis for the study of IBDs in recent years indicating that RS can enhance diagnostic accuracy due to the detailed molecular information it is capable to uncover in a rapid manner. In the field of IBDs, whilst several researchers aimed to identify either CD or UC from patient samples, [40], [200] others aimed to investigate and discriminate both UC and CD from healthy patient samples [189], [191], [193], [194] **(Fig.2.7)**.

One of the earliest studies published in this field by Bi *et* al. examines the structural and compositional changes of both UC and CD colon tissue specimens to differentiate between the two IBD subtypes. The authors used a custom-designed Raman fibre-optic probe to analyse the tissue samples *in-vitro.* The probe was placed in contact with the mucosal surface of the sample for 3s to obtain spectra [200]. After statistical analysis, closely examining both the spectral peak differences and molecular origins, a significant difference was observed between the nucleic acid, phenylalanine, and lipid regions of the spectra. Assignments for the identified Raman bands are summarised in **Table SB.1**. This study importantly demonstrated a higher lipid content

with a lower phenylalanine and nucleic acid content in UC tissue samples, indicating the potential of these characteristics in disease discrimination. The significant variation noted in lipid content is most likely due to CD, often presenting with transmural inflammation, which involves all layers of the bowel wall, including the submucosa and therefore, brings about changes in the lipid content due to inflammation, fibrosis, and the formation of granulomas. UC on the other hand primarily involves inflammation of the mucosal layer thus, the submucosal changes are generally less pronounced than in CD. To quantify and compare these changes, further studies such as, histological examination or lipidomic analyses would be required. Limitations of this study include the exclusion of healthy controls and the use of statistical chemometric methods only. The inclusion of multivariate analysis could further offer the advantages of dimensionality reduction, pattern recognition and effective visualisation.

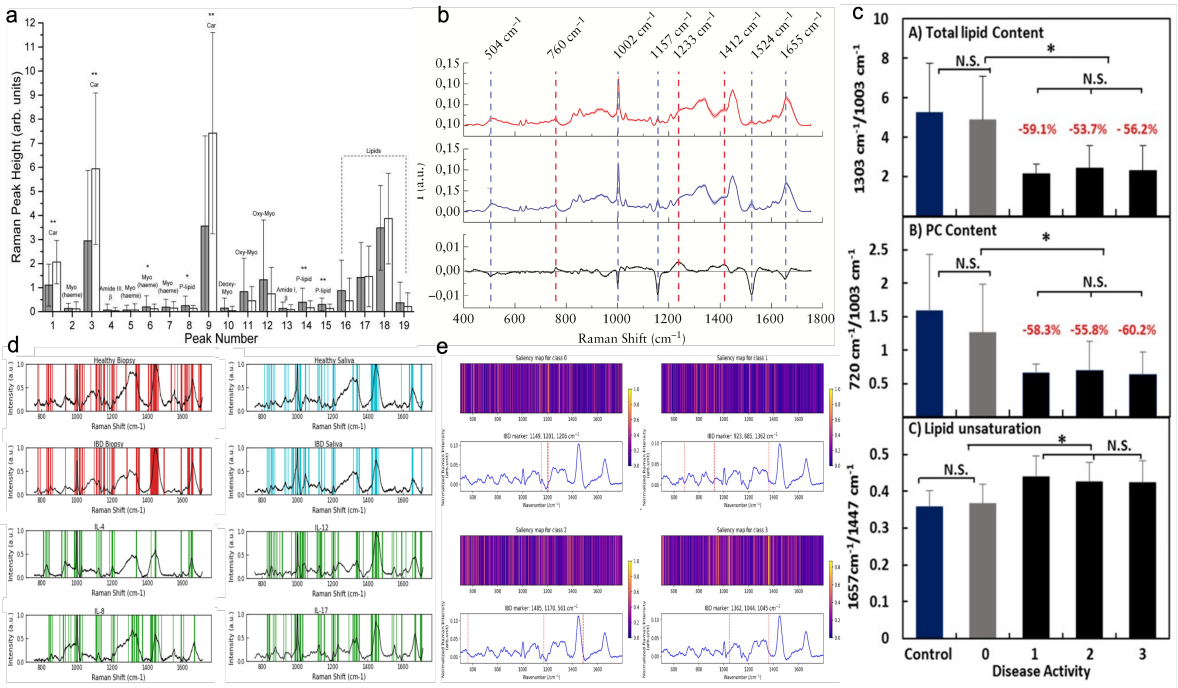

Figure 2.7. IBD discrimination *via*: **(a)** Average peak intensities for areas of endoscopic inflammation or endoscopic healing with highlighted peaks of interest indicating

standard deviations and *p*-values. (*p** 0.10 ≥ p > 0.05 and *p*** ≤ 0.05) [188]. **(b)** Average Raman spectra obtained from patients with CD and healthy controls. Top spectra (red) indicate CD and middle spectra (blue) healthy controls with bottom spectra (black) illustrating the spectral differences between the two sample groups. Dotted lines highlight significant areas of difference between the sample types [36]. **(c)** Differences in total lipid counts and phosphatidylcholines (PCs) relative to proteins as determined by peak height ratios for differing disease severities [196]. **(d)** Barcodes derived from tissue biopsy, saliva and potential IBD biomarkers (e) IL-4, (f) IL-12 (g) IL-8 and (h) IL-17, highlighting the dominant, statistically significant peaks. **(e)** Saliency map visualisations (top) and mean Raman fingerprint regions corresponding to Mayo endoscopy score. A high saliency score on the map is indicated in yellow and represents Raman regions with important contributions in the determination of score classification. Most regions are attributed to proteins, cholesterol, and DNA [195].

Bielecki *et al.* reported a Raman microspectroscopic approach to diagnose IBD. The authors firstly collected Raman measurements and morphological images (obtained *via* haematoxylin and eosin staining) from UC, CD, and healthy control (HC) specimens [40]. Subsequently, a unique classification methodology, support vector machine, capable of analysing histological and spectral data to identify and visualise the tissue morphology and ultimately differentiate the epithelial structures from connective tissue and mucus, was employed. This approach demonstrated significant differences between each of the three groups (UC, CD and HC) with the classifier achieving an accuracy of 98.9%, paving the way for the automatic and objective classification of IBD *via* RS even in patients with minimal or moderate inflammation. Significant changes

were observed in the heme bands, which were associated with increased levels of inflammation in IBD patients compared to HC. Additionally, UC is distinguished by a distinct hyperaemic colonic mucosa, which differs from the observed in CD. The fully automated methodology established by the authors has the potential to be used *in-vivo*. It also paves the way for further investigation using unclassified IBD samples and perhaps, patient samples of varying inflammatory status.

Pence *et* al. exploited the lack of a gold standard in discriminating UC from CD and developed a novel colonoscopy-coupled fibre optic Raman probe as a minimally invasive diagnostic tool. The authors examined spectral signatures from UC, CD and HC patient tissue samples. All samples were compared with tissue pathology markers and assessed using endoscopic examination to confirm IBD subtype and severity [193]. To separate spectra into their respective groups, a sparse multinomial logistic regression (SMLR) algorithm was applied with variable classification performance of 62-86.2% for sensitivity and of 22.9-74.5% for specificity, with the most successful comparison for control *versus* active IBD tissue types. A poorer classification performance was observed when comparing healthy controls and inactive IBD subtypes, indicating active inflammation playing a key role. The study built on previous work where it was recognised that increased contact pressure can lead to spectral differences with attempts made to reduce this effect [255]. Overall, the authors demonstrated that patients with inactive disease exhibited significantly stronger spectra, particularly in bands associated with lipid features. Active disease spectra indicated the broadening of peaks, often associated with an increase in protein content, and possibly originating from increased levels of fibrin and collagen in the bowel wall.

The authors also noted the potential impact of age, gender, BMI, diet, and previous therapeutic treatments, which most likely impact the classification algorithm. Also, as the probe was in direct contact with the colon, blood and mucus were still present thus, leading to bias in the obtained spectra.

Several studies aimed to diagnose either CD or UC with Veenstra *et* al. demonstrating one of the first such applications. The authors collected UC and HC tissue samples and examined the mucosal and serosal surfaces, achieving a classification sensitivity of 82% and specificity of 89% from the mucosal surface and 87% sensitivity and 93% specificity from the serosal surface [201]. Their study indicated the possibility of *real-time* diagnosis with the ability to observe changes within UC colon tissue which are not apparent in histological examination. Spectral areas of interest revealed significant changes in peaks assigned as proteins, lipids, and nucleic acids due to the increase in levels of inflammation. Moreover, Tefas *et* al. applied a sub-category of RS known as, surface enhanced Raman spectroscopy (SERS) to provide a significantly improved, non-invasive method for UC diagnosis. Plasma samples were collected from UC and HC subjects and analysed using silver nanoparticles as a plasmonic substrate to significantly enhance signal and detection [197]. The primary advantage of SERS over conventional RS is in its ability to offer a significant increase in Raman intensity, leading to the identification of molecules not previously observed in Raman spectra, however, SERS results in a more complex spectral interpretation particularly, in the case of biological samples, which contain a wide array of metabolites as interferants. Two multivariate classification algorithms were employed for spectral analysis, including principal component analysis-linear discriminant analysis (PCA-LDA), achieving 86%

sensitivity and 92% specificity and partial least squares discriminant analysis (PLS-DA), achieving 89 and 94% sensitivity and specificity, respectively. In UC, there is a known disruption in the gastrointestinal mucosa, which typically maintains the functional status of the bowel[256]. This disruption results in widespread inflammation and thus, a change in metabolic pathways observed *via* changes in Raman or SERS spectra and therefore, enables an insight into the mechanisms of IBD, facilitating the potential development of new diagnostic and therapeutic measures. The major drawback of the study is in the overall sample size as well as the exclusion of patients with mild disease. However, it does represent a promising non-invasive approach with the ability to act as a screening tool for colorectal cancer in the longer-term.

Similarly, Li *et* al. demonstrated the non-invasive diagnosis of CD based on SERS combined with a principal component analysis-support vector machine (PCA-SVM) approach. Urine samples were collected from patients with inactive and active forms of CD as well as HC subjects and combined with silver nanoparticles to form a SERS substrate for analysis. Both PCA-SVM and PCA-LDA were employed to establish classification models to distinguish between CD and HC sample types [198]. Their PCA-SVM algorithm reached a higher classification accuracy than PCA-LDA of 82.5% *versus* 69.9% between CD and HC groups, respectively. In terms of disease severity, the classification accuracy was higher when comparing active CD/HC subjects than inactive CD/HC subjects at 86.8% and 76.5%, respectively. These findings indicate the potential of RS to effectively identify metabolic changes in patients' urine and to monitor the progress and recurrence of disease.

Morasso *et* al. applied a similar approach to classification through a combination of RS and PCA-LDA in the diagnosis of CD. Dry plasma samples from CD and HC subjects were analysed with spectral differences between the two groups determined as well as those differences between CD patients with differing disease pattern [36]. The developed PCA-LDA model classified CD and HC subjects with a sensitivity of 80% and specificity of 85.7%. Biochemical variations were identified between the CD and HC sample types with the primary difference arising in the carotenoids, compared to the HC, CD spectra, which presented less intense peaks. Further differences were found with respect to β-sheet secondary structure of proteins and between the lipid and aromatic amino acid bands. This data combined, indicates the systemic metabolic alterations present in patients with CD, where for instance, the altered levels of lipids could indicate malnutrition due to reduced food intake to alleviate abdominal discomfort or due to the decreased absorptive surface of the bowel. Moreover, the value of this study lies in its ability to find a use as a screening tool in the diagnosis, or exclusion of CD. At present, no accurate blood biomarkers of IBD are available and those described markers, including anti-Saccharomyces cerevisiae antibody (ASCA) and perinuclear antineutrophil cytoplasmic antibodies (pANCAs), suffer from poor sensitivities and therefore, many patients undergo invasive colonoscopy procedures to rule out CD. A multivariate analysis of blood plasma could potentially present an exciting advance in the development of an effective first-line screening tool for the discrimination of CD.

2.8.4 Raman Spectroscopic Assessment of IBD Severity

Assessing disease severity in IBD is crucial for the evaluation of patient care and response to treatment. Due to the relapsing behaviour of IBD, it is essential that an

optimal treatment strategy is chosen and routinely evaluated. Clinically, the assessment of disease severity is validated using symptom scores such as Mayo Endoscopic score (MES), Robarts Histological index (RHI) and Ulcerative colitis endoscopic index of severity score (UCEIS). However, these scoring systems are often subject to bias, interobserver variability and the heterogeneity in patient clinical presentation [29], [195], [257]. Conventional IBD treatments typically result in non-specific inhibition of inflammation, which leads to a reduction in clinical symptoms. Consequently, treatment endpoints have historically centred in assessing the severity of symptoms, despite their limited correlation with mucosal inflammation or the overall impact of the disease related morbidity and mortality [258]. To this end, it is essential that alternative methods for histopathological diagnosis of disease severity and activity are developed to shorten the time taken to make an informed diagnosis. Such methods, capable of *real-time* assessment of IBD severity will have inherent advantages, including for instance, monitoring the progression of disease to determine whether current treatments are effective and/or predicting the outcomes of the disease, *i.e.,* as an indicator of long-term outcomes. Further, since severe or poorly controlled IBD may lead to complications including fistulas and structures, early recognition of severe disease state has the ability to prompt interventions to mitigate such risks [259]. It would also lead to improving the quality of life of the patients. Given that the more severe form of IBD often leads to much more pronounced symptoms and thus, limiting the daily activities to a greater degree, tailored treatment plans based on disease severity would also improve the overall quality of life. Further advantages of direct disease assessment include education - where the patients make a more informed decision related to their care and lifestyle and healthcare costs with severe IBD

associated with more frequent hospitalisations, surgeries, and expensive medicines – where the accurate severity assessments would optimise the more cost-effective strategies [260], [261].

Moreover, Ding *et* al. studied the *in-vivo* analysis of mucosal lipids to determine histological disease activity in UC by way of an endoscope-coupled RS. The researchers systematically investigated lipids in inflamed colon tissue, which was correlated with histological assessment of inflammatory status at the same location and observed that inflamed colon tissue (histology grade 1-3) presented a substantial decrease (50-60%) in phosphatidylcholine content and total lipid count when compared to quiescent and control tissue (grade 0), indicating that active inflammation reduces the overall level of lipids in the colonic mucosa [196]. However, for the lipid count comparing control and grade 0 tissue, no significant differences were observed and similarly, between the 3 inflammatory grades (1-3) thus indicating, lipid content could serve as a potential spectral marker in the assessment of disease activity. The advantage of the presented technique over current qualitative methods *e.g.,* mass spectrometry, is in its suitability and ease of use for the *in-situ* analysis of the colonic mucosa during endoscopy, eliminating the need for invasive biopsies. A further study with a similar focus in the field of RS assessment of IBD disease severity by Kirchberger-Tolstik *et* al. correlates Raman spectra obtained from tissue with the four-level Mayo score used to indicate endoscopic disease severity. A one-dimensional deep convolution neural-network (1D-CNN) was applied to produce a predictive model of the Raman spectroscopic data [195], achieving a sensitivity and specificity of 78% and 93%, respectively. The main observable differences between spectra from normal

and inflamed tissue indicate the presence of increased levels of proteins, lipids, and DNA. The authors demonstrate the ability of the methodology to classify fresh biopsy samples as well as diseases beyond the IBD and describe a low-risk diagnostic procedure with a *real-time* to results workflow.

2.8.5 Mucosal Healing in IBD

Endoscopic mucosal healing (MH) is defined as “complete absence of all inflammatory and ulcerative lesions” upon endoscopy [258]. In recent years, MH has been proposed as a superior measure compared to clinical symptoms in predicting the effectiveness of treatments as well as disease trajectory. The primary objective of treatment for IBD patients has traditionally been to induce and maintain symptomatic improvement, or ideally, to achieve remission, leading to improved quality of life for patients. However, due to varying interpretations of MH, challenges associated with reproducibility of IBD scores among different observes and potential microscopic inflammation in mucosa which appears healthy during endoscopy, there is an urgent need for enhanced techniques to aid in defining treatment endpoints for IBD [262].

Raman spectroscopic analysis has been in recent years proposed as a suitable methodology in differentiating mucosal healing from inflamed tissue. The first study published on this topic by Addis *et* al., investigated colon biopsy samples from quiescent (mucosally healed) and inflamed tissue in patients with UC. The suitability of RS in discriminating tissue, which has been described as histologically healed and tissue for which histological inflammation is still present, was assessed [188]. The authors identified clear differences between the two tissue types with significant

changes observed between the carotenoid peaks and two different phospholipid peaks [188]. These changes were attributed to the role of carotenoids as a defence mechanism against inflammation and that during inflammation, tissue loses integrity due to ulceration with such disruption leading to an expected decrease in phospholipids, further consistent with damage to the cell membrane. This study was limited to the single scan measurements of tissue biopsies and endoscopic scoring using the Mayo endoscopic score as opposed to the recently developed Paddington international virtual ChromoendoScopy ScOre (PICaSSO), which correlates strongly with the five most commonly applied histological indices in predicting histological remission [29].

Smith *et al*. built upon this and further identified spectral changes before and after treatment as well as the ability of RS to differentiate between MH and inflammation combined with rapid machine learning classification. The authors identified significant decrease in peak intensity at Raman bands of 1003 and 1252 cm^{-1} post treatment and also when MH was observed [191]. Of further interest was the increase in intensity at 1304 cm^{-1} detected in MH. [213]. The employed SKiNET algorithm achieved sensitivity and specificity of 96.3% and 95% in UC and 96.2% and 88% in CD patients, respectively. The detected spectral intensity increases were attributed to the inflammation present in both UC and CD tissue samples with phenylalanine associated with immune activation and an influx of inflammatory cells, similar to the previous studies, attributed to phospholipids with disruption in the bowel wall of inflamed tissue. Since this study was conducted *ex-vivo,* optimisation of laser power and exposure

times would be a primary requirement to enable translation into *in-vivo* clinical translations.

2.8.6 Biomarkers in IBD via RS

Diagnosis, evaluation of severity and prognosis still pose challenges for physicians in the clinic. The identification of potential biomarkers in IBD would therefore, offer important advantages including for instance, a more objective assessment thus, reducing the reliance on subjective evaluation of symptoms as well as early diagnosis of disease therefore, allowing timely interventions and improved outcomes. These would also offer treatment guidance in tailoring therapeutic strategies, monitoring of disease progression and a reduction in invasive procedures. Their use can therefore lead to an overall more effective and efficient healthcare delivery with improved outcomes for patients with IBD [263], [264].

Currently, blood and stool-based biomarkers offer reproducible and quantifiable resources which aid clinicians in the diagnosis and management of IBD with C-reactive (CRP), faecal calprotectin (FC) and lactoferrin being the only biomarkers routinely used in clinical practice [265], [266], [267]. CRP is an objective marker of inflammation in the assessment of disease activity however, it correlates less well in patients with UC compared to CD. It also has a lower sensitivity for mild or localised disease and as such, is mostly limited to more moderate or severe inflammation. In addition, whilst faecal markers such as FC and lactoferrin appear promising and may be more specific in the detection of gut inflammation, these markers require stool collection, often perceived as unpleasant for the patient. Recent studies also indicate the superior

performance of FC in the detection of UC than for CD. Further possible blood biomarkers of IBD have been described in literature with conflicting results. Perhaps the most well-known biomarkers are ASCA and pANCAs however, their sensitivity is currently poor at only 55% [268] thus, defeating their purpose.

An optimal IBD biomarker detection technique should be simple, straightforward, minimally invasive, cost-effective, and rapid. It must also be reproducible between patients and laboratories and should possess predictive significance concerning the likelihood of disease relapse of recurrence. In pursuit of this objective, RS could aid in IBD biomarker identification *via* the enhanced understanding of the molecular basis of IBDs and combined with machine learning, would uncover complex relationships and patterns within large datasets, leading to the discovery of new candidate biomarkers. Previous studies using techniques such as enzyme linked immunosorbent assay have indicated differing levels of mucosal and systemic concentrations of pro- and anti-inflammatory cytokines in patients with IBD, and here the ability of RS as a non-invasive technique for the multiplexed profiling of saliva and tissue biopsies has successfully established unique molecular barcodes for candidate biomarkers of IBD.

Overall, laboratory biomarkers are valuable and should be integrated into the comprehensive management of IBD. Nonetheless, it's essential to recognise they are not a panacea and until additional data becomes accessible, the use of CRP and other laboratory markers should be regarded as a supplementary aid to clinical observation and physical examination rather than a substitute. Building upon demonstrating relevant RS applications for the clinical and biomedical fields, including biomarker

discovery and patient stratification with a potential to contribute significantly to the development of new diagnostic methodologies and therapeutic monitoring, further studies validating these findings will lay the platform for defining functional relevance in the complex aetiology of IBD as well as cementing Raman spectroscopy as a powerful technique for discovery of biomarkers in other diseases and biological samples with many ramifications.

2.9 Point of Care Devices in Healthcare

Point of care testing (PoCT) devices have emerged as invaluable tools in the diagnosis and management of many healthcare concerns, offering rapid, on-the-spot results that enable timely decision-making and patient care. For the diagnosis and monitoring of IBD several PoCT devices are currently available, each with unique features and applications, significantly impacting healthcare delivery for individuals. One such example is the faecal calprotectin at-home test. Faecal calprotectin is a well-established biomarker for intestinal inflammation, making it a key target for PoCT devices [269]. It enables the quantitative measurement of calprotectin levels in stool samples, providing a non-invasive and cost-effective means of assessing disease activity. Faecal calprotectin levels demonstrate a stronger correlation with endoscopic disease activity compared to symptom-based clinical disease activity indices [270], [271]. Although the optimal cut-ff value for faecal calprotectin levels remains elusive, levels <250 µg/g indicate endoscopic remission in both UC and CD patients [272]. The at home- assay involves patients using a provided extraction tool alongside a manufacturer specific test cassette to collect a stool sample and extract calprotectin [273]. Patients may then use a smartphone equipped with a dedicated application to

capture images of the test strip, enabling them to conduct semi-quantitative faecal calprotectin measurements at home. In addition, endoscopic imaging devices, such as capsule endoscopes have revolutionised the diagnosis and monitoring of IBD at the point of care. These devices enable real-time visualisation of the GI tract, thus allowing clinicians to directly assess mucosal inflammation, disease extent, and complications without the need for traditional endoscopic procedures [274], [275].

CRP is another commonly used biomarker for assessing systemic inflammation in IBD as well as CVD. PoCT devices for CRP measurement offer rapid results from a small blood sample, aiding in the diagnosis, monitoring, and risk stratification of patients with IBD or CVDs [276]. CRP levels >10 mg/mL are associated with disease progression to surgery. Dried blood spot CRP is an accessible remote monitoring tool whereby patients simply draw a small blood sample from their fingertip and place it on the test card [271]. Moreover, several qualitative PoCT devices capable of measuring several biomarkers of CVD including Decision point (Nexus Dx, USA) which requires 6 drops of blood placed on a cassette in order to directly measure cardiac troponin (Tn), creatine kinase (CK), and myoglobin (Myo), MI CK-MB (LifeSign, USA) which requires 120 µL of either white blood cells, platelets or serum, Instant view (Alfa Scientific, USA) which requires 4 drops of serum and Cardiac STATus (Spectral Diagnostics, Canada) which uses 6 drops of serum are currently available technologies used to measure cardiac markers of disease [277], [278], [279], [280].

2.10 Conclusions

Here we have shown that Raman spectroscopy and saliva present a promising combination for translation to real-world use. Having surveyed Raman-saliva applications in both cardiovascular disease, IBD, alongside further applications in healthcare, current challenges with the non-specificity of biomarkers has been highlighted as a particular challenge highlighting the need for interdisciplinary collaborative research to further advance this field. This echoes the views of Poste *et* al. who has pointed out that, 'Biomarker discovery should be a component of large research networks, involving industry and experts in molecular biology, genetics, analytical chemistry, computation, engineering, clinical-trial design, epidemiology, statistics, regulation and health-care economics' [235]. A dedicated commentary on the specific field of EV related studies is included with a focussed discussion on the potential of extracellular vesicles which can be directly interrogated in saliva and blood plasma samples *via* Raman spectroscopy, for the specific detection of disease.

However, the breadth of emerging studies, collectively, emphasise the promise of RS in providing crucial understanding along with timely detection of the molecular and structural changes occurring in, in particular, IBD-affected tissues such as the colon. Raman spectroscopy, by providing unique biomolecular spectral fingerprints of target analytes with a very rapid analytical response, enabling non-destructive, label-free, quantitative analysis of composition and structure with an inherently straightforward detection and no complex sample preparation, renders itself as a powerful *in-situ* technique, yielding quantitative outputs. The non-invasive nature combined with the

ability to offer *real-time* rapid analysis has been cementing Raman spectroscopy recognition as a valuable adjunct to conventional diagnostic methods.

It is worth noting that albeit significant progress has been accomplished in these fields, challenges remain. One potential issue for *real-world* adoption is in a lack of standardisation in methods used within the current literature, from matters pertaining to biofluid collection, storage, and pre-processing, to those concerning Raman measurement protocols and chemometric analyses and classification. Standardisation of protocols and conducting larger-scale clinical studies are vital for advancing the clinical translation of RS in healthcare diagnostic and prognostic applications. Whilst RS has shown promise in research settings, there is limited clinical evidence demonstrating its diagnostic accuracy and clinical utility with many presented studies comprising small sample sizes. Thus, extensive clinical evaluation with much larger sample cohorts is an essential requirement in the clinical translation of RS. A further challenge associated with RS transition from benchtop to clinical laboratory and the bedside, is the complexity and cost of instrumentation. RS systems currently, require specialised personnel to operate, are often expensive and cumbersome to maintain. However, with the development and advancement of portable Raman instrumentation as well as hybrid PoCT technologies many of the existing limitations such as user-friendly interface, flexibility and point-of-care testing are concurrently being addressed, rendering it increasingly valuable tool in healthcare settings. RS, further combined with advanced machine learning techniques, enable the enhancement of the accuracy, reliability, and speed of disease diagnosis and the imminent automated classification of Raman spectra along with assignment to a particular IBD or CVD biomarkers, tissue

type or disease state is perhaps the most important step for the translation of Raman based diagnostics to *real-world*, clinical applications. Moreover, hybrid RS-AI methodologies have the ability to support large-scale screening for IBD and CVD risk factors or early signs of the disease, particularly useful for identifying high-risk populations and implementing preventative measures given both the increased incidence and the prevalence of, for example, IBDs.

Overall, the expanding interdisciplinary collaborations between clinicians, scientists, and engineers, which are pivotal to unlocking the full potential of RS for healthcare, are starting to bridge the gap between clinical research and practice, with the emerging findings being translated into *real-life* benefits for patients including, timely diagnosis, improved disease management and continued bedside monitoring. With the ability of RS to distinguish different diseased states, determine therapeutic response to treatment as well as characterise the complex microenvironment of the gut and beyond, this spectroscopic technique holds a further great potential to transform the early-stage IBD and CVD diagnosis and the long-term monitoring.

2.11 References

[1] J. R. Turner, “Intestinal mucosal barrier function in health and disease,” *Nat Rev Immunol*, vol. 9, no. 11, pp. 799–809, Nov. 2009, doi: 10.1038/nri2653.

[2] J. K. Gustafsson and M. E. V. Johansson, “The role of goblet cells and mucus in intestinal homeostasis,” *Nat Rev Gastroenterol Hepatol*, vol. 19, no. 12, pp. 785–803, Dec. 2022, doi: 10.1038/s41575-022-00675-x.

[3] A. N. Ananthakrishnan *et al.*, “Environmental triggers in IBD: a review of progress and evidence,” *Nat Rev Gastroenterol Hepatol*, vol. 15, no. 1, pp. 39–49, Jan. 2018, doi: 10.1038/nrgastro.2017.136.

[4] A. Levine *et al.*, “Pediatric modification of the Montreal classification for inflammatory bowel disease,” *Inflamm Bowel Dis*, vol. 17, no. 6, pp. 1314–1321, Jun. 2011, doi: 10.1002/ibd.21493.

[5] A. N. Ananthakrishnan, “Epidemiology and risk factors for IBD,” *Nat Rev Gastroenterol Hepatol*, vol. 12, no. 4, pp. 205–217, Apr. 2015, doi: 10.1038/nrgastro.2015.34.

[6] N. A. Molodecky *et al.*, “Increasing Incidence and Prevalence of the Inflammatory Bowel Diseases With Time, Based on Systematic Review,” *Gastroenterology*, vol. 142, no. 1, pp. 46-54.e42, Jan. 2012, doi: 10.1053/j.gastro.2011.10.001.

[7] E. Louis, C. Van Kemseke, and C. Reenaers, “Necessity of phenotypic classification of Inflammatory Bowel Disease,” *Best Pract Res Clin Gastroenterol*, vol. 25, pp. S2–S7, Apr. 2011, doi: 10.1016/S1521-6918(11)70003-8.

[8] K. Bassler, J. Schulte-Schrepping, S. Warnat-Herresthal, A. C. Aschenbrenner, and J. L. Schultze, “The Myeloid Cell Compartment—Cell by Cell,” *Annu Rev Immunol*, vol. 37, no. 1, pp. 269–293, Apr. 2019, doi: 10.1146/annurev-immunol-042718-041728.

[9] Herrero-Fernandez, Gomez-Bris, Somovilla-Crespo, and Gonzalez-Granado, “Immunobiology of Atherosclerosis: A Complex Net of Interactions,” *Int J Mol Sci*, vol. 20, no. 21, p. 5293, Oct. 2019, doi: 10.3390/ijms20215293.

[10] S. Ghilas, R. O’Keefe, L. A. Mielke, D. Raghu, M. Buchert, and M. Ernst, “Crosstalk between epithelium, myeloid and innate lymphoid cells during gut homeostasis and disease,” *Front Immunol*, vol. 13, Sep. 2022, doi: 10.3389/fimmu.2022.944982.

[11] J. Satsangi, “The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications,” *Gut*, vol. 55, no. 6, pp. 749–753, Jun. 2006, doi: 10.1136/gut.2005.082909.

[12] E. A. Mann and S. A. Saeed, “Gastrointestinal infection as a trigger for inflammatory bowel disease,” *Curr Opin Gastroenterol*, vol. 28, no. 1, pp. 24–29, Jan. 2012, doi: 10.1097/MOG.0b013e32834c453e.

[13] M. F. Neurath and M. Leppkes, “Resolution of ulcerative colitis,” *Semin Immunopathol*, vol. 41, no. 6, pp. 747–756, Nov. 2019, doi: 10.1007/s00281-019-00751-6.

[14] B. O. Schroeder *et al.*, “Bifidobacteria or Fiber Protects against Diet-Induced Microbiota-Mediated Colonic Mucus Deterioration,” *Cell Host Microbe*, vol. 23, no. 1, pp. 27-40.e7, Jan. 2018, doi: 10.1016/j.chom.2017.11.004.

[15] V. Yadav, F. Varum, R. Bravo, E. Furrer, D. Bojic, and A. W. Basit, “Inflammatory bowel disease: exploring gut pathophysiology for novel therapeutic targets,” *Translational Research*, vol. 176, pp. 38–68, Oct. 2016, doi: 10.1016/j.trsl.2016.04.009.

[16] A. Saez, B. Herrero-Fernandez, R. Gomez-Bris, H. Sánchez-Martinez, and J. M. Gonzalez-Granado, “Pathophysiology of Inflammatory Bowel Disease: Innate Immune System,” *Int J Mol Sci*, vol. 24, no. 2, p. 1526, Jan. 2023, doi: 10.3390/ijms24021526.

[17] K. Nijakowski *et al.*, “Potential Salivary Markers for Differential Diagnosis of Crohn’s Disease and Ulcerative Colitis,” *Life*, vol. 11, no. 9, p. 943, Sep. 2021, doi: 10.3390/life11090943.

[18] M. Cappello and G. C. Morreale, “The Role of Laboratory Tests in Crohn’s Disease,” *Clin Med Insights Gastroenterol*, vol. 9, p. CGast.S38203, Jan. 2016, doi: 10.4137/CGast.S38203.

[19] M. Scharl *et al.*, “Protection of Epithelial Barrier Function by the Crohn’s Disease Associated Gene Protein Tyrosine Phosphatase N2,” *Gastroenterology*, vol. 137, no. 6, pp. 2030-2040.e5, Dec. 2009, doi: 10.1053/j.gastro.2009.07.078.

[20] M. Daperno *et al.*, “Development and validation of a new, simplified endoscopic activity score for Crohn’s disease: the SES-CD,” *Gastrointest Endosc*, vol. 60, no. 4, pp. 505–512, Oct. 2004, doi: 10.1016/S0016-5107(04)01878-4.

[21] P. Parronchi *et al.*, “Type 1 T-helper cell predominance and interleukin-12 expression in the gut of patients with Crohn’s disease.,” *Am J Pathol*, vol. 150, no. 3, pp. 823–32, Mar. 1997, doi: 9060820.

[22] H. Schmitt, M. F. Neurath, and R. Atreya, “Role of the IL23/IL17 Pathway in Crohn’s Disease,” *Front Immunol*, vol. 12, Mar. 2021, doi: 10.3389/fimmu.2021.622934.

[23] J. Jansson *et al.*, “Metabolomics Reveals Metabolic Biomarkers of Crohn’s Disease,” *PLoS One*, vol. 4, no. 7, p. e6386, Jul. 2009, doi: 10.1371/journal.pone.0006386.

[24] K. Nijakowski *et al.*, “Potential salivary markers for differential diagnosis of crohn’s disease and ulcerative colitis,” *Life*, vol. 11, no. 9, pp. 1–10, 2021, doi: 10.3390/life11090943.

[25] V. Marcil *et al.*, “Association Between the PTPN2 Gene and Crohn’s Disease,” *Inflamm Bowel Dis*, vol. 19, no. 6, pp. 1149–1155, May 2013, doi: 10.1097/MIB.0b013e318280b181.

[26] S. P. L. Travis *et al.*, “Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS),” *Gut*, vol. 61, no. 4, pp. 535–542, Apr. 2012, doi: 10.1136/gutjnl-2011-300486.

[27] F. Casellas *et al.*, “Fecal excretion of human deoxyribonucleic acid as an index of inflammatory activity in ulcerative colitis,” *Clinical Gastroenterology and Hepatology*, vol. 2, no. 8, pp. 683–689, Aug. 2004, doi: 10.1016/S1542-3565(04)00291-5.

[28] W. Xu *et al.*, “The Mayo Endoscopic Score Is a Novel Predictive Indicator for Malignant Transformation in Ulcerative Colitis: A Long-Term Follow-Up Multicenter Study,” *Front Surg*, vol. 9, Mar. 2022, doi: 10.3389/fsurg.2022.832219.

[29] P. J. Trivedi *et al.*, “The Paddington International Virtual Chromoendoscopy Score in ulcerative colitis exhibits very good inter-rater agreement after computerized module training: a multicenter study across academic and community practice (with video),” *Gastrointest Endosc*, vol. 88, no. 1, pp. 95-106.e2, Jul. 2018, doi: 10.1016/j.gie.2018.02.044.

[30] E. A. Scoville *et al.*, “Alterations in lipid, amino acid, and energy metabolism distinguish Crohn’s disease from ulcerative colitis and control subjects by serum metabolomic profiling,” *Metabolomics*, vol. 14, no. 1, p. 17, Jan. 2018, doi: 10.1007/s11306-017-1311-y.

[31] R. M. Feakins, “Ulcerative colitis or Crohn’s disease? Pitfalls and problems,” *Histopathology*, vol. 64, no. 3, pp. 317–335, Feb. 2014, doi: 10.1111/his.12263.

[32] T. Okada *et al.*, “IL-8 and LYPD8 expression levels are associated with the inflammatory response in the colon of patients with ulcerative colitis,” *Biomed Rep*, Feb. 2020, doi: 10.3892/br.2020.1280.

[33] F. Heller *et al.*, “Interleukin-13 Is the Key Effector Th2 Cytokine in Ulcerative Colitis That Affects Epithelial Tight Junctions, Apoptosis, and Cell Restitution,” *Gastroenterology*, vol. 129, no. 2, pp. 550–564, Aug. 2005, doi: 10.1053/j.gastro.2005.05.002.

[34] S. Saeid Seyedian, F. Nokhostin, and M. Dargahi Malamir, “A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease,” *J Med Life*, vol. 12, no. 2, pp. 113–122, Apr. 2019, doi: 10.25122/jml-2018-0075.

[35] N. F. Rasmussen *et al.*, “Clinical procedures used to diagnose inflammatory bowel disease: real-world evidence from a Danish nationwide population-based study,” *BMJ Open Gastroenterol*, vol. 9, no. 1, p. e000958, Aug. 2022, doi: 10.1136/bmjgast-2022-000958.

[36] C. Morasso *et al.*, “Raman Analysis Reveals Biochemical Differences in Plasma of Crohn’s Disease Patients,” *J Crohns Colitis*, vol. 14, no. 11, pp. 1572–1580, Nov. 2020, doi: 10.1093/ecco-jcc/jjaa080.

[37] G. L. Austin, H. H. Herfarth, and R. S. Sandler, “A Critical Evaluation of Serologic Markers for Inflammatory Bowel Disease,” *Clinical Gastroenterology and Hepatology*, vol. 5, no. 5, pp. 545–547, May 2007, doi: 10.1016/j.cgh.2007.03.006.

[38] G. L. Austin, N. J. Shaheen, and R. S. Sandler, “Positive and Negative Predictive Values: Use of Inflammatory Bowel Disease Serologic Markers,” *Am J Gastroenterol*, vol. 101, no. 3, pp. 413–416, Mar. 2006, doi: 10.1111/j.1572-0241.2006.00475.x.

[39] I. J. Pence *et al.*, “Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy,” *Biomed Opt Express*, vol. 8, no. 2, p. 524, 2017, doi: 10.1364/boe.8.000524.

[40] C. Bielecki *et al.*, “Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells,” *J Biomed Opt*, vol. 17, no. 7, p. 0760301, 2012, doi: 10.1117/1.jbo.17.7.076030.

[41] C. Morasso *et al.*, “Raman analysis reveals biochemical differences in plasma of crohn’s disease patients,” *J Crohns Colitis*, vol. 14, no. 11, pp. 1572–1580, 2020, doi: 10.1093/ecco-jcc/jjaa080.

[42] K. Ember *et al.*, “Saliva-based detection of COVID-19 infection in a real-world setting using reagent-free Raman spectroscopy and machine learning,” *J Biomed Opt*, vol. 27, no. 02, 2022, doi: 10.1117/1.jbo.27.2.025002.

[43] N. M. Ralbovsky, L. Halámková, K. Wall, C. Anderson-Hanley, and I. K. Lednev, “Screening for Alzheimer’s Disease Using Saliva: A New Approach Based on Machine Learning and Raman Hyperspectroscopy,” *Journal of Alzheimer’s Disease*, vol. 71, no. 4, pp. 1351–1359, 2019, doi: 10.3233/JAD-190675.

[44] Dr. G. Calado *et al.*, “Raman Spectral Study of Saliva: A New Tool for Detection of Malignant and Premalignant Oral Lesions,” *Oral Surg Oral Med Oral Pathol Oral Radiol*, vol. 128, no. 1, pp. e90–e90, 2019, doi: 10.1016/j.oooo.2019.02.230.

[45] I. M. Colceriu-Şimon *et al.*, “The effects of low-dose irradiation on human saliva: A surface-enhanced Raman spectroscopy study,” *Diagnostics*, vol. 9, no. 3, p. 9030101, 2019, doi: 10.3390/diagnostics9030101.

[46] I. Miletich, “Introduction to Salivary Glands: Structure, Function and Embryonic Development,” 2010, pp. 1–20. doi: 10.1159/000313703.

[47] S. P. Humphrey and R. T. Williamson, “A review of saliva: Normal composition, flow, and function,” *J Prosthet Dent*, vol. 85, no. 2, pp. 162–169, 2001, doi: 10.1067/mpr.2001.113778.

[48] T. Pfaffe, J. Cooper-White, P. Beyerlein, K. Kostner, and C. Punyadeera, “Diagnostic Potential of Saliva: Current State and Future Applications,” *Clin. Chem.*, vol. 57, no. 5, pp. 675–687, 2011, doi: 10.1373/clinchem.2010.153767.

[49] W. M. Edgar, “Saliva: its secretion, composition and functions,” *Br Dent J*, vol. 172, no. 8, pp. 305–312, Apr. 1992, doi: 10.1038/sj.bdj.4807861.

[50] P. D. V. de Almeida, A. M. T. Grégio, M. Â. N. Machado, A. A. S. De Lima, and L. R. Azevedo, “Saliva composition and functions: A comprehensive review,” *Journal of Contemporary Dental Practice*, vol. 9, no. 3, pp. 72–80, 2008, doi: 10.5005/jcdp-9-3-72.

[51] W. M. Edgar, “Saliva and dental health. Clinical implications of saliva: report of a consensus meeting,” *Br Dent J*, vol. 169, no. 4, pp. 96–98, Aug. 1990, doi: 10.1038/sj.bdj.4807284.

[52] R. D. Pullan *et al.*, “Thickness of adherent mucus gel on colonic mucosa in humans and its relevance to colitis.,” *Gut*, vol. 35, no. 3, pp. 353–359, Mar. 1994, doi: 10.1136/gut.35.3.353.

[53] M. W. J. Dodds, D. A. Johnson, and C.-K. Yeh, “Health benefits of saliva: a review,” *J Dent*, vol. 33, no. 3, pp. 223–233, Mar. 2005, doi: 10.1016/j.jdent.2004.10.009.

[54] M. DECHAUME, S. GOGUEL, and P. POULLAIN, “[Human saliva: physical properties; chemical composition; cytology; bacteriology; serological properties; role].,” *Revue Stomatol*, vol. 51, no. 8–9, pp. 521–52, 1950, doi: 15443942.

[55] G. Iorgulescu, “Saliva between normal and pathological. Important factors in determining systemic and oral health.,” *J Med Life*, vol. 2, no. 3, pp. 303–7, 2009, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20112475

[56] M. Tiwari, “Science behind human saliva,” *J Nat Sci Biol Med*, vol. 2, no. 1, p. 53, 2011, doi: 10.4103/0976-9668.82322.

[57] A. van Nieuw Amerongen, J. G. M. Bolscher, and E. C. I. Veerman, “Salivary Proteins: Protective and Diagnostic Value in Cariology?,” *Caries Res*, vol. 38, no. 3, pp. 247–253, 2004, doi: 10.1159/000077762.

[58] N. Jacobsen and A. Hensten-Pettersen, “Salivary Amylase,” *Caries Res*, vol. 4, no. 3, pp. 193–199, 1970, doi: 10.1159/000259641.

[59] J. L. Chicharro, A. Lucía, M. Pérez, A. F. Vaquero, and R. Ureña, “Saliva composition and exercise,” *Sports Medicine*, vol. 26, no. 1, pp. 17–27, 1998, doi: 10.2165/00007256-199826010-00002.

[60] A. R. M. Radzol, K. Y. Lee, W. Mansor, and N. Ariffin, “Biostatistical Analysis of Principle Component of Salivary Raman Spectra for NS1 Infection,” in *2016 IEEE EMBS CONFERENCE ON BIOMEDICAL ENGINEERING AND SCIENCES (IECBES)*, in IEEE EMBS Conference on Biomedical Engineering and Sciences. 2016, pp. 13–18.

[61] C. Dawes, “The effects of flow rate and duration of stimulation on the concentrations of protein and the main electrolytes in human parotid saliva,” *Arch Oral Biol*, vol. 14, no. 3, pp. 277–294, Mar. 1969, doi: 10.1016/0003-9969(69)90231-3.

[62] K. Li *et al.*, “A signature of saliva-derived exosomal small RNAs as predicting biomarker for esophageal carcinoma: a multicenter prospective study,” *Mol Cancer*, vol. 21, no. 1, p. 21, Dec. 2022, doi: 10.1186/s12943-022-01499-8.

[63] D. P. Lima, D. G. Diniz, S. A. S. Moimaz, D. H. Sumida, and A. C. Okamoto, “Saliva: reflection of the body,” *International Journal of Infectious Diseases*, vol. 14, no. 3, pp. e184–e188, Mar. 2010, doi: 10.1016/j.ijid.2009.04.022.

[64] C. K. Muro, L. de Souza Fernandes, and I. K. Lednev, “Sex Determination Based on Raman Spectroscopy of Saliva Traces for Forensic Purposes,” *Anal Chem*, vol. 88, no. 24, pp. 12489–12593, 2016, doi: 10.1021/acs.analchem.6b03988.

[65] S. P. Humphrey and R. T. Williamson, “A review of saliva: Normal composition, flow, and function,” *J Prosthet Dent*, vol. 85, no. 2, pp. 162–169, Feb. 2001, doi: 10.1067/mpr.2001.113778.

[66] P. N. De Aza, Z. B. Luklinska, M. R. Anseau, M. Hector, F. Guitián, and S. De Aza, “Reactivity of a wollastonite-tricalcium phosphate Bioeutectic ceramic in human parotid saliva.,” *Biomaterials*, vol. 21, no. 17, pp. 1735–1741, Sep. 2000, doi: 10.1016/s0142-9612(00)00058-2.

[67] S. Watanabe, M. Ohnishi, K. Imai, E. Kawano, and S. Igarashi, “Estimation of the total saliva volume produced per day in five-year-old children,” *Arch Oral Biol*, vol. 40, no. 8, pp. 781–782, 1995, doi: 10.1016/0003-9969(95)00026-L.

[68] C. Dawes, “Circadian rhythms in human salivary flow rate and composition,” *J Physiol*, vol. 220, no. 3, pp. 529–545, 1972, doi: 10.1113/jphysiol.1972.sp009721.

[69] V. D’Elia, G. Montalvo García, and C. García Ruiz, “Spectroscopic Trends for the Determination of Illicit Drugs in Oral Fluid,” *Appl Spectrosc Rev*, vol. 50, no. 9, pp. 775–796, Oct. 2015, doi: 10.1080/05704928.2015.1075206.

[70] W. Zhang, S. Guo, W. S. Pereira Carvalho, Y. Jiang, and M. J. Serpe, “Portable point-of-care diagnostic devices,” *Analytical Methods*, vol. 8, no. 44, pp. 7847–7867, 2016, doi: 10.1039/C6AY02158A.

[71] J. M. Yoshizawa, C. A. Schafer, J. J. Schafer, J. J. Farrell, B. J. Paster, and D. T. W. Wong, “Salivary Biomarkers: Toward Future Clinical and Diagnostic Utilities,” *Clin Microbiol Rev*, vol. 26, no. 4, pp. 781–791, Oct. 2013, doi: 10.1128/CMR.00021-13.

[72] J. K. M. Aps and L. C. Martens, “Review: The physiology of saliva and transfer of drugs into saliva,” *Forensic Sci Int*, vol. 150, no. 2–3, pp. 119–131, Jun. 2005, doi: 10.1016/j.forsciint.2004.10.026.

[73] S. Bonassi, M. Neri, and R. Puntoni, “Validation of biomarkers as early predictors of disease,” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 480–481, pp. 349–358, Sep. 2001, doi: 10.1016/S0027-5107(01)00194-4.

[74] A. Roi, L. C. Rusu, C. I. Roi, R. E. Luca, S. Boia, and R. I. Munteanu, “A New Approach for the Diagnosis of Systemic and Oral Diseases Based on Salivary Biomolecules,” *Dis Markers*, vol. 2019, pp. 1–11, Feb. 2019, doi: 10.1155/2019/8761860.

[75] N. L. Anderson and N. G. Anderson, “The Human Plasma Proteome,” *Molecular & Cellular Proteomics*, vol. 1, no. 11, pp. 845–867, Nov. 2002, doi: 10.1074/mcp.R200007-MCP200.

[76] T. Pfaffe, J. Cooper-White, P. Beyerlein, K. Kostner, and C. Punyadeera, “Diagnostic Potential of Saliva: Current State and Future Applications,” *Clin*

Chem, vol. 57, no. 5, pp. 675–687, May 2011, doi: 10.1373/clinchem.2010.153767.

[77] R. S. Khan, H. U. Rehman, and I. U. Rehman, “Saliva for the diagnosis of COVID-19,” *Appl Spectrosc Rev*, vol. 55, no. 9–10, pp. 805–809, Nov. 2020, doi: 10.1080/05704928.2020.1809442.

[78] A. Dadas, J. Washington, R. Diaz-Arrastia, and D. Janigro, “Biomarkers in traumatic brain injury (TBI): a review,” *Neuropsychiatr Dis Treat*, vol. Volume 14, pp. 2989–3000, Nov. 2018, doi: 10.2147/NDT.S125620.

[79] X. Li, T. Yang, and J. Ding, “[Surface enhanced Raman spectroscopy (SERS) of saliva for the diagnosis of lung cancer].,” *Guang Pu Xue Yu Guang Pu Fen Xi*, vol. 32, no. 2, pp. 391–393, Feb. 2012.

[80] A. C. Hernández-Arteaga *et al.*, “Determination of Salivary Sialic Acid Through Nanotechnology: A Useful Biomarker for the Screening of Breast Cancer.,” *Arch Med Res*, vol. 50, no. 3, pp. 105–110, Apr. 2019, doi: 10.1016/j.arcmed.2019.05.013.

[81] R. S. Vasan, “Biomarkers of Cardiovascular Disease,” *Circulation*, vol. 113, no. 19, pp. 2335–2362, May 2006, doi: 10.1161/CIRCULATIONAHA.104.482570.

[82] R. Farah, H. Haraty, Z. Salame, Y. Fares, D. M. Ojcius, and N. Said Sadier, “Salivary biomarkers for the diagnosis and monitoring of neurological diseases,” *Biomed J*, vol. 41, no. 2, pp. 63–87, 2018, doi: 10.1016/j.bj.2018.03.004.

[83] C.-Z. Zhang *et al.*, “Saliva in the diagnosis of diseases,” *Int J Oral Sci*, vol. 8, no. 3, pp. 133–137, Sep. 2016, doi: 10.1038/ijos.2016.38.

[84] J. Liu and Y. Duan, “Saliva: A potential media for disease diagnostics and monitoring,” *Oral Oncol.*, vol. 48, no. 7, pp. 569–577, 2012, doi: 10.1016/j.oraloncology.2012.01.021.

[85] A. Roi, L. C. Rusu, C. I. Roi, R. E. Luca, S. Boia, and R. I. Munteanu, “A New Approach for the Diagnosis of Systemic and Oral Diseases Based on Salivary Biomolecules,” *Dis Markers*, vol. 2019, p. 8761860, 2019, doi: https://doi.org/10.1155/2019/8761860.

[86] F. Bermejo-Pareja, D. Antequera, T. Vargas, J. A. Molina, and E. Carro, “Saliva levels of Abeta1-42 as potential biomarker of Alzheimer’s disease: a pilot study,” *BMC Neurol*, vol. 10, no. 108, pp. 1–7, 2010, doi: 10.1186/1471-2377-10-108.

[87] S. Prasad, A. K. Tyagi, and B. B. Aggarwal, “Detection of inflammatory biomarkers in saliva and urine: Potential in diagnosis, prevention, and treatment for chronic diseases,” *Exp. Biol. Med.*, vol. 241, pp. 783–799, 2016, doi: 10.1177/1535370216638770.

[88] N. H. Othman, K. Y. Lee, A. R. M. Radzol, W. Mansor, and N. I. A. Hisham, “PCA-QDA Model Selection for Detecting NS1 Related Diseases from SERS Spectra of Salivary Mixtures,” in *WORLD CONGRESS ON MEDICAL PHYSICS AND BIOMEDICAL ENGINEERING 2018, VOL 1*, Lhotska, L and Sukupova, L and Lackovic, I and Ibbott, GS, Ed., in IFMBE Proceedings, vol. 68. 2019, pp. 623–627. doi: 10.1007/978-981-10-9035-6_116.

[89] X. Zheng *et al.*, “Salivary exosomal PSMA7: a promising biomarker of inflammatory bowel disease,” *Protein Cell*, vol. 8, no. 9, pp. 686–695, Sep. 2017, doi: 10.1007/s13238-017-0413-7.

[90] S. Hosseini, P. Vázquez-Villegas, M. Rito-Palomares, and S. O. Martinez-Chapa, “Advantages, Disadvantages and Modifications of Conventional ELISA,” 2018, pp. 67–115. doi: 10.1007/978-981-10-6766-2_5.

[91] W. C. Rodrigues, P. Catbagan, S. Rana, G. Wang, and C. Moore, “Detection of synthetic cannabinoids in oral fluid using ELISA and LC-MS-MS,” *J Anal Toxicol*, vol. 37, no. 8, pp. 526–533, 2013, doi: 10.1093/jat/bkt067.

[92] S. Hosseini, P. Vázquez-Villegas, M. Rito-Palomares, and S. O. Martinez-Chapa, *Enzyme-linked Immunosorbent Assay (ELISA)*. in SpringerBriefs in Applied Sciences and Technology. Singapore: Springer Singapore, 2018. doi: 10.1007/978-981-10-6766-2.

[93] Q. SONG, T. LANGE, A. SPAHR, G. ADLER, and G. BODE, “Characteristic distribution pattern of Helicobacter pylori in dental plaque and saliva detected with nested PCR,” *J Med Microbiol*, vol. 49, no. 4, pp. 349–353, Apr. 2000, doi: 10.1099/0022-1317-49-4-349.

[94] S. A. Wickline and G. M. Lanza, “Molecular imaging, targeted therapeutics, and nanoscience,” *J Cell Biochem*, vol. 87, no. S39, pp. 90–97, 2002, doi: 10.1002/jcb.10422.

[95] X. Sun, E. Salih, F. G. Oppenheim, and E. J. Helmerhorst, “Activity-based mass spectrometric characterization of proteases and inhibitors in human saliva,” *Proteomics Clin Appl*, vol. 3, no. 7, pp. 810–820, Jul. 2009, doi: 10.1002/prca.200800242.

[96] S. Fu, Y. Zhang, Y. Li, L. Luo, Y. Zhao, and Y. Yao, “Extracellular vesicles in cardiovascular diseases.,” *Cell Death Discov*, vol. 6, p. 68, 2020, doi: 10.1038/s41420-020-00305-y.

[97] N. Comfort *et al.*, “Isolation and characterization of extracellular vesicles in saliva of children with asthma,” *Extracell Vesicles Circ Nucl Acids*, 2021, doi: 10.20517/evcna.2020.09.

[98] A. Gámez-Valero, S. I. Lozano-Ramos, I. Bancu, R. Lauzurica-Valdemoros, and F. E. Borràs, “Urinary extracellular vesicles as source of biomarkers in kidney diseases,” *Frontiers in Immunology*, vol. 6, no. JAN. Frontiers Media S.A., 2015. doi: 10.3389/fimmu.2015.00006.

[99] M. Palviainen *et al.*, “Extracellular vesicles from human plasma and serum are carriers of extravesicular cargo—Implications for biomarker discovery,” *PLoS One*, vol. 15, no. 8, p. e0236439, Aug. 2020, doi: 10.1371/journal.pone.0236439.

[100] M. Yáñez-Mó *et al.*, “Biological properties of extracellular vesicles and their physiological functions,” *Journal of Extracellular Vesicles*, vol. 4, no. 2015. Co-Action Publishing, pp. 1–60, 2015. doi: 10.3402/jev.v4.27066.

[101] F. T. Borges *et al.*, “TGF-β1-Containing exosomes from injured epithelial cells activate fibroblasts to initiate tissue regenerative responses and fibrosis,” *Journal of the American Society of Nephrology*, vol. 24, no. 3, pp. 385–392, 2013, doi: 10.1681/ASN.2012101031.

[102] D. D. Taylor and C. Gercel-Taylor, “MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer,” *Gynecol Oncol*, vol. 110, no. 1, pp. 13–21, 2008, doi: 10.1016/j.ygyno.2008.04.033.

[103] biomarkers Thompson and M. W. martin, “Extracellular vesicles in neurodegenerative disease-pathogenesis to,” 2016.

[104] F. A. Alzahrani *et al.*, “The Potential Use of Mesenchymal Stem Cells and Their Derived Exosomes as Immunomodulatory Agents for COVID-19 Patients,” *Stem Cells Int*, vol. 2020, pp. 1–11, 2020, doi: 10.1155/2020/8835986.

[105] S. Caruso and I. K. H. Poon, “Apoptotic cell-derived extracellular vesicles: More than just debris,” *Frontiers in Immunology*, vol. 9, no. JUN. Frontiers Media S.A., 2018. doi: 10.3389/fimmu.2018.01486.

[106] G. Van Niel, G. D’Angelo, and G. Raposo, “Shedding light on the cell biology of extracellular vesicles,” *Nature Reviews Molecular Cell Biology*, vol. 19, no. 4. Nature Publishing Group, pp. 213–228, 2018. doi: 10.1038/nrm.2017.125.

[107] G. Hu, K. M. Drescher, and X. M. Chen, “Exosomal miRNAs: Biological properties and therapeutic potential,” *Front Genet*, vol. 3, no. APR, 2012, doi: 10.3389/fgene.2012.00056.

[108] D. D. Gonda *et al.*, “Neuro-oncologic applications of exosomes, microvesicles, and other nano-sized extracellular particles,” *Neurosurgery*, vol. 72, no. 4, pp. 501–510, 2013, doi: 10.1227/NEU.0b013e3182846e63.

[109] H. Shao, H. Im, C. M. Castro, X. Breakefield, R. Weissleder, and H. Lee, “New Technologies for Analysis of Extracellular Vesicles,” *Chemical Reviews*, vol. 118, no. 4. American Chemical Society, pp. 1917–1950, 2018. doi: 10.1021/acs.chemrev.7b00534.

[110] K. W. Witwer *et al.*, “Standardization of sample collection, isolation and analysis methods in extracellular vesicle research,” *J Extracell Vesicles*, vol. 2, no. 1, 2013, doi: 10.3402/jev.v2i0.20360.

[111] A. V Vlassov, S. Magdaleno, R. Setterquist, and R. Conrad, “Exosomes: Current knowledge of their composition, biological functions, and diagnostic and therapeutic potentials,” *Biochimica et Biophysica Acta - General Subjects*, vol. 1820, no. 7. pp. 940–948, 2012. doi: 10.1016/j.bbagen.2012.03.017.

[112] G. Müller, “Novel tools for the study of cell type-specific exosomes and microvesicles,” *J Bioanal Biomed*, vol. 4, no. 4, pp. 46–60, 2012, doi: 10.4172/1948-593X.1000063.

[113] Z. Zhao, H. Wijerathne, A. K. Godwin, and S. A. Soper, “Isolation and analysis methods of extracellular vesicles (EVs),” *Extracell Vesicles Circ Nucl Acids*, 2021, doi: 10.20517/evcna.2021.07.

[114] S. Xue, X. Cai, W. Li, Z. Zhang, W. Dong, and G. Hui, “Elevated Plasma Endothelial Microparticles in Alzheimer’s Disease,” *Dement Geriatr Cogn Disord*, vol. 34, no. 3–4, pp. 174–180, 2012, doi: 10.1159/000343491.

[115] I. E. Hoefer *et al.*, “Novel methodologies for biomarker discovery in atherosclerosis,” *Eur Heart J*, vol. 36, no. 39, pp. 2635–2642, Oct. 2015, doi: 10.1093/eurheartj/ehv236.

[116] B. Levänen *et al.*, “Altered microRNA profiles in bronchoalveolar lavage fluid exosomes in asthmatic patients,” *Journal of Allergy and Clinical Immunology*, vol. 131, no. 3, pp. 894-903.e8, Mar. 2013, doi: 10.1016/j.jaci.2012.11.039.

[117] A. Haghikia *et al.*, “Regulated microRNAs in the CSF of patients with multiple sclerosis: A case-control study,” *Neurology*, vol. 79, no. 22, pp. 2166–2170, Nov. 2012, doi: 10.1212/WNL.0b013e3182759621.

[118] A. I. Masyuk, T. V. Masyuk, and N. F. LaRusso, “Exosomes in the pathogenesis, diagnostics and therapeutics of liver diseases,” *J Hepatol*, vol. 59, no. 3, pp. 621–625, Sep. 2013, doi: 10.1016/j.jhep.2013.03.028.

[119] L. Zu *et al.*, “Proteomic research of high-glucose-activated endothelial microparticles and related proteins to Alzheimer’s disease,” *Diab Vasc Dis Res*, vol. 12, no. 6, pp. 467–470, Nov. 2015, doi: 10.1177/1479164115597865.

[120] A. I. Masyuk, T. V. Masyuk, and N. F. LaRusso, “Exosomes in the pathogenesis, diagnostics and therapeutics of liver diseases,” *J Hepatol*, vol. 59, no. 3, pp. 621–625, Sep. 2013, doi: 10.1016/j.jhep.2013.03.028.

[121] Y. Couch *et al.*, “Inflammatory Stroke Extracellular Vesicles Induce Macrophage Activation,” *Stroke*, vol. 48, no. 8, pp. 2292–2296, Aug. 2017, doi: 10.1161/STROKEAHA.117.017236.

[122] K. B. Johnsen, J. M. Gudbergsson, T. L. Andresen, and J. B. Simonsen, “What is the blood concentration of extracellular vesicles? Implications for the use of extracellular vesicles as blood-borne biomarkers of cancer,” *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1871, no. 1, pp. 109–116, Jan. 2019, doi: 10.1016/j.bbcan.2018.11.006.

[123] A. W. Brenner, G. H. Su, and F. Momen-Heravi, “Isolation of Extracellular Vesicles for Cancer Diagnosis and Functional Studies,” 2019, pp. 229–237. doi: 10.1007/978-1-4939-8879-2_21.

[124] O. Ruhen and K. Meehan, “Tumor-Derived Extracellular Vesicles as a Novel Source of Protein Biomarkers for Cancer Diagnosis and Monitoring,” *Proteomics*, vol. 19, no. 1–2, p. 1800155, Jan. 2019, doi: 10.1002/pmic.201800155.

[125] Y. Yoshioka *et al.*, “Ultra-sensitive liquid biopsy of circulating extracellular vesicles using ExoScreen,” *Nat Commun*, vol. 5, p. 3591, 2014, doi: 10.1038/ncomms4591.

[126] Z. Zhao, Y. Yang, Y. Zeng, and M. He, “A microfluidic ExoSearch chip for multiplexed exosome detection towards blood-based ovarian cancer diagnosis,” *Lab-on-a-Chip*, vol. 16, pp. 489–496, 2016, doi: 10.1039/x0xx00000x.

[127] D. Maiolo *et al.*, “Colorimetric nanoplasmonic assay to determine purity and titrate extracellular vesicles,” *Anal Chem*, vol. 87, no. 8, pp. 4168–4176, 2015, doi: 10.1021/ac504861d.

[128] J. C. Akers, D. Gonda, R. Kim, B. S. Carter, and C. C. Chen, “Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies,” *J Neurooncol*, vol. 113, no. 1, pp. 1–11, May 2013, doi: 10.1007/s11060-013-1084-8.

[129] D. Duijvesz *et al.*, “Differential tissue expression of extracellular vesicle-derived proteins in prostate cancer,” *Prostate*, vol. 79, no. 9, pp. 1032–1042, Jun. 2019, doi: 10.1002/pros.23813.

[130] B. W. Sódar *et al.*, “Low-density lipoprotein mimics blood plasma-derived exosomes and microvesicles during isolation and detection,” *Sci Rep*, vol. 6, no. 1, p. 24316, Apr. 2016, doi: 10.1038/srep24316.

[131] Z. Varga *et al.*, “Size Measurement of Extracellular Vesicles and Synthetic Liposomes: The Impact of the Hydration Shell and the Protein Corona,” *Colloids Surf B Biointerfaces*, vol. 192, p. 111053, Aug. 2020, doi: 10.1016/j.colsurfb.2020.111053.

[132] T. S. Martins, M. Vaz, and A. G. Henriques, “A review on comparative studies addressing exosome isolation methods from body fluids,” *Anal Bioanal Chem*, vol. 415, no. 7, pp. 1239–1263, Mar. 2023, doi: 10.1007/s00216-022-04174-5.

[133] Y. Yuana *et al.*, “Handling and storage of human body fluids for analysis of extracellular vesicles,” *J Extracell Vesicles*, vol. 4, no. 1, Jan. 2015, doi: 10.3402/jev.v4.29260.

[134] M. Wu *et al.*, “Separating extracellular vesicles and lipoproteins via acoustofluidics.,” *Lab Chip*, vol. 19, no. 7, pp. 1174–1182, Mar. 2019, doi: 10.1039/c8lc01134f.
[135] M. A. Livshits *et al.*, “Correction: Corrigendum: Isolation of exosomes by differential centrifugation: Theoretical analysis of a commonly used protocol,” *Sci Rep*, vol. 6, no. 1, p. 21447, Mar. 2016, doi: 10.1038/srep21447.
[136] S. Guan, H. Yu, G. Yan, M. Gao, W. Sun, and X. Zhang, “Characterization of Urinary Exosomes Purified with Size Exclusion Chromatography and Ultracentrifugation,” *J Proteome Res*, vol. 19, no. 6, pp. 2217–2225, Jun. 2020, doi: 10.1021/acs.jproteome.9b00693.
[137] E. A. Mol, M.-J. Goumans, P. A. Doevendans, J. P. G. Sluijter, and P. Vader, “Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation,” *Nanomedicine*, vol. 13, no. 6, pp. 2061–2065, Aug. 2017, doi: 10.1016/j.nano.2017.03.011.
[138] K. Iwai, T. Minamisawa, K. Suga, Y. Yajima, and K. Shiba, “Isolation of human salivary extracellular vesicles by iodixanol density gradient ultracentrifugation and their characterizations,” *J Extracell Vesicles*, vol. 5, no. 1, p. 30829, Jan. 2016, doi: 10.3402/jev.v5.30829.
[139] Z. Onódi *et al.*, “Isolation of High-Purity Extracellular Vesicles by the Combination of Iodixanol Density Gradient Ultracentrifugation and Bind-Elute Chromatography From Blood Plasma,” *Front Physiol*, vol. 9, Oct. 2018, doi: 10.3389/fphys.2018.01479.
[140] D. W. Greening, R. Xu, H. Ji, B. J. Tauro, and R. J. Simpson, “A Protocol for Exosome Isolation and Characterization: Evaluation of Ultracentrifugation, Density-Gradient Separation, and Immunoaffinity Capture Methods,” 2015, pp. 179–209. doi: 10.1007/978-1-4939-2550-6_15.
[141] D. Wang and W. Sun, “Urinary extracellular microvesicles: Isolation methods and prospects for urinary proteome,” *Proteomics*, vol. 14, no. 16, pp. 1922–1932, Aug. 2014, doi: 10.1002/pmic.201300371.
[142] M. L. Heinemann *et al.*, “Benchtop isolation and characterization of functional exosomes by sequential filtration,” *J Chromatogr A*, vol. 1371, pp. 125–135, Dec. 2014, doi: 10.1016/j.chroma.2014.10.026.
[143] R. J. Lobb *et al.*, “Optimized exosome isolation protocol for cell culture supernatant and human plasma,” *J Extracell Vesicles*, vol. 4, no. 1, p. 27031, Jan. 2015, doi: 10.3402/jev.v4.27031.
[144] J. Q. Gerlach *et al.*, “Surface Glycosylation Profiles of Urine Extracellular Vesicles,” *PLoS One*, vol. 8, no. 9, p. e74801, Sep. 2013, doi: 10.1371/journal.pone.0074801.
[145] J. B. Lang *et al.*, “Comparative analysis of extracellular vesicle isolation methods from human AML bone marrow cells and AML cell lines,” *Front Oncol*, vol. 12, Oct. 2022, doi: 10.3389/fonc.2022.949261.
[146] M. L. Alvarez, M. Khosroheidari, R. Kanchi Ravi, and J. K. DiStefano, “Comparison of protein, microRNA, and mRNA yields using different methods of urinary exosome isolation for the discovery of kidney disease biomarkers,” *Kidney Int*, vol. 82, no. 9, pp. 1024–1032, Nov. 2012, doi: 10.1038/ki.2012.256.
[147] Z. Andreu *et al.*, “Comparative analysis of EV isolation procedures for miRNAs detection in serum samples,” *J Extracell Vesicles*, vol. 5, no. 1, p. 31655, Jan. 2016, doi: 10.3402/jev.v5.31655.

[148] J. Ko, E. Carpenter, and D. Issadore, “Detection and isolation of circulating exosomes and microvesicles for cancer monitoring and diagnostics using micro-/nano-based devices,” *Analyst*, vol. 141, no. 2, pp. 450–460, 2016, doi: 10.1039/C5AN01610J.

[149] B. J. Tauro *et al.*, “Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes,” *Methods*, vol. 56, no. 2, pp. 293–304, Feb. 2012, doi: 10.1016/j.ymeth.2012.01.002.

[150] B. J. Tauro *et al.*, “Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes,” *Methods*, vol. 56, no. 2, pp. 293–304, Feb. 2012, doi: 10.1016/j.ymeth.2012.01.002.

[151] A. N. Böing, E. van der Pol, A. E. Grootemaat, F. A. W. Coumans, A. Sturk, and R. Nieuwland, “Single-step isolation of extracellular vesicles by size-exclusion chromatography,” *J Extracell Vesicles*, vol. 3, no. 1, p. 23430, Jan. 2014, doi: 10.3402/jev.v3.23430.

[152] D. D. Taylor and S. Shah, “Methods of isolating extracellular vesicles impact down-stream analyses of their cargoes,” *Methods*, vol. 87, pp. 3–10, Oct. 2015, doi: 10.1016/j.ymeth.2015.02.019.

[153] P. Han, P. M. Bartold, C. Salomon, and S. Ivanovski, “Salivary Outer Membrane Vesicles and DNA Methylation of Small Extracellular Vesicles as Biomarkers for Periodontal Status: A Pilot Study,” *Int J Mol Sci*, vol. 22, no. 5, p. 2423, Feb. 2021, doi: 10.3390/ijms22052423.

[154] C. Gai *et al.*, “Salivary extracellular vesicle-associated miRNAs as potential biomarkers in oral squamous cell carcinoma,” *BMC Cancer*, vol. 18, no. 1, p. 439, Dec. 2018, doi: 10.1186/s12885-018-4364-z.

[155] P. Han, A. Lai, C. Salomon, and S. Ivanovski, “Detection of Salivary Small Extracellular Vesicles Associated Inflammatory Cytokines Gene Methylation in Gingivitis,” *Int J Mol Sci*, vol. 21, no. 15, p. 5273, Jul. 2020, doi: 10.3390/ijms21155273.

[156] L. A. Aqrawi *et al.*, “Identification of potential saliva and tear biomarkers in primary Sjögren’s syndrome, utilising the extraction of extracellular vesicles and proteomics analysis,” *Arthritis Res Ther*, vol. 19, no. 1, p. 14, Dec. 2017, doi: 10.1186/s13075-017-1228-x.

[157] Y. Cheng *et al.*, “Potential biomarkers to detect traumatic brain injury by the profiling of salivary extracellular vesicles,” *J Cell Physiol*, vol. 234, no. 8, pp. 14377–14388, Aug. 2019, doi: 10.1002/jcp.28139.

[158] Z. Cao *et al.*, “α-Synuclein in salivary extracellular vesicles as a potential biomarker of Parkinson’s disease,” *Neurosci Lett*, vol. 696, pp. 114–120, Mar. 2019, doi: 10.1016/j.neulet.2018.12.030.

[159] N. Comfort *et al.*, “Isolation and characterization of extracellular vesicles in saliva of children with asthma,” *Extracell Vesicles Circ Nucl Acids*, 2021, doi: 10.20517/evcna.2020.09.

[160] Y. E. Park *et al.*, “Microbial changes in stool, saliva, serum, and urine before and after anti-TNF-α therapy in patients with inflammatory bowel diseases,” *Sci Rep*, vol. 12, no. 1, p. 6359, Apr. 2022, doi: 10.1038/s41598-022-10450-2.

[161] F. Zhao *et al.*, “Extracellular vesicles package dsDNA to aggravate Crohn’s disease by activating the STING pathway,” *Cell Death Dis*, vol. 12, no. 9, p. 815, Aug. 2021, doi: 10.1038/s41419-021-04101-z.

[162] D. M. Goss, S. A. Vasilescu, G. Sacks, D. K. Gardner, and M. E. Warkiani, “Microfluidics facilitating the use of small extracellular vesicles in innovative approaches to male infertility,” *Nat Rev Urol*, vol. 20, no. 2, pp. 66–95, Feb. 2023, doi: 10.1038/s41585-022-00660-8.

[163] S. R. Vitale *et al.*, “Detection of tumor-derived extracellular vesicles in plasma from patients with solid cancer,” *BMC Cancer*, vol. 21, no. 1, p. 315, Dec. 2021, doi: 10.1186/s12885-021-08007-z.

[164] J. Chen *et al.*, “Rapid and efficient isolation and detection of extracellular vesicles from plasma for lung cancer diagnosis,” *Lab Chip*, vol. 19, no. 3, pp. 432–443, 2019, doi: 10.1039/C8LC01193A.

[165] D. Osti *et al.*, “Clinical Significance of Extracellular Vesicles in Plasma from Glioblastoma Patients,” *Clinical Cancer Research*, vol. 25, no. 1, pp. 266–276, Jan. 2019, doi: 10.1158/1078-0432.CCR-18-1941.

[166] K. Liang *et al.*, “Nanoplasmonic quantification of tumour-derived extracellular vesicles in plasma microsamples for diagnosis and treatment monitoring,” *Nat Biomed Eng*, vol. 1, no. 4, p. 0021, Feb. 2017, doi: 10.1038/s41551-016-0021.

[167] E. S. H. Cheow, W. C. Cheng, C. N. Lee, D. de Kleijn, V. Sorokin, and S. K. Sze, “Plasma-derived Extracellular Vesicles Contain Predictive Biomarkers and Potential Therapeutic Targets for Myocardial Ischemic (MI) Injury,” *Molecular & Cellular Proteomics*, vol. 15, no. 8, pp. 2628–2640, Aug. 2016, doi: 10.1074/mcp.M115.055731.

[168] S. Oggero *et al.*, “Role of plasma extracellular vesicles in prediction of cardiovascular risk and alterations in response to statin therapy in hypertensive patients,” *J Hypertens*, vol. 40, no. 8, pp. 1522–1529, Aug. 2022, doi: 10.1097/HJH.0000000000003178.

[169] R. Zhou, E. Bozbas, K. Allen-Redpath, and P. Yaqoob, “Circulating Extracellular Vesicles Are Strongly Associated With Cardiovascular Risk Markers,” *Front Cardiovasc Med*, vol. 9, May 2022, doi: 10.3389/fcvm.2022.907457.

[170] M. E. Kranendonk *et al.*, “Extracellular vesicle markers in relation to obesity and metabolic complications in patients with manifest cardiovascular disease,” *Cardiovasc Diabetol*, vol. 13, no. 1, p. 37, Dec. 2014, doi: 10.1186/1475-2840-13-37.

[171] A. Gualerzi *et al.*, “Raman spectroscopy as a quick tool to assess purity of extracellular vesicle preparations and predict their functionality,” *J Extracell Vesicles*, vol. 8, no. 1, 2019, doi: 10.1080/20013078.2019.1568780.

[172] E. A. Kwizera *et al.*, “Molecular detection and analysis of exosomes using surface-enhanced Raman scattering gold nanorods and a miniaturized device,” *Theranostics*, vol. 8, no. 10, pp. 2722–2738, 2018, doi: 10.7150/thno.21358.

[173] M. Hardy, L. Kelleher, P. de Carvalho Gomes, E. Buchan, H. O. M. Chu, and P. Goldberg Oppenheimer, “Methods in Raman spectroscopy for saliva studies–a review,” *Applied Spectroscopy Reviews*, vol. 57, no. 3. Taylor and Francis Ltd., pp. 177–233, 2022. doi: 10.1080/05704928.2021.1969944.

[174] T. Deckert-Gaudig, A. Taguchi, S. Kawata, and V. Deckert, “Tip-enhanced Raman spectroscopy-from early developments to recent advances,” *Chem Soc Rev*, vol. 46, no. 13, pp. 4077–4110, 2017, doi: 10.1039/c7cs00209b.

[175] S. Mahapatra, L. Li, J. F. Schultz, and N. Jiang, “Tip-enhanced Raman spectroscopy: Chemical analysis with nanoscale to angstrom scale resolution,” *J Chem Phys*, vol. 153, no. 1, p. 10902, 2020, doi: 10.1063/5.0009766.

[176] E. J. Matheson, S. M. Hamilton, and K. Kyser, “Shallow groundwater salinization of the Niagara Peninsula, Ontario, Canada,” *Geochemistry: Exploration, Environment, Analysis*, vol. 18, no. 2, pp. 155–174, 2018, doi: 10.1144/geochem2017-072.

[177] R. S. Krishnan and R. K. Shankar, “Raman effect: History of the discovery,” *Journal of Raman Spectroscopy*, vol. 10, no. 1, pp. 1–8, Jan. 1981, doi: 10.1002/jrs.1250100103.

[178] C. V. RAMAN and K. S. KRISHNAN, “A New Type of Secondary Radiation,” *Nature*, vol. 121, no. 3048, pp. 501–502, Mar. 1928, doi: 10.1038/121501c0.

[179] P. Vandenabeele, H. G. M. Edwards, and J. Jehlička, “The role of mobile instrumentation in novel applications of Raman spectroscopy: Archaeometry, geosciences, and forensics,” *Chemical Society Reviews*, vol. 43, no. 8. Royal Society of Chemistry, pp. 2628–2649, 2014. doi: 10.1039/c3cs60263j.

[180] H. Götz, M. A. Klukowska, H. Duschner, and D. J. White, “Physical, morphological, and micro-Raman chemical studies on bleaching strip effects on enamel, coronal dentin, and root dentin.,” *J Clin Dent*, vol. 18, no. 4, pp. 112–119, 2007.

[181] B. Yu, M. Ge, P. Li, Q. Xie, and L. Yang, “Development of surface-enhanced Raman spectroscopy application for determination of illicit drugs: Towards a practical sensor,” *Talanta*, vol. 191, pp. 1–10, 2019, doi: 10.1016/j.talanta.2018.08.032.

[182] A. Douglas, F. Skoog, J. Holler, and R. Stanley, *Principles of Instrumental Analysis*. 2018.

[183] J. M. Chalmers, H. G. M. Edwards, and M. D. Hargreaves, Eds., *Infrared and Raman Spectroscopy in Forensic Science*. Wiley, 2012. doi: 10.1002/9781119962328.

[184] F. A. Willard, H H; Merritt, Jr, L L; Dean, J A; Settle, Jr, “Instrumental Methods of Analysis,” in *Instrumental Methods of Analysis*, 7th ed., Washington, United States: Osti Gov, 1988.

[185] G. S. Bumbrah and R. M. Sharma, “Raman spectroscopy – Basic principle, instrumentation and selected applications for the characterization of drugs of abuse,” *Egypt J Forensic Sci*, vol. 6, no. 3, pp. 209–215, Sep. 2016, doi: 10.1016/j.ejfs.2015.06.001.

[186] R. L. McCreery, *Raman spectrscopy for chemical analysis*. John Wiley & Sons, Ltd, 2005.

[187] E. Smith and G. Dent, *Modern Raman spectroscopy: A Practical Approach*. John Wiley & Sons, Ltd, 2019.

[188] J. Addis, N. Mohammed, O. Rotimi, D. Magee, A. Jha, and V. Subramanian, “Raman spectroscopy of endoscopic colonic biopsies from patients with ulcerative colitis to identify mucosal inflammation and healing,” *Biomed Opt Express*, vol. 7, no. 5, p. 2022, 2016, doi: 10.1364/boe.7.002022.

[189] O. Chernavskaia *et al.*, “Beyond endoscopic assessment in inflammatory bowel disease: Real-time histology of disease activity by non-linear multimodal imaging,” *Sci Rep*, vol. 6, no. July, pp. 1–11, 2016, doi: 10.1038/srep29239.

[190] K. Eberhardt, C. Stiebing, C. Matthäus, M. Schmitt, and J. Popp, “Advantages and limitations of Raman spectroscopy for molecular diagnostics: an update,” *Expert Rev Mol Diagn*, vol. 15, no. 6, pp. 773–787, Jun. 2015, doi: 10.1586/14737159.2015.1036744.

[191] S. C. L. Smith *et al.*, “Raman spectroscopy accurately differentiates mucosal healing from non-healing and biochemical changes following biological therapy in inflammatory bowel disease,” *PLoS One*, vol. 16, no. 6 June, pp. 1–16, 2021, doi: 10.1371/journal.pone.0252210.

[192] A. Downes and A. Elfick, “Raman Spectroscopy and Related Techniques in Biomedicine,” *Sensors*, vol. 10, no. 3, pp. 1871–1889, Mar. 2010, doi: 10.3390/s100301871.

[193] I. J. Pence *et al.*, “Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy,” *Biomed Opt Express*, vol. 8, no. 2, p. 524, Feb. 2017, doi: 10.1364/BOE.8.000524.

[194] G. Acri *et al.*, “Raman Spectroscopy as Noninvasive Method of Diagnosis of Pediatric Onset Inflammatory Bowel Disease,” *Applied Sciences*, vol. 10, no. 19, p. 6974, Oct. 2020, doi: 10.3390/app10196974.

[195] T. Kirchberger-Tolstik *et al.*, “Towards an Interpretable Classifier for Characterization of Endoscopic Mayo Scores in Ulcerative Colitis Using Raman Spectroscopy,” *Anal Chem*, vol. 92, no. 20, pp. 13776–13784, 2020, doi: 10.1021/acs.analchem.0c02163.

[196] H. Ding *et al.*, “In vivo analysis of mucosal lipids reveals histological disease activity in ulcerative colitis using endoscope-coupled Raman spectroscopy,” *Biomed Opt Express*, vol. 8, no. 7, p. 3426, 2017, doi: 10.1364/boe.8.003426.

[197] C. Tefas *et al.*, “Surface-enhanced Raman scattering for the diagnosis of ulcerative colitis: will it change the rules of the game?,” *Anal Bioanal Chem*, vol. 413, no. 3, pp. 827–838, Jan. 2021, doi: 10.1007/s00216-020-03036-2.

[198] B. Li *et al.*, “Non-invasive diagnosis of Crohn’s disease based on SERS combined with PCA-SVM,” *Analytical Methods*, vol. 13, no. 44, pp. 5264–5273, 2021, doi: 10.1039/d1ay01377g.

[199] Y. Zhu *et al.*, “Surface-enhanced Raman spectroscopy analysis reveals biochemical difference in urine of patients with perianal fistula,” *Asian J Surg*, Jun. 2023, doi: 10.1016/j.asjsur.2023.05.137.

[200] X. Bi, A. Walsh, A. Mahadevan-Jansen, and A. Herline, “Development of Spectral Markers for the Discrimination of Ulcerative Colitis and Crohn’s Disease Using Raman Spectroscopy,” *Dis Colon Rectum*, vol. 54, no. 1, pp. 48–53, Jan. 2011, doi: 10.1007/DCR.0b013e3181fcf68d.

[201] M. Veenstra *et al.*, “Raman Spectroscopy in the Diagnosis of Ulcerative Colitis,” *European Journal of Pediatric Surgery*, vol. 25, no. 01, pp. 56–59, Aug. 2014, doi: 10.1055/s-0034-1387951.

[202] Y. Wu *et al.*, “The Specific Changes of Urine Raman Spectra Can Serve as Novel Diagnostic Tools for Disease Characteristics in Patients with Crohn’s Disease,” *J Inflamm Res*, vol. Volume 15, pp. 897–910, Feb. 2022, doi: 10.2147/JIR.S341871.

[203] S. He *et al.*, “Baseline correction for Raman spectra using an improved asymmetric least squares method,” *Anal. Methods*, vol. 6, no. 12, pp. 4402–4407, 2014, doi: 10.1039/C4AY00068D.

[204] H. Hu, J. Bai, G. Xia, W. Zhang, and Y. Ma, “Improved Baseline Correction Method Based on Polynomial Fitting for Raman Spectroscopy,” *Photonic Sensors*, vol. 8, no. 4, pp. 332–340, Dec. 2018, doi: 10.1007/s13320-018-0512-y.

[205] X. Wang and X. Chen, “Baseline Correction Based on a Search Algorithm from Artificial Intelligence,” *Appl Spectrosc*, vol. 75, no. 5, pp. 531–544, May 2021, doi: 10.1177/0003702820977512.

[206] J. Peng, S. Peng, A. Jiang, J. Wei, C. Li, and J. Tan, “Asymmetric least squares for multiple spectra baseline correction,” *Anal Chim Acta*, vol. 683, no. 1, pp. 63–68, Dec. 2010, doi: 10.1016/j.aca.2010.08.033.

[207] Åsmund Rinnan, F. van den Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra,” *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, Nov. 2009, doi: 10.1016/j.trac.2009.07.007.

[208] P. Heraud, B. R. Wood, J. Beardall, and D. McNaughton, “Effects of pre-processing of Raman spectra on in vivo classification of nutrient status of microalgal cells,” *J Chemom*, vol. 20, no. 5, pp. 193–197, May 2006, doi: 10.1002/cem.990.

[209] J. Laumer and S. K. O’Leary, “An approach to the spectral smoothing of Raman data applied to the specific case of thin-film carbon,” *Journal of Materials Science: Materials in Electronics*, vol. 29, no. 12, pp. 10026–10036, Jun. 2018, doi: 10.1007/s10854-018-9046-8.

[210] H. G. Schulze, S. Rangan, J. M. Piret, M. Blades, and R. Turner, “EXPRESS: Smoothing Raman Spectra with Contiguous Single-Channel Fitting of Voigt Distributions: An Automated, High Quality Procedure,” *Appl Spectrosc*, p. 000370281879495, Jul. 2018, doi: 10.1177/0003702818794957.

[211] C. Bielecki *et al.*, “Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells,” *J Biomed Opt*, vol. 17, no. 7, p. 0760301, Jul. 2012, doi: 10.1117/1.JBO.17.7.076030.

[212] I. Notingher *et al.*, “Multivariate analysis of Raman spectra for in vitro non-invasive studies of living cells,” *J Mol Struct*, vol. 744–747, pp. 179–185, Jun. 2005, doi: 10.1016/j.molstruc.2004.12.046.

[213] C. Banbury *et al.*, “Development of the Self Optimising Kohonen Index Network (SKiNET) for Raman Spectroscopy Based Detection of Anatomical Eye Tissue,” *Sci Rep*, vol. 9, no. 1, p. 10812, Dec. 2019, doi: 10.1038/s41598-019-47205-5.

[214] K. Hernández-Vidales, E. Guevara, V. Olivares-Illana, and F. J. González, “Characterization of wild-type and mutant p53 protein by Raman spectroscopy and multivariate methods,” *Journal of Raman Spectroscopy*, vol. 50, no. 10, pp. 1388–1394, Oct. 2019, doi: 10.1002/jrs.5655.

[215] K. Mehta, A. Atak, A. Sahu, S. Srivastava, and M. K. C, “An early investigative serum Raman spectroscopy study of meningioma,” *Analyst*, vol. 143, no. 8, pp. 1916–1923, 2018, doi: 10.1039/C8AN00224J.

[216] A. Daniel, A. Prakasarao, and S. Ganesan, “Near-infrared Raman spectroscopy for estimating biochemical changes associated with different pathological conditions of cervix,” *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 190, pp. 409–416, Feb. 2018, doi: 10.1016/j.saa.2017.09.014.

[217] A. Bonifacio, S. Cervo, and V. Sergo, “Label-free surface-enhanced Raman spectroscopy of biofluids: Fundamental aspects and diagnostic applications,”

Anal Bioanal Chem, vol. 407, no. 27, pp. 8265–8277, Nov. 2015, doi: 10.1007/s00216-015-8697-z.

[218] M. Paraskevaidi *et al.*, "Raman spectroscopic techniques to detect ovarian cancer biomarkers in blood plasma," *Talanta*, vol. 189, no. June, pp. 281–288, 2018, doi: 10.1016/j.talanta.2018.06.084.

[219] "Modern Raman Imaging Techniques," *Spectroscopy*, no. September 2020, 2020.

[220] E. J. Blackie, E. C. Le Ru, and P. G. Etchegoin, "Single-molecule surface-enhanced raman spectroscopy of nonresonant molecules," *J Am Chem Soc*, vol. 131, no. 40, pp. 14466–14472, 2009, doi: 10.1021/ja905319w.

[221] E. C. Le Ru, E. Blackie, M. Meyer, and P. G. Etchegoint, "Surface enhanced raman scattering enhancement factors: A comprehensive study," *Journal of Physical Chemistry C*, vol. 111, no. 37, pp. 13794–13803, 2007, doi: 10.1021/jp0687908.

[222] M. Moskovits, "Persistent misconceptions regarding SERS," *Physical Chemistry Chemical Physics*, vol. 15, no. 15, p. 5301, 2013, doi: 10.1039/c2cp44030j.

[223] S. A. Maier, *Plasmonics: Fundamentals and Applications*. New York, NY: Springer US, 2007. doi: 10.1007/0-387-37825-1.

[224] J. Langer *et al.*, "Present and Future of Surface-Enhanced Raman Scattering," *ACS Nano*, vol. 14, no. 1, pp. 28–117, Jan. 2020, doi: 10.1021/acsnano.9b04224.

[225] M. Baia, S. Astilean, and T. Iliescu, *Raman and SERS Investigations of Pharmaceuticals*. Heidelberg: Springer, 2008.

[226] R. A. Álvarez-Puebla, "Effects of the Excitation Wavelength on the SERS Spectrum," *J Phys Chem Lett*, vol. 3, pp. 857–866, 2012.

[227] I. Valpapuram *et al.*, "Waveguiding and SERS Simplified Raman Spectroscopy on Biological Samples," *BIOSENSORS-BASEL*, vol. 9, no. 1, Mar. 2019, doi: 10.3390/bios9010037.

[228] A. Stefancu *et al.*, "SERS-based liquid biopsy of saliva and serum from patients with Sjogren's syndrome," *Anal Bioanal Chem*, vol. 411, no. 22, pp. 5877–5883, Sep. 2019, doi: 10.1007/s00216-019-01969-x.

[229] L. Wu, Z. Wang, S. Zong, and Y. Cui, "Rapid and reproducible analysis of thiocyanate in real human serum and saliva using a droplet SERS-microfluidic chip," *Biosens Bioelectron*, vol. 62, pp. 13–18, 2014, doi: 10.1016/j.bios.2014.06.026.

[230] M. Eryılmaz, E. Acar Soykut, D. Çetin, İ. H. Boyacı, Z. Suludere, and U. Tamer, "SERS-based rapid assay for sensitive detection of Group A Streptococcus by evaluation of the swab sampling technique.," *Analyst*, vol. 144, no. 11, pp. 3573–3580, Jun. 2019, doi: 10.1039/c9an00173e.

[231] S. Han, A. K. Locke, L. A. Oaks, Y.-S. L. Cheng, and G. L. Cote, "Development of a free-solution SERS-based assay for point-of-care oral cancer biomarker detection using DNA-conjugated gold nanoparticles," in *OPTICAL DIAGNOSTICS AND SENSING XVIII: TOWARD POINT-OF-CARE DIAGNOSTICS*, Cote, GL, Ed., in Proceedings of SPIE, vol. 10501. 2018. doi: 10.1117/12.2290516.

[232] O. Durucan, K. Wu, M. Viehrig, T. Rindzevicius, and A. Boisen, "Nanopillar-Assisted SERS Chromatography," *ACS Sens*, vol. 3, no. 12, pp. 2492–2498, Dec. 2018, doi: 10.1021/acssensors.8b00887.

[233] M. Velicka, S. Adomaviciute, E. Zacharovas, and V. Sablinskas, “Application of label-free SERS and EC-SERS for detection of traces of drugs in biological fluids,” in *PLASMONICS IN BIOLOGY AND MEDICINE XVII*, VoDinh, T and Ho, HPA and Ray, K, Ed., in Proceedings of SPIE, vol. 11257. 2020. doi: 10.1117/12.2542306.

[234] M. Blanco-Formoso and R. A. Alvarez-Puebla, “Cancer Diagnosis through SERS and Other Related Techniques,” *Int J Mol Sci*, vol. 21, no. 6, Mar. 2020, doi: 10.3390/ijms21062253.

[235] G. Poste, “Bring on the biomarkers,” *Nature*, vol. 469, no. 7329, pp. 156–157, 2011, doi: 10.1038/469156a.

[236] A. Dadas, J. Washington, R. Diaz-Arrastia, and D. Janigro, “Biomarkers in traumatic brain injury (TBI): A review,” *Neuropsychiatr Dis Treat*, vol. 14, pp. 2989–3000, 2018, doi: 10.2147/NDT.S125620.

[237] L. E. Jones *et al.*, “Infrared and Raman screening of seized novel psychoactive substances: a large scale study of >200 samples,” *Analyst*, vol. 141, pp. 902–909, 2016, doi: 10.1039/c5an02326b.

[238] D. F. Ransohoff and M. L. Gourlay, “Sources of Bias in Specimens for Research About Molecular Markers for Cancer,” *Journal of Clinical Oncology*, vol. 28, no. 4, pp. 698–704, 2010, doi: 10.1200/JCO.2009.25.6065.

[239] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.

[240] J. I. Githaiga, H. K. Angeyo, K. A. Kaduki, W. D. Bulimo, and D. K. Ojuka, “Quantitative Raman spectroscopy of breast cancer malignancy utilizing higher-order principal components: A preliminary study,” *Sci Afr*, vol. 14, 2021, doi: 10.1016/j.sciaf.2021.e01035.

[241] W. Wu, H. Gong, M. Liu, G. Chen, and R. Chen, “Noninvasive breast tumors detection based on saliva protein surface enhanced Raman spectroscopy and regularized multinomial regression,” in *Proceedings - 2015 8th International Conference on BioMedical Engineering and Informatics, BMEI 2015*, Institute of Electrical and Electronics Engineers Inc., 2016, pp. 214–218. doi: 10.1109/BMEI.2015.7401503.

[242] S. Feng *et al.*, “Surface-enhanced Raman spectroscopy of saliva proteins for the noninvasive differentiation of benign and malignant breast tumors,” *Int J Nanomedicine*, vol. 10, pp. 537–547, 2015, doi: 10.2147/IJN.S71811.

[243] A. Hernández-Arteaga *et al.*, “Diagnosis of breast cancer by analysis of sialic acid concentrations in human saliva by surface-enhanced Raman spectroscopy of silver nanoparticles,” *Nano Res*, vol. 10, no. 11, pp. 3662–3670, 2017, doi: 10.1007/s12274-017-1576-5.

[244] A. C. Society, “Cancer Statistics.” 2021.

[245] Y. Wang *et al.*, “A New Method for the Early Detection of the Lung Cancer by the Saliva Tests Using Surface-Enhanced Raman Spectroscopy,” in *INTERNATIONAL SYMPOSIUM ON SIGNAL PROCESSING BIOMEDICAL ENGINEERING, AND INFORMATICS (SPBEI 2013)*, 2014, pp. 866–873.

[246] K. Qian, Y. Wang, L. Hua, A. Chen, and Y. Zhang, “New method of lung cancer detection by saliva test using surface-enhanced Raman spectroscopy,” *Thorac Cancer*, vol. 9, no. 11, pp. 1556–1561, Nov. 2018, doi: 10.1111/1759-7714.12837.

[247] G. Cao *et al.*, “A potential method for non-invasive acute myocardial infarction detection based on saliva Raman spectroscopy and multivariate analysis,” *Laser Phys Lett*, vol. 12, no. 12, p. 125702, 2015, doi: 10.1088/1612-2011/12/12/125702.
[248] V. Moisoiu *et al.*, “Combining surface-enhanced Raman scattering (SERS) of saliva and two-dimensional shear wave elastography (2D-SWE) of the parotid glands in the diagnosis of Sjogren’s syndrome,” *SPECTROCHIMICA ACTA PART A-MOLECULAR AND BIOMOLECULAR SPECTROSCOPY*, vol. 235, Jul. 2020, doi: 10.1016/j.saa.2020.118267.
[249] G. Eom *et al.*, “Diagnosis of tamiflu-resistant influenza virus in human nasal fluid and saliva using surface-enhanced raman scattering,” *ACS Sens*, vol. 4, no. 9, pp. 2282–2287, Sep. 2019, doi: 10.1021/acssensors.9b00697.
[250] C. Carlomagno *et al.*, “Human salivary Raman fingerprint as biomarker for the diagnosis of Amyotrophic Lateral Sclerosis,” *Sci Rep*, vol. 10, no. 1, Jun. 2020, doi: 10.1038/s41598-020-67138-8.
[251] B. N. Zamora-Mendoza *et al.*, “Surface-enhanced raman spectroscopy: A non invasive alternative procedure for early detection in childhood asthma biomarkers in saliva,” *Photodiagnosis Photodyn Ther*, vol. 27, pp. 85–91, 2019, doi: 10.1016/j.pdpdt.2019.05.009.
[252] O. Žukovskaja, I. J. Jahn, K. Weber, D. Cialla-May, and J. Popp, “Detection of Pseudomonas aeruginosa metabolite pyocyanin in water and saliva by employing the SERS technique,” *Sensors (Switzerland)*, vol. 17, no. 8, 2017, doi: 10.3390/s17081704.
[253] N. M. Ćulum *et al.*, “Characterization of ovarian cancer-derived extracellular vesicles by surface-enhanced Raman spectroscopy,” *Analyst*, vol. 146, no. 23, pp. 7194–7206, 2021, doi: 10.1039/D1AN01586A.
[254] M. Sbroscia *et al.*, “Thyroid cancer diagnosis by Raman spectroscopy,” *Sci Rep*, vol. 10, no. 1, p. 13342, Aug. 2020, doi: 10.1038/s41598-020-70165-0.
[255] I. J. Pence, E. Vargis, and A. Mahadevan-Jansen, “Assessing Variability of in Vivo Tissue Raman Spectra,” *Appl Spectrosc*, vol. 67, no. 7, pp. 789–800, Jul. 2013, doi: 10.1366/12-06773.
[256] Y.-Z. Zhang, “Inflammatory bowel disease: Pathogenesis,” *World J Gastroenterol*, vol. 20, no. 1, p. 91, 2014, doi: 10.3748/wjg.v20.i1.91.
[257] R. K. Pai *et al.*, “Definitions of response and remission for the Robarts Histopathology Index,” *Gut*, vol. 68, no. 11, pp. 2101–2102, Nov. 2019, doi: 10.1136/gutjnl-2018-317547.
[258] P. Rutgeerts, S. Vermeire, and G. Van Assche, “Mucosal healing in inflammatory bowel disease: impossible ideal or therapeutic target?,” *Gut*, vol. 56, no. 4, pp. 453–455, Apr. 2007, doi: 10.1136/gut.2005.088732.
[259] S. Chang, “Disease monitoring in inflammatory bowel disease,” *World J Gastroenterol*, vol. 21, no. 40, p. 11246, 2015, doi: 10.3748/wjg.v21.i40.11246.
[260] S. Nikolaus and S. Schreiber, “Diagnostics of Inflammatory Bowel Disease,” *Gastroenterology*, vol. 133, no. 5, pp. 1670–1689, Nov. 2007, doi: 10.1053/j.gastro.2007.09.001.
[261] W. Strober, I. Fuss, and P. Mannon, “The fundamental basis of inflammatory bowel disease,” *Journal of Clinical Investigation*, vol. 117, no. 3, pp. 514–521, Mar. 2007, doi: 10.1172/JCI30587.

[262] P. Rutgeerts *et al.*, “Adalimumab Induces and Maintains Mucosal Healing in Patients With Crohn’s Disease: Data From the EXTEND Trial,” *Gastroenterology*, vol. 142, no. 5, pp. 1102-1111.e2, May 2012, doi: 10.1053/j.gastro.2012.01.035.

[263] M. Norouzinia, V. Chaleshi, A. H. M. Alizadeh, and M. R. Zali, “Biomarkers in inflammatory bowel diseases: Insight into diagnosis, prognosis and treatment,” *Gastroenterol Hepatol Bed Bench*, vol. 10, no. 3, pp. 155–167, 2017, doi: 10.22037/ghfbb.v0i0.1163.

[264] D. Liu, V. Saikam, K. A. Skrada, D. Merlin, and S. S. Iyer, “Inflammatory bowel disease biomarkers,” *Med Res Rev*, vol. 42, no. 5, pp. 1856–1887, Sep. 2022, doi: 10.1002/med.21893.

[265] Y. Fengming and W. Jianbing, “Biomarkers of Inflammatory Bowel Disease,” *Dis Markers*, vol. 2014, pp. 1–11, 2014, doi: 10.1155/2014/710915.

[266] P. Chen *et al.*, “Serum Biomarkers for Inflammatory Bowel Disease,” *Front Med (Lausanne)*, vol. 7, Apr. 2020, doi: 10.3389/fmed.2020.00123.

[267] B. E. Sands, “Biomarkers of Inflammation in Inflammatory Bowel Disease,” *Gastroenterology*, vol. 149, no. 5, pp. 1275-1285.e2, 2015, doi: 10.1053/j.gastro.2015.07.003.

[268] G. E. Reese *et al.*, “Diagnostic Precision of Anti-Saccharomyces cerevisiae Antibodies and Perinuclear Antineutrophil Cytoplasmic Antibodies in Inflammatory Bowel Disease,” *Am J Gastroenterol*, vol. 101, no. 10, pp. 2410–2422, Oct. 2006, doi: 10.1111/j.1572-0241.2006.00840.x.

[269] K. Theede, S. Holck, P. Ibsen, S. Ladelund, I. Nordgaard-Lassen, and A. M. Nielsen, “Level of Fecal Calprotectin Correlates With Endoscopic and Histologic Inflammation and Identifies Patients With Mucosal Healing in Ulcerative Colitis,” *Clinical Gastroenterology and Hepatology*, vol. 13, no. 11, pp. 1929-1936.e1, Nov. 2015, doi: 10.1016/j.cgh.2015.05.038.

[270] J.-F. Lin *et al.*, “Meta-analysis,” *Inflamm Bowel Dis*, vol. 20, no. 8, pp. 1407–1415, Aug. 2014, doi: 10.1097/MIB.0000000000000057.

[271] C. Maaser *et al.*, “ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications,” *J Crohns Colitis*, vol. 13, no. 2, pp. 144-164K, Feb. 2019, doi: 10.1093/ecco-jcc/jjy113.

[272] F. D’Amico *et al.*, “International consensus on methodological issues in standardization of fecal calprotectin measurement in inflammatory bowel diseases,” *United European Gastroenterol J*, vol. 9, no. 4, pp. 451–460, May 2021, doi: 10.1002/ueg2.12069.

[273] D. Labaere *et al.*, “Comparison of six different calprotectin assays for the assessment of inflammatory bowel disease.,” *United European Gastroenterol J*, vol. 2, no. 1, pp. 30–7, Feb. 2014, doi: 10.1177/2050640613518201.

[274] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, “Wireless capsule endoscopy,” *Nature*, vol. 405, no. 6785, pp. 417–417, May 2000, doi: 10.1038/35013140.

[275] I. Hilmi and T. Kobayashi, “Capsule endoscopy in inflammatory bowel disease: when and how,” *Intest Res*, vol. 18, no. 3, pp. 265–274, Jul. 2020, doi: 10.5217/ir.2019.09165.
[276] L. G. D’Cruz *et al.*, “Assessment of a dried blood spot C-reactive protein method to identify disease flares in rheumatoid arthritis patients,” *Sci Rep*, vol. 10, no. 1, p. 21089, Dec. 2020, doi: 10.1038/s41598-020-77826-0.
[277] Nexus-Dx, “No Title.”
[278] S. Diagnostics, “No Title,” Cardiac STATus is a point-of-care cardiac panel test kit. [Online]. Available: https://www.hospitalnetwork.com/doc/point-of-care-cardiac-test-kit-0001
[279] Alfa Scientific, “No Title,” Instant-view® Cardiac 3-in-1 Whole Blood/Serum Test. [Online]. Available: https://www.alfascientific.com/products/cardiac-3-in-1/
[280] LifeSign, “No Title,” LifeSign MI CK-MB/Myo/TnI.

CHAPTER 3

SPECTROSCOPIC MOLECULAR-FINGERPRINT PROFILING OF SALIVA

This results chapter is based on the following published article;

1. **Buchan E.,** Kelleher L., Clancy M., Rickard J.J.S., and Goldberg Oppenheimer P. [2021], 'Spectroscopic molecular-fingerprint profiling of saliva.' *Analytica Chimica Acta,* **1185,** 339074.

Author contributions - **E Buchan:** Conceptualization, Investigation, Methodology, Formal analysis, Writing – original draft, review and editing. **L Kelleher:** SKiNET algorithm data classification. **M Clancy:** Methodology, Formal analysis, Software. **J Rickard:** Validation, Writing – original draft, Writing – review & editing, Data curation. **P Goldberg Oppenheimer:** Conceptualization, Methodology, Writing – review & editing, Resources, Funding acquisition, Supervision.

3.1 Abstract

Saliva analysis has been gaining interest as a potential non-invasive source of disease indicative biomarkers due to being a complex biofluid correlating with blood-based constituents on a molecular level. For saliva to cement its usage for analytical applications, it is paramount to gain underpinning molecular knowledge and establish a 'baseline' of the salivary composition in healthy individuals as well as characterize how these factors are impacting its performance as potential analytical biofluid. Here, we have systematically studied the molecular spectral fingerprint of saliva, including the changes associated with gender, age, and time. *Via* hybrid artificial neural network algorithms and Raman spectroscopy, we have developed a non-destructive molecular profiling approach enabling the assessment of salivary spectral changes yielding the determination of gender and age of the biofluid source. Our classification algorithm successfully identified the gender and age from saliva with high classification accuracy. Discernible spectral molecular 'barcodes' were subsequently constructed for each class and found to primarily stem from amino acid, protein, and lipid changes in saliva. This unique combination of Raman spectroscopy and advanced machine learning techniques lays the platform for a variety of applications in forensics and biosensing.

3.2 Introduction

In the past decade saliva has been emerging as a potential candidate bodily fluid for medical diagnostics [1]–[3]. Owing to its non-invasive collection, simple, or no, pre-processing required prior to analysis and yet, a similar rich molecular composition as of blood, this complex biofluid has been increasingly gaining scientific interest [4], [5]. Like blood, saliva is rich in biochemical information and is comprised of a complex

mixture of proteins, lipids and many biological and inorganic components including hormones, antibodies, and growth factors [1], [6], [7] as well as a breadth of disease biomarkers [4], [8]–[10]. Thus, saliva is an attractive candidate biofluid to complement the routine use of blood for patient screening in diagnostics[11], dentistry [12], [13] or for evidence collection in forensics [14]–[16]. In contrast to blood however, saliva exhibits added advantages including the straightforward and less-demanding collection procedures, alleviating the discomfort to the subject in question and, the less physically intrusive but more cost-effective analyses options. Saliva procurement is non-invasive and painless, does not require highly trained personnel, and the collected samples are easily stored and transported. Importantly, in contrast to blood, saliva does not clot with time. Additionally, salivary secretions from individuals diagnosed with blood borne pathogens such as HIV have negligible rates of oral transmission due to factors that inhibit the infectivity of pathogens and therefore, are considerably safer to handle.

Saliva therefore, has been increasingly exploited for biomedical applications in diagnostic assays including, the detection of acute myocardial infarction and inflammatory disorders [17]–[20]. Salivary research is also ideal for investigations relating to cancer research and diagnostics. It has been applied in oral, gastric, breast, lung, prostate and ovarian cancers to name a few [4], [5], [8], [10], [21], [22].

Concurrently, Raman spectroscopy (RS), a form of a sensitive vibrational spectroscopy, which by exciting molecular bonds within a sample, provides a unique biomolecular spectral fingerprint of target analytes. Spectroscopic ‘fingerprints’ of

Raman spectra have been shown to identify the health/disease state of the subject from which the biofluid has been collected [4], [5], [8], [9]. In contrast to the *in-vitro* bioassays, the availability of inexpensive, portable RS instruments makes this technique particularly attractive for point-of-care sampling, analysis and screening of biofluids such as, saliva.

There are a number of Raman spectroscopy studies where saliva has been deployed towards exploring applications in healthcare, in areas such as CVD and IBD [27], [28], therapeutics in chemotherapy [29]–[31], detection of oral cancer [21], [32], [33] and dental disease diagnostics [12], [34], [35]. SERS and microfluidics were further used to enhance the spectral response of saliva and to create small test platforms to be used [8], [36], [37]. However, none of these studies paid attention to the protein content of saliva samples and understanding of the role of proteins in the differentiation of abnormal cells from the control, protein backbone vibrations and the amide III regions. Therefore, understanding the protein content of saliva is necessary to take full advantage of the synergistic functions of saliva. The inherent inhomogeneity of saliva is most often overlooked by researchers, hence missing the fundamental characteristics of the sample. The utilisation of Raman spectroscopic profiling of healthy saliva also remains rather unexplored. Currently, there are no confirmatory tests specific to saliva in the field of forensic sciences. Although the presumptive testing is based on the activity of amylase in the sample, it is found in two different forms within the human body and is therefore, not exclusive to saliva.

Whilst research continues unravelling saliva as a potential alternative to blood- and tissue-based diagnostics, it is imperative to establish a 'baseline' of inherent characteristics of this biofluid in healthy subjects, where the detected changes *via* Raman spectroscopy could be attributed to the underpinning variations in gender, age and function of temporal evolution. The only study, which has examined the gender related spectroscopic variations, is by Muro *et al.* who used internal validation of saliva samples *via* Support Vector Machine Discriminant Analysis (SVMDA) for forensics discrimination [38]. While the SVMDA model proved successful, no comparison of spectral changes due to gender was established to indicate spectral regions of change, and the authors concluded that there is a need for advanced multivariate analysis due to the high standard deviation of measurements [38].

Here, we present a proof-of-concept Raman spectroscopy profiling and classification of saliva as a function of age and gender as well as temporal evaluation, towards establishing an important baseline of healthy saliva in a form of a "multi-biochemical barcode", as a characteristic tool for ongoing and future spectroscopic studies for diagnostic. We examine and systematically evaluate what is the spectral variability between healthy individuals, divided to female and male groups, as well as the main three representative age groups of young, middle-age and older adults and how these affect Raman spectra while gaining insights on the biochemical interpretation of spectral data. The effect of time from collection of saliva on spectral inter-variability is further examined to establish whether saliva as a biofluid exhibits major temporal changes in comparison to blood, which has been shown to considerably vary as a function of time from collection. The acquired spectral data is classified using our new

artificial neural networks algorithm, SKiNET as a decision support tool [39]. SKiNET is based on the separation of data classes in a self-organising map (SOM) with characterisation using a self-organising map discriminant index (SOMDI) enabling the subsequent classification of the tested data. The SOM is a model inspired by nature and the way that neurons in the visual cortex are spatially organised according to the type of visual stimuli. The SOM defines a 2D map of neurons, typically arranged as a grid of hexagons. Each neuron is assigned a weight vector, which is initialised randomly and has a length equal to the number of variables in a spectrum. The weight vector affects which neuron will be activated for a given sample and neighbouring neurons will have similar weights. Spatial clustering is therefore observed in the trained map, with spectra that exhibit distinct properties activating different neurons. To understand which features in the data cause certain neurons to activate over others, SOMDI is used. The SOMDI introduces class vectors as labels for each spectrum and corresponding weight vectors for each neuron, without influencing the training process. These allow for the identification of what type of data a given neuron activates, which can be used to inspect the weights across all neurons and extract prominent features belonging to each class. Neurons (hexagons) are coloured according to the modal class they activate, from the training set of Raman spectra. Neurons that have no majority class or activate none of the training data are shown in white. Coloured circles within each neuron represent spectra from the training data that have been activated for that neuron. For each class, there is a clearly defined block of neurons, with many of these activating only a single tissue type. An approximately even distribution in the number of neurons required to identify each class observed. The SOMDI provides a representation of weights associated with neurons that identify a particular class. A

higher SOMDI intensity indicates a greater importance of particular inverse centimetres along the axis of a spectrum. Well defined peaks can be easily resolved, which are either more prominent or unique to each class. Automated classification of Raman spectra and assignment to a particular tissue type or disease state is perhaps the most important step for the translation of Raman based diagnostic techniques to real world, clinical applications.

Through inspection of key differences between neuron weights and class weight vectors, the algorithm enables identification of the key spectral changes. Training parameters used for the SOM included the grid size of 7, the learning rate of 0.2 and 10 epochs. From the separation of classes, it is evident that there are characteristic differences due to the obvious classification of certain neurons. As such, there is a clear basis for differentiation enabling characteristic weight vectors to be derived in the SOMDI. The identified barcodes from this study act as a reference, constituting a solid basis towards developing standard protocols as an essential prerequisite for reliable studies aimed at establishing the feasibility of Raman spectroscopy as analytical tool for diagnostics and forensics. Molecular barcodes can further be constructed for distinguishing between disease and healthy states associated with spectral changes *via* an easy subtraction of the variations from the reference sample spectra. In conjunction with the emergence of state-of-the-art machine learning techniques, the development of reliable, rapid spectroscopic analytical tools, ultimately promises improved forensic science technologies, helping to identify a possible victim, suspect or criminal as well as a better quality and timeliness of disease diagnostics and tailored treatments.

3.4 Results and Discussion

We have established a process aimed to identify how age and gender of the participants vary and impact the characteristic Raman fingerprints of the extracted saliva samples as well as to uncover the effects of time on saliva. Namely it includes saliva collection, centrifugation deposition onto an aluminium slide and measuring the corresponding Raman spectra (**Fig. 3.1**).

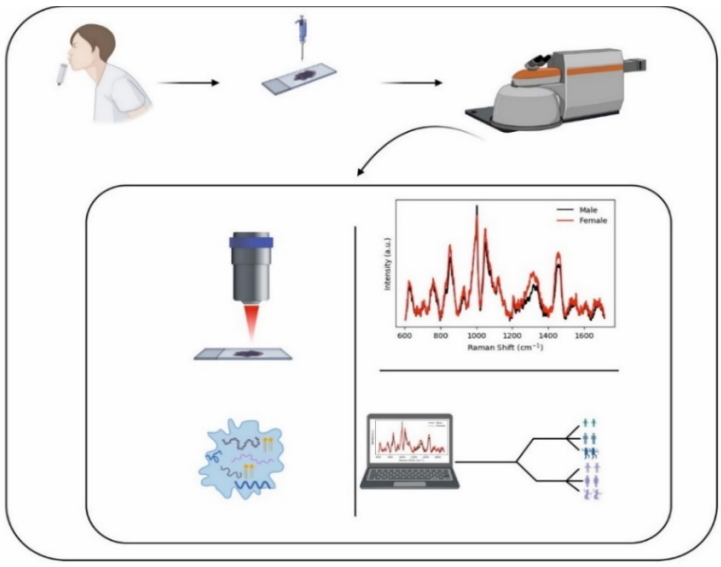

Figure 3.1. Schematic representation of the experimental procedure for collecting and measuring saliva *via* Raman spectroscopy with the subsequent analysis of age, gender and time parameters. (Figure created in Biorendor.com)

The measurement procedure was optimised to achieve high signal-to-noise spectra, with a 785 nm laser used for probing the samples composition due to its low levels fluorescence induction, minimal sample damage and good response in the fingerprint region [42]. Fresh saliva samples, from an overall of 38 female and 32 male participants with an age groups ranging between 22-70 y (<39>±15 y) and 24-80 y (<40>±16 y) respectively, were analysed using our recently developed neural network

SKiNET algorithm [41]. The derived weight vectors characteristic to each class are shown in **Fig.6.2c-e**. The greater the normalised intensity of each wavenumber, the more important that scatter band is to a class' spectral fingerprint.

Participants have been further sub-divided into age groups consisting of young adults (20-25 and 26-30 years), mid-aged adults (31-55 years) and older adults (56+ years) and the representative mean spectra of male (*n*=32) and female (*n*=38) saliva for each group are shown in **Fig.3.2a-b**. Age selection criteria was based on healthy adults with no known medical conditions that fit into each of the age ranges (*see* **'Materials and Methods'**) [15, 38, 68, 70]. The most prominent spectral features of the saliva for these groups are found at 630, 760, 855, 930, 1003, 1051, 1125, 1205, 1304, 1335, 1455, and 1655 cm^{-1}, with the associated assignments summarised in **Table 3.1**.

Table 3.1. Characteristic Raman peak assignments for the identified peak wavenumbers [9,10, 33,48].

Peak Wavenumber (cm^{-1})	**Assignment**
628	C-S stretch and C-C twist Protein; tyrosine
760	Ring breathing mode tryptophan; proteins
855	C-C; ring breathing mode tyrosine
930	C-C stretch amino acids (proline, hydroxyproline and valine); proteins
960	Calcium-phosphate stretching band (cholesterol), *α*-helix Proline, Valine (*n* (C-C))
1003	Symmetric ring breathing mode (phenylalanine, tryptophan)
1051	C-O and C-N stretch
1070	HCO_3
1076	C-C (Lipids); symmetric stretch of phosphates in hydroxyapatite
1087	Vibration of DNA backbone; C-N stretch (proteins); PO_3^{-2} anti-symmetric stretch

1125	C-C skeletal stretch (lipids); C-N stretch (proteins)
1205	Amide III; CH_2 wagging and vibrations (glycine, proline, tyrosine and phenylalanine)
1300	Amide III; CH_2 twisting and wagging (lipids and/or proteins)
1337/1339	CH_2/CH_3 wagging and twisting (proteins, nucleic acid, lipids), nucleic acid bases (*n* (C-H))
1456	CH_2 and CH_3 deformation vibrations (proteins and lipids)
1655	Amide I region; C=C stretch (lipids); C=O stretch (proteins)

Gender based spectral differences of male saliva (black) and female saliva (red) are found to vary significantly (**Fig.3.2e**). SOMDI scores show that it is intensity changes of the saliva rather than peak shifts that are responsible for the main differences. The male spectra were found to be distinct from female spectra due to the increased response at 630, 760, and 1003 cm^{-1} bands, with the female spectra showing an increased 855 cm^{-1} and 1300-1400 cm^{-1} response. Subtle differences are further noticeable at the 1051 and 1455 cm^{-1} shifts. The overall gender of saliva samples (independent of age) has been classified with an accuracy of 93.0±0.5% with a 10-fold cross validation. Protein, lipid, and amino acid assigned peaks are identified to represent the main changes in the spectra, accordingly.

The peak at 630 cm^{-1}, attributed to phenylalanine and linked to metabolites concentration, is decreased in the female spectra where salivary metabolites such as, the taurine and lactate are commonly identified [43]. The 1003 cm^{-1} peak is typically assigned to phenylalanine as well, but may also be attributed to lysine, which is known to be one of the most abundant amino acids found in saliva [44]. Phenylalanine is found in both the sequences for amylase and lipase, two of the major components of saliva. Phenylalanine, as an essential amino acid typically takes 2-3 hours to be fully absorbed in the body and up to 6 hours to breakdown. As preliminary studies aimed to identify

optimal salivary collection times showed no significant change in this peak after eating in individuals that were re-sampled, we determined that the effect due to ingestion of phenylalanine was negligible and thus was due to differences between males and females. A subtle peak shift is observed for the amide III region (1205-1300 cm^{-1}) showing that the composition of saliva in both males and females is distinct from each other. A difference in salivary composition can be linked to flow rate, where Inoue *et al.* showed that the flow rate of saliva in females is considerably lower than in males as their submandibular glands, where unstimulated saliva is mainly derived from, are much smaller than their male counterparts [45]. Lastly, the Raman spectra of saliva show a lack of response in the amide I region. Amide I along with aromatic breathing and C-H stretching peaks typically indicate a significant presence of protein in a sample [8], [13].

Of a further consideration in a gender related spectroscopic study of female saliva is the effect of oestrogen on the salivary metabolome [46]. Salivary flow rate is significantly reduced in post-menopausal women in comparison with those who are menstruating (typically, ages of 45-50 are considered as the average menopause range) [47-49]. It has also been observed that salivary pH is significantly lower in post-menopausal women and the pH of the mouth plays a key role in the salivary Raman response [47]. The normal pH of saliva is 6 to 7 however, this can vary from 5.3 (low flow) to 7.8 (peak flow). The buffer capacity of saliva includes protein (amylase, mucin and immunoglobulin A), bicarbonate and phosphate buffer systems and their main role is to maintain a constant pH [50-51]. The concentration of bicarbonate detected in saliva is dependent on salivary flow rate. Thus, when the flow rate is increased, both

bicarbonate levels and the pH also increase, leading to variance in the saliva spectra. Bicarbonate is observed at 1070 cm^{-1} and **Fig. 3.2e** shows a higher peak intensity in males than females indicating a higher pH and thus, a faster flow rate. In addition, women are affected more often than men with autoimmune diseases, which are known to affect salivary gland function [52]. Therefore, this could also have an effect on the salivary flow rate, as for example in the Sjögren's syndrome, where one of the most common symptoms is a dry mouth [53].

The peak at 1340 cm^{-1}, known to be associated with collagen, has a higher intensity in females than males. Collagenase 3 (MMP-13) expression was first observed in breast cancer and is readily associated with diseases such as the osteoarthritis, which females are more likely to develop than men.

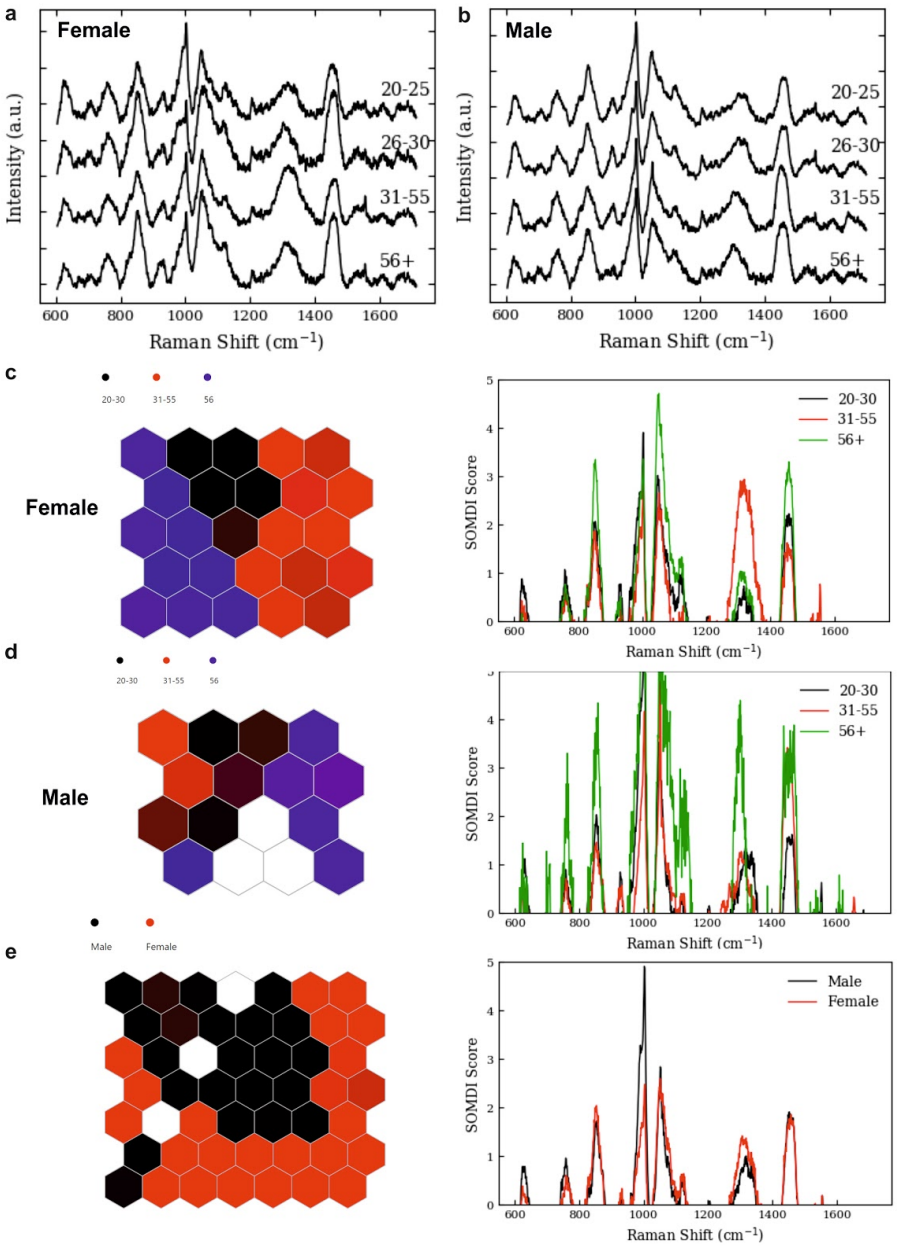

Figure 3.2. Representative mean spectra categorised by age for **(a)** female and **(b)** male saliva. SOM (left) and SOMDI (right) of **(c)** female and **(d)** male saliva, successfully classifying the saliva samples according to age with an accuracy of 70.0±1.2% and 78.0±0.1%, respectively. The sub-classification according to age for female and male samples are represented by black hexagons for 20-30 years, red hexagons for 31-55 years and purple hexagons for the age group above 56 years. **(e)** Classification of female (red) *versus* male (black) saliva with an accuracy of 93.0±0.5%. The biggest changes are observed at 630, 760, 855 and 1300 cm^{-1} bands.

A recent study by Virtanen *et al.* indicated an increased MMP-13 levels in female saliva [55]. The link between MMP-13 and its initial detection in breast cancer may potentially explain the increase observed in women and it is plausible that its metabolism could be an indication of hormonal associations. MMP-13 is also known to initiate bone resorption and activate proMMP-9. Both matrix metalloproteinases have been associated with periodontitis. Hernandez *et* al. identified elevated levels of MMP-13 in patients with chronic periodontitis indicating MMP-13 may play a role in the destruction of periodontal soft tissue [56].

Interestingly, age related spectral changes of saliva split the male and female datasets further into chronological age categories of young (20-30 years), mid (31-55 years) and older (56+ years) adults. These results, for the first time, indicate that it is possible to detect and monitor the age of a person using Raman spectroscopy of fresh saliva for both male and females.

Age based results empirically showed that the 20-25 and 26-30 age groups are closely clustered together and therefore, these have been grouped together for the further analysis. The variation in signal intensity for the four most prominent peaks as a function of the various ages is shown in **Fig.3.3a-b** for female and male saliva. The largest changes noted are for the 56+ age category, where the female spectral fingerprints are found to change the most. The largest differences in peak intensity are observed in the 56+ age group at 855, 1076 and 1456 cm^{-1}, accounting for the ring breathing mode and C-C stretching, respectively. For female saliva samples, there is a clear separation according to the age groups of 20-30 y (black), 31-55 y (red), and

>56y (green), classified with an accuracy of 70.0±1.2% (**Fig.6.2c, left**). The SOMDI, akin to a PCA loading plot, shows in order of decreasing influence that the 1051, 1003, 1456, 855cm^{-1} peaks as well as the region of 1300-1400 cm^{-1} account for the most variance in the classes. It is further noted that the older adult category, *i.e.,* >56 y, is the most distinctive age group, where an increase in peak intensity in the 1076, 1455, and 855cm^{-1} is identified. The younger adult category of female spectra is distinctive by the decreased intensity in the 1300-1400 cm^{-1} region and the increased intensity at the 1003, 630, and 760cm^{-1} peaks. Similarly, male age groups of saliva also exhibit the biggest variance in the ring breathing mode, C-N stretches for the >56y group with an increased peak intensity at 1455, 1051, and 855 cm^{-1}. However, unique to the >56y male group, is the increase in the 760 cm^{-1} and in the 1300-1400 cm^{-1} peaks, attributed to the tryptophan ring breathing mode and amide III region (proteins). For the 20-30 y and 31-55 y age groups, a Raman shift occurs in the 1200-1300cm^{-1} region to 1300-1375 cm^{-1}.

It is well-established that as people age, the rate, volume and composition of saliva changes [57]. Histomorphometric examinations of healthy salivary gland tissues have revealed a decrease in the number of acinar cells as in individual ages. Xu *et* al. hypothesised that despite a loss of acinar cells, a secretory reserve exists to maintain the function of the salivary gland, thus potentially accounting for salivary changes with age. We indeed observe this effect by the shift in the amide III peak of saliva for both genders although, it is more distinct in the male saliva. Both lysine (1003 cm^{-1}) and glycine (1327 cm^{-1}) are known to be the most abundant amino acids in saliva and increase in concentration in males and females as they age [57]. In addition, Nagler *et*

al. discuss differences in salivary gland size and the implication this would have on salivary components in the ageing process of both males and females [58]. The authors determined that older patients tend to commonly experience xerostomia, causing dry mouth and affecting salivary production and speech, causing oral infections and dental caries. At least 25% of the elderly population have xerostomia and many display symptoms without realising it. Although all patients are screened for health conditions, elderly patients may not have been officially diagnosed with xerostomia and thus, it may explain the significant variation in the salivary spectra of the elderly.

Furthermore, the intensity changes observed in the 630 and 1003 cm^{-1} peaks between the 20-30 y and the >56 y groups indicate that the concentration of the associated amino acids is decreasing with increasing age. This is attributed to the reduction of the excretion rate of saliva with age and subsequently resulting in the decrease in the amino acids concentration [59]. In addition, a strong intensity change is observed between the age classes, with a higher response in the 56+ group at 1076 cm^{-1}. This band is attributed to lipids, with the spectra indicative of increased biological matter in older saliva (56+). Underpinning this, might be the lower saliva production rates, which can act as a discernible factor when compared to saliva of younger people. A meta-study by Affoo *et* al. found that saliva flow rate was significantly lower in older adults with such a decrease specifically attributed to the saliva from submandibular and sublingual salivary glands [60]. This decrease in flow rate can be also attributed to the loss of acinar cells, increase in adipose tissue and neurophysiological deterioration. In histopathological samples of salivary glands, glands of young adults show a more even

and compact lobar structure with a uniform appearance of parenchymal elements when compared to those of an older individual [57]. With age, histological studies indicate that although the number of salivary glands remain constant, the volume of fat and fibromuscular tissue increases in both the parotid and submandibular glands. Conversely, the authors also state that in elderly patients a reduced volume of acinar cell secretion is observed and is considered as one of the main causes of dry mouth. Combined, these changes can also lead to salivary gland hypofunction [57]. Also, with age the number of salivary glycoproteins increases as indicated in a mass spectroscopy study by Sun *et* al., where the *N*-glycoproteins were found to be related to innate immunity an individual has built over their lifetime against microorganisms and oral cavity protection [61]. In our case, the detected increase in proteins and the higher response in the over 56 group may indicate that a higher level of innate immunity has been established, thus a decrease in salivary flow rate appears to have an indirect influence on the overall quality of an individual's saliva.

One of the major components of saliva is amylase. However, two different forms are found in the human body, the AMYI and the AMY2. AMYI (sAA) is detected more in saliva than in any of the other biofluids and therefore, has the potential to act as a unique biomarker. Amylase contains phenylalanine in its sequence, resulting in a strong Raman peak at 1003 cm^{-1}. This peak is detected with a higher intensity in younger adults, 20-30 years of age, as opposed to the older 56+ group. Different studies have conflicting opinions regarding levels of AMYI and its age-related differences. Ben-Aryeh *et al.* indicated there was no statistical difference in AMY1 between age groups [62], however Yim *et* al., determined the opposite [63].

Rutherfurd-Markwick *et* al. have shown that the amylase content of saliva increases steadily up to the middle age prior to stabilising and remaining constant in older adults [64]. The authors observed that amylase activity at rest was higher in a group of women than men. AMYI is secreted by the salivary glands but at birth AMYI is not detected. Levels increase as an individual ages to adulthood. Similarly, salivary glands continually grow throughout childhood and mature during adulthood and thus, levels of AMYI indicate developmental differences and potentially influence Raman spectra of saliva. Other studies have found that diet and stress affect AMYI levels. Perry *et al.* have shown that populations or individuals with a high starch diet had a higher AMYI level than those with low-starch diets [65]. Furthermore, AMYI is secreted by the salivary glands in response to stressors. Chatterton *et al.* have shown that psychosocial and physical stressors can rapidly increase levels of AMYI.

In addition to amylase, lysozyme makes up another major component of saliva. The lysozyme peak was observed at 1343 $cm^{-1}$with distinct intensity changes between the age groups. In both males and females, the peak intensity was higher in older adults (30+). Lysozyme, a salivary protein plays a role in the nonimmunological bacterial defence system. A study by Lira-Junior *et al.* found that elderly individuals (>64 years of age) had elevated counts of 24 out of the 41 investigated bacteria when compared to younger adults (<64 years of age) [66]. An increase in bacteria in older adults may account for the increase observed in lysozyme due to lysozyme providing antimicrobial activity. The lysine peak found at 1000 cm^{-1} also increased in intensity as individuals aged. The peak was stronger for 56+ age group in both males and females. A study

by Tanaka *et al.* observed similar results with significantly increased lysine levels (p<0.05) with ageing regardless of gender [67].

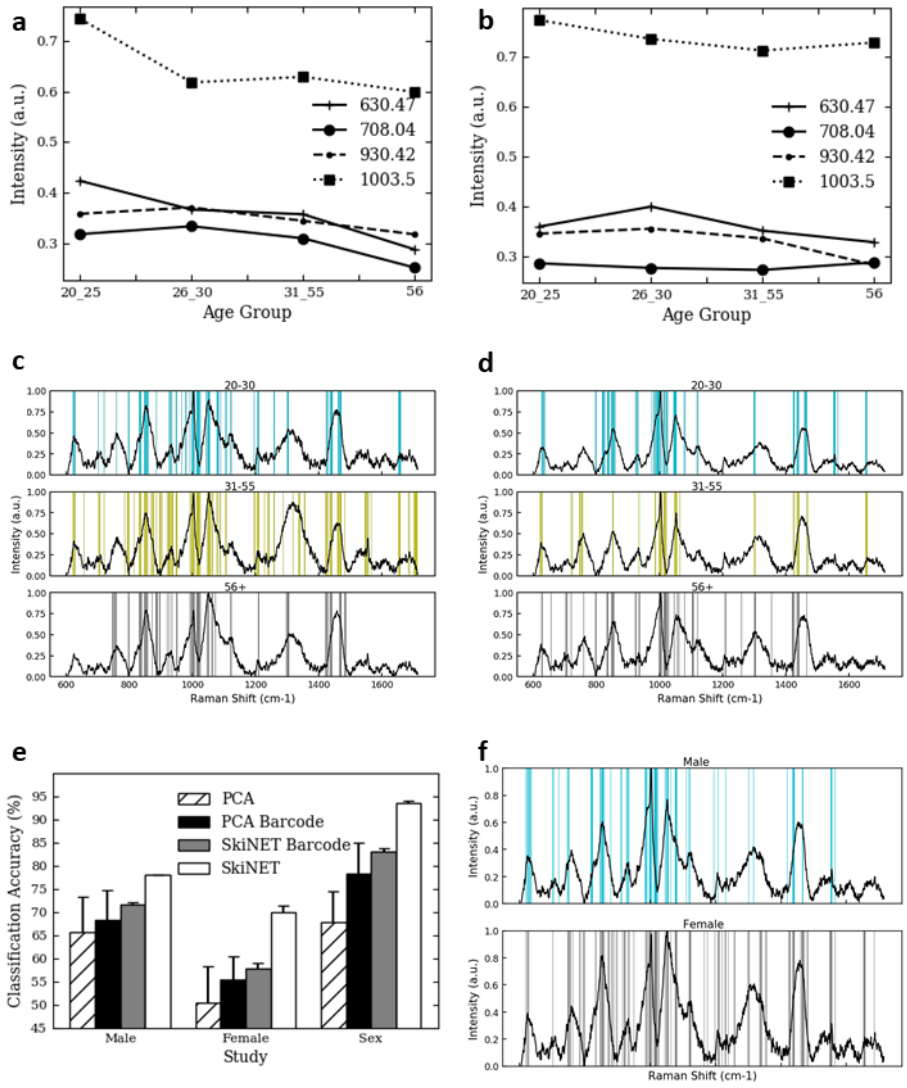

Figure 3.3. Variation in Raman intensity with age for **(a)** female and **(b)** male saliva. Barcodes derived from Raman spectra of saliva for age groups of 20-30y, 31-55y and >56y for **(c)** females and (**d)** males. **(e)** Comparison of PCA and SKiNET based classification of Raman profiling of saliva according to gender and age. **(f)** Barcoding analysis of male and female saliva for overall age groups.

Total protein concentration in saliva also varies with age [66]. Average protein concentration has been shown to decrease from 3.25 mg/ml for 20-30 years young adult to 2.71 mg/ml for >56 years adult groups [67]. Furthermore, females exhibited a

considerably higher total protein concentrations than males, of 3.06 and 2.60 mg/ml, respectively [68]. Mucins (observed in the amide I region) are also more readily detected in young adults (20-30 years) than older adults (>56 years) [69]. Lower molecular weight proteins such as, the Statherin, Histatin (997cm^{-1}) and Cystin (678 cm^{-1}) are also detected in young adults but not in older adults' saliva. Peroxidases, proline rich proteins (856 and 920 cm^{-1}) and IgA (1480 cm^{-1}) Raman bands are present in saliva across all age groups however, the concentrations differ with age as evidenced in **Fig.3.2c-d**. This indicates that there is an *age-dependant* influence on the specific secretion of one component but at other bands (930 and 1200 cm^{-1}) where a lower intensity is observed, the concentration effect is due to the drop in salivary output. Moreover, the higher peak intensities observed at 855, 1003 and 1449 cm^{-1} and an additional peak at 1550 cm^{-1} in older females (56+ age group), indicate the presence of different proteins, and age-related differences in the female parotid gland. This is in agreement with a study by Ambatipudi *et al.* who quantitatively analysed the variation and abundance of salivary proteins [70]. The study identified 338 proteins from the 20-30 years age group and 460 from the 55-65 years age group. The authors further classified each protein with the largest percentage, 40, belonging to the immune function group.

Subsequently, spectral barcodes were generated to represent the Raman bands of the highest intensity and most significance, with spectral differences compared for age and gender. The second derivative of each spectrum was taken, with absolute values above 40% of the maximum peak assigned a value of 1. The cut-off threshold of 40% was selected from empirical testing, where bars overlaid below the threshold resulted

in an overly crowded spectrum. The overcrowding is due to the noise in the derivative spectrum and the need for a higher cut-off to allow only the most significant peaks to be identified. The balance of selecting the threshold value needs to be tested for each study, accordingly [71]–[73]. Barcodes derived from the mean Raman spectra for male and female saliva across all age groups are shown in **Fig.3.3f** and the barcodes derived for the specific age groups are shown in **Fig.3.3c-d**. These are found to be in an agreement with the SOMDI spectral features of the most significant classifications shown in **Fig.3.2**.

While considering the barcoded spectra for gender classification, an overlap of the male and female spectral features can be seen with the most notable differences within the female barcodes is the appearance of more bands in the amide I and III regions as well as in the 800-860 cm^{-1} region. The high frequency of these bands correlates with the higher classification significance shown by SOMDI in **Fig.3.2e, right**. In the characteristic male barcode, the increased bands frequency found in the 900-1000 cm^{-1} region is also found in the corresponding SOMDI. Similarly, for the classification of according to the age groups, the frequency of spectral bands in **Fig.3.3, c-d** is in correspondence to the Raman shift values in SOMDI (**Fig.3.2c-d, right)**. Therefore, these barcodes can be used to identify both the gender and age of the subject according to the unique salivary spectral profile.

We have further compared the classification accuracy of the Raman spectra based on standard and barcode-based approaches using SKiNET and PCA (**Fig.3.3e**). Standard spectral classification *via* the artificial neural network SKiNET shows the highest

accuracy for both gender and age classifications. In contrast, standard Principal Component Analysis (PCA) was found to exhibit the lowest accuracy, with a classification decrease of 27% for gender, 13% for male age groups and 26% for female age groups relative to SKiNET. A more enhanced classification is found when PCA is utilised in conjunction with the barcoded spectral approach. It is possible that the performance increase is due to the PCA classifying fewer complex spectra since the barcoded spectra pre-identifies the significant spectral features. SKiNET's inherent ability to classify complex data with a higher accuracy whilst embracing and making use of the noisy data, enables it to achieve an overall classification accuracy of 93% for gender discrimination, specifically identifying males with 78% and females with 70% accuracy according to their ages.

We have further studied the temporal profiles of saliva over a period of 7-days (**Fig. 3.4a**). Saliva (males, 20-30 y) was stored at room temperature in a non-airtight microslide storage box during the study. The stability (or variability) of saliva as a function of time is an important factor in its assessment for suitability to act as a potential diagnostic biofluid and could find further uses in forensic applications, where hours or days may pass before the sample is acquired and analysed. The spectroscopic variation of the identified peaks at 630, 1003, 1051, 1458 cm^{-1} responsible for gender and age classification is shown in **Fig.3.4c-f**. The *C-C* stretching around the 1125 cm^{-1} and in the Amide III, region exhibit the largest variance across the whole-time range (**Fig.3.4b**). The largest variation as a function of time is found at 855 cm^{-1}, followed by the 1458 cm^{-1} peaks. These peaks are attributed to ring breathing mode tyrosine and CH_2 and CH_3 deformation vibrations (proteins and lipids).

Tyrosine -rich peptide statherin is a potent inhibitor of calcium phosphate precipitation. Human saliva is usually supersaturated with calcium phosphate salts that form the dental enamel. Therefore, tyrosine alongside proline rich proteins work to inhibit precipitation from calcium phosphate solutions. These inhibitors adsorb from the saliva onto surfaces such as, the tooth enamel and thus, work to recalcify the enamel and prevent the formation of mineral deposits on the tooth.

Numerous components of saliva have been studied over time under different conditions. At room temperature, analytes such as alpha amylase, cholinesterase, aspartate aminotransferase and catalase are all stable. However, lipase, creatine kinase and lactate dehydrogenase are considerably less stable after 2-3 days [74]. Salivary enzymes show the greatest degree of variation at room temperature however, the reason for these enzymatic activity changes is largely unknown and requires further study. Protein breakdown is expected to occur over the first four hours after sampling and therefore, this is where we anticipate seeing the majority of protein variation. This was observed in **Fig.3.4 c, d and f**, where the lowest saliva intensity was detected prior to stabilising over the proceeding days. Each of the three peaks are attributed to proteins. Esser *et* al. also shown that immediately after sampling, a large variation between the salivary contents of each patient's salivary profile was observed [74]. It is possible, as opposed to a direct change in the analyte under investigation over a period of time, the identified changes are due to variation within the volunteer groups saliva. Variation could also be attributed to biological variations or differing observable degrees of protein degradation between individuals. The concentration of protein/enzymes within a patient's saliva can also vary the length of time required to

cause any observable breakdown within the saliva sample thus, leading to variation between the samples. After 4 hours, most peptides decrease in abundance since during this time period they are further degraded into single amino acids.

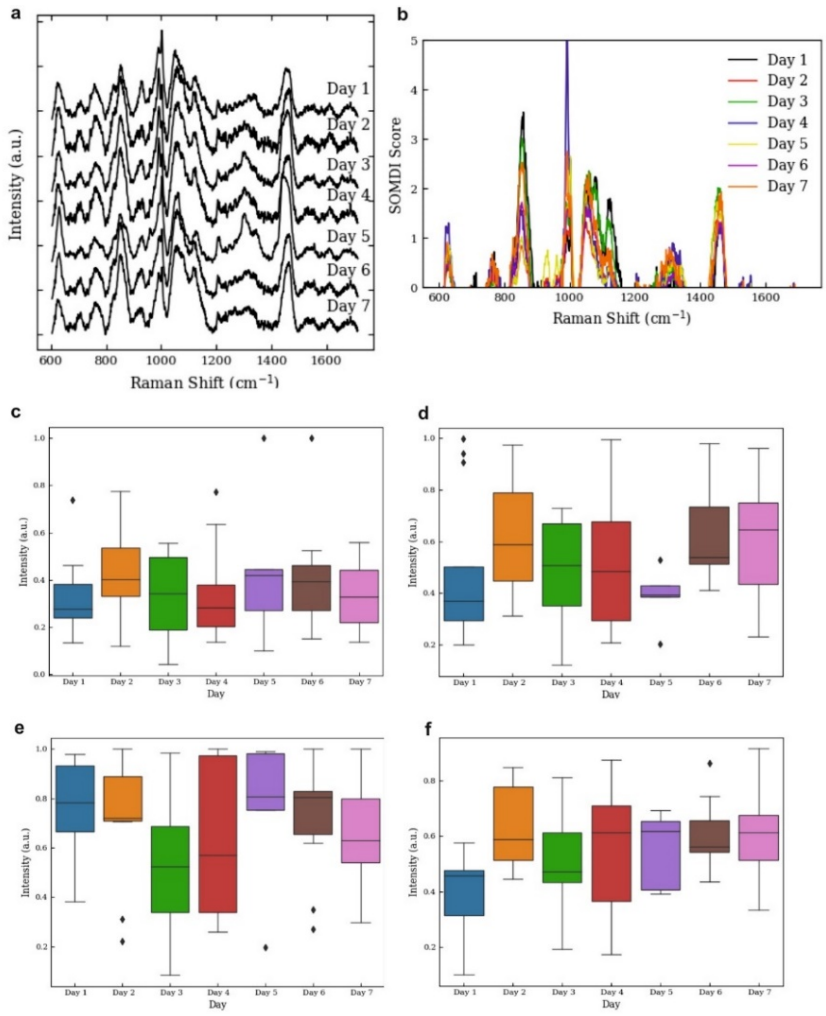

Figure 3.4. Spectroscopic time series profiles of saliva between the ages of 23-27 (*n*=10), with **(a)** mean spectra measured over 7 days and **(b)** representative Raman peaks identified through SOMDI SKiNET analysis. **(c)** Box and whisker plots represent the minima, maxima, interquartile ranges, whiskers and the median in saliva intensity at **(c)** 630 cm^{-1}, **(d)** 855 cm^{-1}, **(e)** 1003 cm^{-1} and **(f)** 1458 cm^{-1} peaks.

This could be observed on the spectroscopic time series as an initial dip in intensity before the increase when salivary stabilisation occurs. Overall, these results indicate that saliva appears to be stable over a time of 1 week with no significant changes in its

composition (p>0.05), making it a highly promising biofluid, which can maintain diagnostic significance over prolonged time. These results are in agreement with electrophoresis and mass spectroscopy observations by Chevalier *et al.*, who studied saliva over 30 days [75], and by Kim *et al.*, who used PCR analysis on saliva DNA found that unstimulated saliva remained unchanged for four weeks after initial room temperature storage. However, after 4 weeks there was a gradual reduction in its content [76].

3.4 Conclusions

In this study, Raman spectroscopy has been established as an efficient technique for profiling the molecular composition of saliva. Gender and age differentiation have been enabled from the spectral fingerprints, identifiable from the representative chronologically defined age groups of young (20-30 years), mid-aged (31-55 years) and older (>56 years) adults, establishing unique *molecular barcodes* for females and males as well as each age group. Saliva also showed no significant change in its composition with time over a period of seven days. This is of a further importance if saliva is to be employed for either diagnostics or forensic purposes and might be beneficial over other human biofluids such as blood or the cerebrospinal fluid, which are known to rapidly and significantly change their properties from the time of excretion. The classification of the molecular spectroscopic profiles has been enhanced using our SKiNET algorithm with gender identification accuracy of 93% and average age sub-classification of 75%.

These results, addressing the underpinning molecular composition of healthy saliva, by considering changes due to gender, age, and time, allow the determination of sex and age of individuals. The identified unique molecular properties of the saliva combined with the remarkable progress in developments of Raman instrumentation, which has made great strides forward in the last few decades in terms of miniaturisation of sensing devices, lays the basis for the design of novel assays for on-site forensic and diagnostic point-of-need applications. Furthermore, the non-destructive nature of Raman spectroscopy, enables preservation of evidence and the subsequent diagnostic analyses of the salivary biofluid by additional techniques. However, with the obvious advantages for forensic applications in generating a primary victim or criminal profile, for diagnostic purposes, our results highlight the importance of the care that must be taken in designing and developing diagnostic assays. For efficient and accurate diagnostic significance of disease detection, the interpretation of Raman spectral differences, either in the form of peak shifts, broadening of peaks, change in intensities or appearance/disappearance of bands, should consider the inherent spectral differences arising from the variation in gender and age of the patients. The further effects detailed interplay of the effects on saliva composition arising from smoking or BMI variation are subject to further studies. Overall, Raman based diagnosis of the various disease pathologies *via* saliva detection is emerging as a promising tool for rapid point-of-care diagnostics. With the addition of the identified molecular profiles, a simple process of a '*background-like subtraction*' of the healthy salivary barcodes will ensure the variation due to the factors such as sex, age, and further, therapeutic treatments and various underlying conditions, will be taken into the

account and reduce the error in the data interpretation of disease associated changes, leading towards improved diagnostic accuracy.

Ultimate integration of such a tool with emerging artificial intelligence techniques, such as our recently developed SkiNET method will further provide important interpretable therapeutic and management guidance. Real-time, rapid *in-vivo* spectroscopic saliva measurements in patients could also enable establishing insights into biological pathways underlying the associated diseases pathophysiology and could conceivably allow tracking the passage and dosage of pharmacological therapeutics and anaesthetics.

3.5 Materials and Methods

3.5.1 Saliva Collection and Sample Preparation

Saliva has been collected from 70 healthy consenting volunteers (32 males and 38 females) with no previous health issues or chronic conditions. Samples were collected from healthy individuals only. All individuals recruited were subject to two-tier screening by the research team at the University Hospital Birmingham. The ethical framework for the laboratory studies for analysis of saliva samples from healthy volunteers using Raman Spectroscopy to explore age related variations comprises the Red Diamond ethics (ERN_18-2050) part of the NIHR portfolio study. Tobacco and alcohol usage were noted as factors preventing participants from donating saliva. The presence of pre-established pathology or chronic pathology (such as periodontitis, diabetes or cancer) was specific exclusion, and if this was identified or suspected in the recruited individual, they were excluded, and samples were dissuaded. Participants did not

consume any food or beverages 45 minutes prior to each sample collection, only drank water for at least 60 minutes prior to collection and avoided consuming foods with strongly coloured dyes on the day of collection. All samples were collected after 11 am to avoid the cortisol wave in the early morning. The passive drool method was used to collect approximately 2 ml of unstimulated saliva (sampling time: 5 minutes) in a 50 ml conical tube (Thermo Fisher, UK). The method allowed for direct collection of saliva without need of parafilm or a citrus liquid droplet to promote stimulated salivary emission. An unstimulated saliva was used herein since the previous studies using the stimulated saliva and chewing of parafilm are known to change salivary composition [77-78]. Stimulated saliva has shown to contain lower quantities of protein and has a lower viscosity than unstimulated saliva. [79-80]. 1 ml of the collected saliva was transferred into a 1.5 ml micro-centrifuge (Thermo Fisher, UK) tube and centrifuged for 30 minutes at 10,000xg to remove the larger debris. After centrifugation, 5 µl of the collected supernatant was pipetted onto a microscope glass slide coated with aluminium foil with subsequent Raman measurements carried out immediately after. [81]

3.5.2 Raman Spectroscopy

Raman spectra were acquired *via* a Renishaw InVia Qontor confocal Raman microscope (Renishaw PLC, UK) equipped with a microscope Leica DMLM with a 50x objective, 785 nm excitation laser with an output power at the sample of 30-50 mW. Raman maps over a 100 µm x 100 µm grid were acquired for each sample with (100 spectra) using a step size of 1.5 µm between points and an acquisition time of 10 s per spectra.

3.5.3 Data Acquisition and Analysis

WiRE 5.1 (Renishaw PLC, UK) was used to acquire the data and for the polynomial (5^{th} order) background subtraction. Next the spectra were normalised using the standard normal variate (SNV) and subsequently, screened to remove cosmic ray peaks using a custom Python script (Python 3.7) [40]. The calcium spectra were identified by the presence of two dominant peaks at 620 and 1002cm^{-1}. Following this, sets of 10 spectra from different map locations were averaged to improve the SNR. Boxplots and *p*-value calculations were generated and calculated using Python. The Student's *t*-test was used to determine p-values. $p<0.05$ was considered significant. Multivariate analysis was performed using SKiNET, an open source analysis tool [41], with the accompanying Raman Toolkit web interface to build SOM models using training data and perform predictions against test data. SKiNET models were optimized by performing 10-fold cross validation on the training data, and tuning the number of neurons, initial learning rate and number of training steps. The final model used a 20 x 20 grid of neurons, 57,600 training steps (5 epochs of the data), with an initial learning rate of 0.2. The initial neighbourhood size was maintained at 2/3 the edge length of the grid and cosine similarity used as the distance metric to determine the best matching unit. The optimized model was subsequently used to classify the previously unused test data. Classification using the test data were repeated 10 times from separate SOM initializations. The number for male/female participants were 32/38, respectively. Female age groups 20-30/31-55/56+ participants were 16/11/11 and male age groups 20-30/31-55/56+ participants were 22/6/4, respectively. To achieve higher accuracy the SOM size, learning rate and number of epochs was empirically tested, with classification accuracy determined using a 10-fold cross validation. Stability of the

model was further verified by running a repeat initialisation of the classification five times. Barcoding was completed in the custom Python script, with barcodes generated using the Savitzky-Golay filter to calculate second derivative of each spectra, with a smoothing window of 21 and polynomial order of 2. The absolute values over 40% of the maximum peak height were assigned a value of 1, with values below 40% assigned a value of 0. These were overlaid on the mean spectra referencing the underpinning peak positions. Principal component analysis (PCA) was carried out in Python using the Sklearn Decomposition package, with pre-processed spectra split into a testing and training group (split of 75% training and 25% test). The model was initially trained and then tested to assess the classification accuracy with the classification repeated five times to obtain the error values.

3.6 References

[1] G. Bahar, R. Feinmesser, T. Shpitzer, A. Popovtzer, and R. M. Nagler, “Salivary analysis in oral cancer patients: DNA and protein oxidation, reactive nitrogen species, and antioxidant profile,” *Cancer*, vol. 109, no. 1, pp. 54–59, 2007.

[2] S. Sindhu and N. Jagannathan, “Saliva: A Cutting Edge in Diagnostic Procedures,” *J. Oral Dis.*, pp. 1–8, 2014.

[3] I. Pence and A. Mahadevan-Jansen, “Clinical instrumentation and applications of Raman spectroscopy,” *Chem. Soc. Rev.*, vol. 45, no. 7, pp. 1958–1979, 2016.

[4] J. M. Connolly, K. Davies, A. Kazakeviciute, A.M. Wheatley, P. Dockery, I. Keogh, M. Olivo, “Non-invasive and label-free detection of oral squamous cell carcinoma using saliva surface-enhanced Raman spectroscopy and multivariate analysis,” *Nanomedicine Nanotechnology, Biol. Med.*, vol. 12, no. 6, pp. 1593–1601, 2016.

[5] Y. Chen, S. Cheng, A. Zhang, J. Song, J. Chang, K. Wang, Y. Zhang, S. Li, H. Liu, G. Alfranca, M.A. Aslam, B. Chu, C. Wang, F. Pan, L. Ma, J.M. de la Fuente, J. Ni, D. Cui, “Salivary analysis based on surface enhanced Raman scattering sensors distinguishes early and advanced gastric cancer patients from healthy persons,” *J. Biomed. Nanotechnol.*, vol. 14, no. 10, pp. 1773–1784, 2018.

[6] F. Zapata, Á. Fernández De La Ossa, and C. García-Ruiz, “Emerging spectrometric techniques for the forensic analysis of body fluids,” *Trends Anal. Chem.*, vol. 64, pp. 53–63, 2015.

[7] C. K. Muro, K. C. Doty, L. De, S. Fernandes, and I. K. Lednev, “Forensic body

fluid identification and differentiation by Raman spectroscopy," *Forensic Chem.*, vol. 1, pp. 31–38, 2016.

[8] X. Lin, D. Lin, X. Ge, S. Qiu, S. Feng, and R. Chen, "Noninvasive detection of nasopharyngeal carcinoma based on saliva proteins using surface-enhanced Raman spectroscopy," *J. Biomed. Opt.*, vol. 22, no. 10, pp. 1–6, 2017.

[9] S. Qiu, Y. Xu, L. Huang, W. Zheng, C. Huang, S. Huang, J. Lin, D. Lin, S. Feng, R. Chen, J. Pan, "Non-invasive detection of nasopharyngeal carcinoma using saliva surface-enhanced Raman spectroscopy," *Oncol. Lett.*, vol. 11, no. 1, pp. 884–890, 2016.

[10] S. Feng, S. Huang, D. Lin, G. Chen, Y. Xu, Y. Li, Z. Huang, J. Pan, R. Chen, H. Zeng, "Surface-enhanced raman spectroscopy of saliva proteins for the noninvasive differentiation of benign and malignant breast tumors," *Int. J. Nanomedicine*, vol. 10, pp. 537–547, 2015.

[11] J. J. Zermeño-Nava, M. U. Martinez-Martinez, A. L. Ramirez-de-Avila, A. C. Hernandez-Arteaga, M. G. Garcia-Valdivieso, A. Hernandez-Cedillo, M. Jose-Yacaman, H. R. Navarro-Contreras, "Determination of sialic acid in saliva by means of surface-enhanced Raman spectroscopy as a marker in adnexal mass patients: ovarian cancer vs benign cases," *J. Ovarian Res.*, vol. 11, no. 61, pp. 1–9, 2018.

[12] P. Seredin, D. Goloshchapov, Y. Ippolitov, and P. Vongsvivut, "Pathology-specific molecular profiles of saliva in patients with multiple dental caries—potential application for predictive, preventive and personalised medical services," *EPMA J.*, vol. 9, no. 2, pp. 195–203, 2018.

[13] A. Hernández-Cedillo, G. Garcia-Valdivieso, A. C. Hernandez-Arteaga, N. Patino-Marin, A. A. Vertiz-Hernandez, M. Jose-Yacaman, H. R. Navarro-Contreras, "Determination of sialic acid levels by using surface-enhanced Raman spectroscopy in periodontitis and gingivitis," *Oral Dis.*, vol. 25, no. 6, pp. 1627–1633, 2019.

[14] V. Sikirzhytski, A. Sikirzhytskaya, and I. K. Lednev, "Multidimensional Raman spectroscopic signature of sweat and its potential application to forensic body fluid identification," *Anal. Chim. Acta*, vol. 718, pp. 78–83, 2012.

[15] K. Virkler and I. K. Lednev, "Forensic body fluid identification: The Raman spectroscopic signature of saliva," *Analyst*, vol. 135, pp. 512–517, 2010.

[16] K. Virkler and I. K. Lednev, "Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene," *Forensic Sci. Int.*, vol. 188, pp. 1–17, 2009.

[17] G. Cao, M. Chen, Y. Chen, Z. Huang, J. Lin, J. Lin, Z. Xu, S. Wu, W. Huang, G. Weng, G. Chen, "A potential method for non-invasive acute myocardial infarction detection based on saliva Raman spectroscopy and multivariate analysis," *Laser Phys. Lett.*, vol. 12, no. 12, 2015.

[18] C. Carlomagno, P. I. Banfi, A. Gualerzi, S. Picciolini, E. Volpato, M. Meloni, A. Lax, E. Colomvo, N. Ticozzi, F. Verde, V. Silani, M. Bedoni, "Human salivary Raman fingerprint as biomarker for the diagnosis of Amyotrophic Lateral Sclerosis," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, 2020.

[19] G. Eom, A. Hwang, H. Kim, S. Yang, D. K. Lee, S. Song, K. Ha, J. Jeong, J. Jung, E. Lim, T. Kang, "Diagnosis of tamiflu-resistant influenza virus in human nasal fluid and saliva using surface-enhanced raman scattering," *ACS Sensors*, vol. 4, no. 9, pp. 2282–2287, 2019.

[20] S. Hu, Y. Gao, Y. Wu, X. Guo, Y. Ying, Y. Wen, H. Yang, “Raman tracking the activity of urease in saliva for healthcare,” *Biosens. Bioelectron. J.*, vol. 129, pp. 24–28, 2019.

[21] A. Falamas, C. I. Faur, M. Baciut, H. Rotaru, M. Chirila, S. Cinta Pinzaru, M. Hedesiu, “Raman Spectroscopic Characterization of Saliva for the Discrimination of Oral Squamous Cell Carcinoma,” *Anal. Lett.*, pp. 1–13, 2020.

[22] X. Li, T. Yang, J. Lin, “Spectral analysis of human saliva for detection of lung cancer using surface-enhanced Raman spectroscopy,” *J. Biomed. Opt.*, vol. 17, no. 3, pp. 1–5, 2012.

[23] V. Sikirzhytski, A. Sikirzhytskaya, and I. K. Lednev, “Multidimensional Raman spectroscopic signatures as a tool for forensic identification of body fluid traces: A review,” *Appl. Spectrosc.*, vol. 65, no. 11, pp. 1223–1232, 2011.

[24] K. Virkler and I. K. Lednev, “Raman spectroscopy offers great potential for the nondestructive confirmatory identification of body fluids,” *Forensic Sci. Int. J.*, vol. 181, pp. 1–5, 2008.

[25] D. Nilendu, A. Kundu, A. Chand, and A. Johnson, “Forensic implications of saliva: An overview,” *Indian J. Forensic Med. Toxicol.*, vol. 14, no. 1, pp. 189–194, 2020.

[26] S. Pandit, D. Desai, P. Jeergal, and S. Venkatesh, “Awareness of forensic odontology among police personnel: A new ray of hope in forensic odontology,” *J. Forensic Dent. Sci.*,. 8, no. 1, pp. 1–6, 2016.

[27] S. Farquharson, C. Shende, A. Sengupta, H. Huang, and F. Inscore, “Rapid detection and identification of overdose drugs in saliva by surface-enhanced raman scattering using fused gold colloids,” *Pharmaceutics*, vol. 3, no. 3, pp. 425–439, 2011.

[28] V. Sikirzhytski, K. Virkler, and I. K. Lednev, “Discriminant analysis of Raman spectra for body fluid identification for forensic purposes,” *Sensors*, vol. 10, no. 4, pp. 2869–2884, 2010.

[29] C. Shende, F. Inscore, P. Maksymiuk, and S. Farquharson, “Five minute analysis of chemotherapy drugs in saliva,” *Opt. Methods Life Sci.*, vol. 6386, pp. 1–6, 2006.

[30] A. D. Gift, C. S. Shende, F. E. Inscore, and S. Farquharson, “Analysis of chemotherapy drug 5- fluorouracil and its metabolites by surface-enhanced Raman spectroscopy,” *Smart Med. Biomed. Sens. Technol. II*, vol. 5588, no. December 2004, pp. 70–77, 2004.

[31] S. Farquharson, A. D. Gift, C. Shende, P. Maksymiuk, F. E. Inscore, and J. Murran, “Detection of 5-fluorouracil in saliva using surface-enhanced Raman spectroscopy,” *Vib. Spectrosc.*, vol. 38, pp. 79–84, 2005.

[32] S. Han, L. A. Oaks, A. K. Locke, Y.-S. L. Cheng, and G. L. Coté, “Development of a free-solution SERS-based assay for point-of-care oral cancer biomarker detection using DNA-conjugated gold nanoparticles,” in *SPIE*, 2018, vol. 10501, pp. 1–6.

[33] P. Rekha, P. Aruna, E. Brindha, D. Koteeswaran, M. Baludavid, and S. Ganesan, “Near-infrared Raman spectroscopic characterization of salivary metabolites in the discrimination of normal from oral premalignant and malignant conditions,” *J. Raman Spectrosc.*, vol. 47, no. 7, pp. 763–772, 2016.

[34] R. Condò, L. Cerroni, G. Pasquantonio, M. Mancini, A. Pecora, A. Convertino, V. Mussi, A. Rinaldi, L. Maiolo, “A Deep Morphological Characterization and

Comparison of Different Dental Restorative Materials," *Biomed Res. Int.*, pp. 1–16, 2017.

[35] D. E. Romonti, A. V Gomez Sanchez, I. Milošev, I. Demetrescu, and S. Ceré, "Effect of anodization on the surface characteristics and electrochemical behaviour of zirconium in artificial saliva," *Mater. Sci. Eng. C*, vol. 62, pp. 458–466, 2016.

[36] I. J. Jahn, O. Zukovskaja, X. S. Zheng, K. Weber, T. W. Bocklitz, D. Cialla-May, J. Popp, "Surface-enhanced Raman spectroscopy and microfluidic platforms: Challenges, solutions and potential applications," *Analyst*, vol. 142, no. 7, pp. 1022–1047, 2017.

[37] C. Andreou, M. R. Hoonejani, M. R. Barmi, M. Moskovits, and C. D. Meinhart, "Rapid Detection of Drugs of Abuse in Saliva Using Surface Enhanced Raman Spectroscopy and Microfluidics," *ACS Nano*, vol. 7, no. 8, pp. 7157–7164, 2013.; (b). P. D. Gomes, J. J. Rickard, P. Goldberg Oppenheimer, " Electrofluidynamic Patterning of Tailorable Nanostructured Substrates for Surface-Enhanced Raman Scattering", *ACS Applied Nano Materials*, vol. 3, no. 7, pp. 6774-6784, 2020.

[38] C. K. Muro, L. De, S. Fernandes, and I. K. Lednev, "Sex Determination Based on Raman Spectroscopy of Saliva Traces for Forensic Purposes," *Anal. Chem.*, vol. 88, pp. 12489–12593, 2016.

[39] C. Banbury, R. Mason, I. Styles, N. Eisenstein, M. Clancy, A. Belli, A. Logan, P. G. Oppenheimer, "Development of the self optimising Kohonen Index Network (SKiNET) for Raman spectroscopy Based Detection of Anatomical eye tissue," *Sci. Rep.*, vol. 9, pp. 1–9, 2019.

[40] L. Kelleher, "Raman Spectrscopy Barcode and PCA Analysis Scripts," 2020. [Online]. Available: https://github.com/Liam-Kelleher123/Raman_code_snips.

[41] C. Banbury, "Raman Toolkit - Analysis and Data Management Tool for Raman Spectra," 2018. [Online]. Available: https://github.com/cbanbury/raman-tools.

[42] H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAinsh, N. Stone, F. L. Martin, "Using Raman spectroscopy to characterize biological materials," *Nat. Protoc.*, vol. 11, no. 4, pp. 664–687, 2016.

[43] I. Takeda, C. Stretch, P. Barnaby, K. Bhatnager, K. Rankin, H. Fu, A. Weljie, N. Jha, C. Slupsky, "Understanding the human salivary metabolome," *NMR Biomed.*, vol. 22, no. 6, pp. 577–584, 2009.

[44] I. Reddy, H. Sherlin, P. Ramani, P. Premkummar, A. Natesan, and T. Chandrasekar, "Amino acid profile of saliva from patients with oral squamous cell carcinoma using high performance liquid chromatography," *J. Oral Sci.*, vol. 54, no. 3, pp. 279–283, 2012.

[45] H. Inoue, K. Ono, W. Masuda, Y. Morimoto, T. Tanaka, M. Yokota, K. Inenaga, "Gender difference in unstimulated whole saliva flow rate and salivary gland sizes," *Arch. Oral Biol.*, vol. 51, no. 12, pp. 1055–1060, 2006.

[46] A. Cydejko, A. Kusiak, M. E. Grzybowska, B. Kochanska, J. Ochocinska, A. Maj, D. Sweietlik, "Selected physicochemical properties of saliva in menopausal women-a pilot study," *Int. J. Environ. Res. Public Health*, vol. 17, no. 7, pp. 1–8, 2020.

[47] J. N. Rukmini, R. Sachan, N. Sibi, A. Meghana, and C. Malar, "Effect of menopause on saliva and dental health," *J. Int. Soc. Prev. Community Dent.*, vol.

8, no. 6, pp. 529–533, 2018.
[48] D.R. Mahesh, ‘Evaluation of Salivary Flow Rate, pH and Buffer in Pre, Post & Post-Menopausal Women on HRT,’ *J. Clin and Diagnostic Res*., 2014, 8, pp. 233-236.
[49] National Health Service - Menopause, Last Updated August 2018, Last Accessed February 2021.
[50] C.F. Streckfus, 'Advances in Salivary Diagnostics,' *Springer-Verlag*: Berlin Heidelberg, pp. 1204, 2015.
[51] K.E. Kaczor-Urbanowicz, C.M. Carreras-Presas, K. Aro, M. Tu, F. Garcia-Godoy, D.T.W. Wong, 'Saliva diagnostics - Current reviews and directions,' vol. 242, no. 5, pp. 459-472, 2017.
[52] A. Chandra Sekhar Talari, Z. Movasaghi, S. Rehman, and I. ur Rehman, “Raman Spectroscopy of Biological Tissues,” *Raman Spectrosc. Biol. Tissues*, vol. 50, no. 1, pp. 46–111, 2015.
[53] S. Derruau, J. Robinet, V. Untereiner, O. Piot, G.D. Sockalingum, S. Lorimier, 'Vibrational Spectroscopy saliva profiling as a biometric tool for disease diagnostics: A systematic literature review,' *Molecules,* vol. 25, no. 18, pp. 4142, 2020.
[54] W. Li-Hui, L. Chuan-Quan, Y. Long, L. Ru-Liu, C. Long-Hui, C. Wei-Wen, 'Gender differences in the saliva of young healthy subjects before and after citric acid stimulation,' *Clin Chim Acta,* vol. 460, pp. 142-145, 2016.
[55] R. Leimola-Virtanen, H. Helenius, and M. Laine, “Hormone replacement therapy and some salivary antimicrobial factors in post- and perimenopausal women,” *Maturitas*, vol. 27, no. 2, pp. 145–151, 1997.
[56] M. Hernández Ríos, T. Sorsa, F. Obregón, T. Tervahartiala, M. Antonieta Valenzuela, P. Pozo, N. Dutzan, E. Lesaffre, M. Molas, J. Gamonal, "Proteolytic roles of matrix metalloproteinase (MMP)-13 during progression of chronic periodontitis: initial evidence for MMP-13/MMP-9 activation cascade," *J. Clin Periodontology,* vol. 36, no. 12, pp. 1011-1017.
[57] F. Xu, L. Laguna, and A. Sarkar, “Aging-related changes in quantity and quality of saliva: Where do we stand in our understanding?,” *J. Texture Stud.*, vol. 50, no. 1, pp. 27–35, 2019.
[58] R.M. Nagler, ‘Salivary glands and the aging process: mechanistic aspects, health-status and medicinal-efficacy monitoring,’ *Biogerontology*, vol. 5, pp. 223-233, 2004.
[59] C.K. Yeh, D. A. Johnson, M. W. J. Dodds, "Impact of aging on human salivary gland function: A community-based study," *Aging Clin Exp Res.,* vol.10, pp. 421-428, 1998.
[60] R. H. Affoo, N. Foley, R. Garrick, W. L. Siqueira, and R. E. Martin, “Meta-analysis of salivary flow rates in young and older adults,” *J. Am. Geriatr. Soc.*, vol. 63, no. 10, pp. 2142–2151, 2015S.
[61] S. Sun, F. Zhao, Q. Wang, Y. Zhong, T. Cai, P. We , “Analysis of age and gender associated n-glycoproteome in human whole saliva,” *Clin. Proteomics*, vol. 11, no. 1, pp. 1–10, 2014.
[62] H. Ben-Aryeh, M. Fisher, R. Szargel, and D. Laufer, “Composition of whole unstimulated saliva of healthy children: Changes with age,” *Arch. Oral Biol.*, vol. 35, no. 11, pp. 929–931, 1990.
[63] I. S. Yim, D. A. Granger, and J. A. Quas, “Children’s and adults’ salivary alpha-

amylase responses to a laboratory stressor and to verbal recall of the stressor," *Dev. Psychobiol.*, vol. 52, no. 6, pp. 598–602, 2010.

[64] K. Rutherfurd-Markwick, C. Starck, D. K. Dulson, and A. Ali, "Salivary diagnostic markers in males and females during rest and exercise," *J. Int. Soc. Sports Nutr.*, vol. 14, no. 1, pp. 1–8, 2017.

[65] G. H. Perry, N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Vilanae, J. L. Mountain, R. Misra, N. P. Carter, C. Lee, A. C. Stone, "Diet and the evolution of human amylase gene copy number variation," *Nat. Genet.*, vol. 39, no. 10, pp. 1256–1260, 2007.

[66] R. Lira-Junior, S. Åkerman, B. Klinge, E. A. Boström, and A. Gustafsson, "Salivary microbial profiles in relation to age, periodontal, and systemic diseases," *PLoS One*, vol. 13, no. 3, pp. 1–14, 2018.

[67] S. Tanaka, M. Machino, S. Akita, Y. Yokote, and H. Sakagami, "Changes in salivary amino acid composition during aging," *In Vivo (Brooklyn).*, vol. 24, no. 6, pp. 853–856, 2010.

[68] D. Bhuptani, S. Kumar, M. Vats, and R. Sagav, "Age and gender related changes of salivary total protein levels for forensic application," *J. Forensic Odontostomatol.*, vol. 36, no. 1, pp. 26–33, 2018.

[69] W. I. Chang, J. Y. Chang, Y. Y. Kim, G. Lee, and H. S. Kho, "MUC1 expression in the oral mucosal epithelial cells of the elderly," *Arch. Oral Biol.*, vol. 56, no. 9, pp. 885–890, 2011.

[70] K. S. Ambatipudi, B. Lu, F. K. Hagen, J. E. Melvin, and J. R. Yates, "Quantitative Analysis of Age Specific Variation in the Abundance of Human Female Parotid Salivary Proteins," *J. Protein Chem.*, vol. 8, no. 11, pp. 5093–5102, 2009.

[71] J. J. S. Rickard, V. Di-Pietro, D. J. Smith, D. J. Davies, A. Belli, and P. G. Oppenheimer, "Rapid optofluidic detection of biomarkers for traumatic brain injury via surface-enhanced Raman spectroscopy," *Nat. Biomed. Eng.*, 2020.

[72] V. Mlinar and A. Zunger, "Spectral barcoding of quantum dots: Deciphering structural motifs from the excitonic spectra," *Phys. Rev. B - Condens. Matter Mater. Phys.*, vol. 80, no. 3, pp. 1–7, 2009.

[73] I. S. Patel, W. R. Premasiri, D. T. Moir, and L. D. Ziegler, "Barcoding bacterial cells: a SERS-based methodology for pathogen identification," *J. Raman Spectrosc.*, vol. 39, no. 11, pp. 1660–1672, 2008.

[74] D. Esser, G. Alvarez-Llamas, M.P. de Vries, D. Weening, R.J. Vonk, H. Roelofsen, 'Sample stability and protein composition of saliva: Implications for its use as a diagnostic fluid,' *Biomarker Insights,* vol. 3, pp. 25-27, 2008.

[75] F. Chevalier, C. Hirtz, S. Chay, F. Cuisinier, N. Sommerer, M. Rossignol, D. Deville de Périère, "Proteomic studies of saliva: A proposal for a standardized handling of clinical samples," *Clin. Proteomics*, vol. 3, no. 1–4, pp. 13–21, 2007.

[76] Y. Kim, Y. Kim, and D. Ph, "The Effects of Storage of Human Saliva on DNA Isolation and Stability," vol. 31, no. 1, pp. 1–16, 2006.

[77] E. Neyraud, T. Sayd, M. Morzel, E. Dransfield, 'Proteomic analysis of human whole and parotid salivas following stimulation by different tastes,' *J. Proteome Res.,* vol. 5, pp. 2474-2480, 2006.

[78] R. Schipper, A. Loof, J. de Groot, L. Harthoom, E. Dransfield, W. van Heerde, 'SELDI TOF-MS of saliva: Methodology and pre-treatment effects, *J. Chromatogr B,* vol. 15, pp. 45-53, 2007.

[79] M. Edgar, C. Dawes, D. O'Mullane, 'Saliva and Oral Health, 3rd Ed, London,

British Dental Association, pp. 6-7, 2004.

[80] S. Gomar-Vercher, A. Simón-Soro, J.M. Montiel-Company, J.M. Almerich-Silla, A. Mira, 'Stimulated and unstimulated saliva samples have significantly different bacterial profiles,' *PLoS One,* vol. 13, pp. 1-12, 2018.

[81] I. Maitra, C.L.M. Morais, K.M.G.Lima, K.M. Ashton, R.S. Date, F.L. Martin, 'Raman spectral discrimination in human liquid biopsies of oesophageal transformation to adenocarcinoma,' *J. Biophotonics,* vol. 13, no. 3, 2019.

CHAPTER 4

RAMAN SPECTROSCOPY AIDED DIAGNOSIS OF INFLAMMATORY BOWEL DISEASE *VIA* ASSESSING SALIVA AND ENDOSCOPIC TISSUE BIOPSY: A PILOT STUDY

This results chapter is based on a paper currently in submission;

1. **Buchan E.,** Majumder S., Lorenzo Parigi T., Nardone O.M., Abdawn Z., Hejmadi R., Bossa C., Panton M., Shivaji U., Iacucci M., Goldberg Oppenheimer P. [2023], 'Raman spectroscopy aided diagnosis of inflammatory bowel disease *via* assessing saliva and endoscopic tissue biopsy.'

Author contributions – **E Buchan**: Conceptualisation, Investigation, Methodology, Formal analysis, Writing-original draft, review and editing. **S Majumder**: Writing draft, review and edits. **T Parigi**: Patient recruitment, Writing- review and editing. **Z Abdawn**: Performed/compiled sample histology. **R Hajmadi**: Performed/compiled sample histology, writing-review and editing. **C Bossa**: Patient recruitment**. M Panton**: Patient recruitment. US: Performed/compiled sample histology, writing-review and editing. **M Iacucci**: Conceptualisation, patient recruitment, writing-original draft, review and editing, resources. **P Goldberg Oppenheimer**: Conceptualisation, Methodology, resources, writing-original draft, review and editing, funding acquisition, supervision.

All authors reviewed and approved the final version of the manuscript.

4.1 Abstract

Colonoscopy, the gold standard for diagnosing and monitoring inflammatory bowel disease, is a costly and invasive procedure often poorly tolerated by patients. In recent years, saliva analysis has gained interest as a potential non-invasive source of disease biomarkers due to its correlation with blood constituents at a molecular level. We aimed to investigate the diagnostic performance of Raman spectroscopic analysis on saliva samples compared to colonic tissues in patients with ulcerative colitis (UC), Crohn's disease (CD) and healthy controls. The ultimate goal is to identify biomarkers or distinctive patterns which would help diagnose and monitor IBDs by analysing the differences in Raman spectral signatures. Consecutive patients with IBDs were recruited and underwent colonoscopy. Tissue biopsies were graded according to their histology *via* a selection of PICaSSO Histologic Remission Index (PHRI), Nancy Index Score (NIS), Robarts histopathology index (RHI), Mayo Endoscopic Score (MES), and Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Via* a combination of Raman spectroscopy and a new artificial neural network algorithm (SKiNET), a non-destructive molecular profiling tool enabling the assessment of both salivary and colonic biopsy tissues was developed. Artificial neural network computational tool, SKiNET, was used to identify spectral markers and discriminate healthy and IBD classes. This pilot study characterised the spectral signatures of 101 saliva samples (healthy = 61, UC = 22 and CD =18) and 46 tissue biopsy samples (healthy = 14, UC = 17 and CD = 15). Samples were correlated with tissue pathology markers and endoscopic evaluation. SKiNET yielded 87.5% sensitivity and 89.0% specificity when classifying saliva and 83.9% sensitivity and 84.6% specificity when classifying tissue biopsy. When comparing the spectral fingerprints of saliva to tissue biopsy, the UC spectra exhibit

similar Raman signatures with comparable intensities at the primary peaks of 1003 cm^{-1}, 1445 cm^{-1} and 1656 cm^{-1}. In contrast, CD saliva and tissue biopsy indicate significant variance throughout the Raman spectral fingerprint region with intensity changes observed at each of the primary peaks and additional shoulder peaks. Raman spectroscopic analysis of saliva is a reliable, non-invasive tool for diagnosing and differentiating UC from CD. This unique combination of Raman spectroscopy and advanced machine learning represents significant progress towards improved, non-invasive, and rapid classification of IBD. It lays the platform for various applications in disease diagnosis and monitoring.

4.2 Introduction

Inflammatory bowel disease (IBD) is a group of chronic inflammatory conditions that includes ulcerative colitis (UC) and Crohn's disease (CD) (1). Although both are known to cause inflammation in the digestive system, UC affects the large intestine whilst CD can affect any part of the digestive tract from the mouth to the rectum [1]. Colonoscopy remains the most important investigation for diagnosing and monitoring IBD despite being an invasive and costly procedure, often poorly tolerated by patients. Therefore, accurate and non-invasive alternatives for IBD diagnosis and monitoring are needed to reduce the burden of invasive testing [2].

Thus far, the only biomarkers used in clinical practice are circulating C-reactive protein (CRP), and faecal calprotectin (FC). However, the utility of CRP is mostly limited to moderate or severe settings, as well as having a low sensitivity for mild localised diseases. In contrast, FC requires the collection of stool samples, which can be

impractical and often perceived as unpleasant. Moreover, no biomarker can differentiate IBD from other causes of intestinal inflammation, *e.g.,* UC from CD.

More recently, saliva has gained interest as a readily available biofluid, which can be easily collected in any setting without needles or invasive procedures [3]. Saliva contains a host of hormones, antibodies, enzymes and growth factors, many of which enter from the blood *via* transcellular or paracellular routes, meaning that many constituents are also present in blood. Therefore, saliva can reflect an individual's physiological state [4]–[7]. Although saliva has potential diagnostic properties, a challenge with this biofluid is the relatively low concentration of biomarkers compared with the levels typically found in blood. Nevertheless, new, more sensitive and rapid methodologies such as, Raman spectroscopy (RS), have emerged as a useful tool to overcome these issues [8]–[13].

RS is a novel method used to study the vibrational modes of molecules by analysing the light scattered from a biological sample [14]. RS has been used to analyse tissues and cells in *real-time*, without additional staining or labelling, successfully distinguishing between healthy and diseased samples based on spectral fingerprints. Of a particular benefit for clinical settings, RS is capable of revealing altered biochemistry within an individual’s body, specific to a particular disease, which precedes macroscopic tissue changes [15-17]. It has, thus, proved to be a versatile tool in clinical diagnostics, aiding in the detection of a variety of diseases ranging from cancer [18] through infectious [19], [20] to neurodegenerative diseases [12].

In the current study, we investigated the potential of Raman spectroscopy to reflect salivary changes consistent with IBD, establishing if it can be a useful diagnostic adjunct. We further examined and systematically evaluated Raman spectral fingerprints of saliva and endoscopic tissue biopsies from patients diagnosed with IBD. Finally, we have evaluated the spectral variability between CD and UC types and determined the subsequent ability to discriminate between the disease subtypes.

4.3 Results

Saliva was collected from 101 participants (22 UC, 18 CD and 61 healthy controls), and colon tissue biopsies were collected from 13 IBD patients (UC 8, CD 5) and 6 healthy controls.

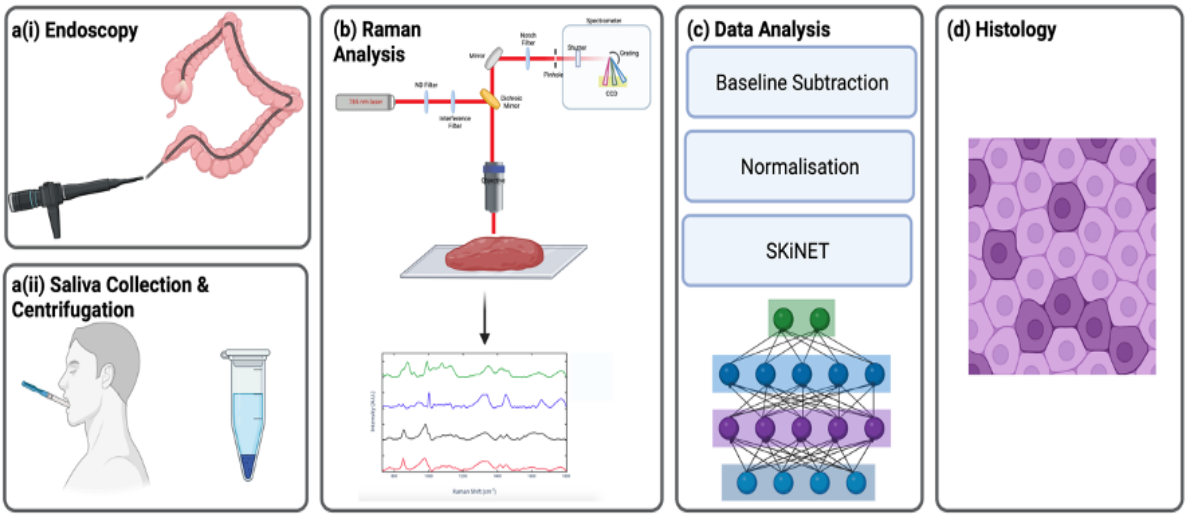

Figure 4.1. Experimental roadmap indicating **(a)** endoscopic biopsy collection **(i)**, saliva collection and centrifugation **(ii)**, Raman analysis **(b)** indicating spectral fingerprints of healthy and diseased sample types, data analysis **(c)** including baseline subtraction, normalisation and SKiNET classification and **(d)** tissue biopsy histology correlation. (Figure created in Biorendor.com)

Patients' demographic, endoscopic and histological characteristics are summarized in **Table 4.1**. An overview of the experimental roadmap is presented in **Fig. 4.1,** showing the established process including saliva collection, endoscopy for tissue biopsy retrieval, Raman spectral analysis, histological tissue examination and artificial neural network algorithm-based data classification.

4.3.1 Raman Spectroscopy of Saliva

Fig.4.2 demonstrates the Raman spectral fingerprints of saliva from UC (n=22), CD (n=18) and healthy patients (n=61). A representative mean spectrum in the fingerprint spectral region of 700-1700 cm^{-1} for healthy and IBD saliva is shown in **Fig.4.2a**, with further sub-classification of UC and CD in **Fig.4.2b** with the derived weight vectors characteristic for each class in **Fig.4.2c-h.**

Each sample was analysed *via* mapping of 100 spectra, providing a total of 10,100 spectra used to train the SKiNET model, and out of these, 20% were set aside for testing. In the testing phase, SKiNET distinguished healthy from IBD patients with a sensitivity of 87.5 (86.1 – 88.9) %, specificity of 89 (86.9 – 91.1) %, positive predictive value (PPV) of 89 (87.3 – 90.7) %, negative predictive value (NPV) of 87.3 (84.9 – 89.6) % and an accuracy of 89.9 (87.6 - 92.2) %. Performance did not vary significantly when restricting the analysis to only one type of IBD, UC or CD *vs*. controls, indicating the reliability of this method for both conditions (**Table 4.2**).

Table 4.1. Overview of patient demographics of IBD patients.

Characteristic	Total (n=33)
Age (median, std, yrs)	43.7 ± 14.9
Sex	
Female	na
Male	na
Disease duration (years)	72.6 ± 8.8
UC patients	16 (48.5%)
UC Montreal disease extension	
Proctosigmoiditis	7 (43.7%)
Left colitis	1 (6.25%)
Pancolitis	4 (25%)
CD patients	17(51.5%)
CD Montreal disease localisation	
Ileal	5 (29.4%)
Colonic	2 (11.8%)
Ileocolonic	8 (47%)
CD Montreal disease behaviour	
Inflammatory	na
Stricturing	na
Penetrating	na
Perianal disease	2 (11.8%)
CDAI	na
pMAYO	na
SES-CD > 2	na
Rutgeerts score ≥ 2	na
MES >1	5 (8 na)
MES ≤1	3
UCEIS ≤1	
UCEIS > 1	2
PICASSO	6 (8 na)
Baseline anti-TNF-α therapy	6(18.1%)
Baseline AZA therapy	4(12.1%)
Baseline Vedolizumab	3(9.09%)
Baseline ustekinumab	5 (15.1%)
Baseline tofacitinib	4(12.1%)
Baseline mesalamine therapy	5 (15.1%)
Baseline risankizumab	1 (3.03%)
CRP (mg/L)	10.3 ± 9.8
FC (µg/g)	298 ± 8.9

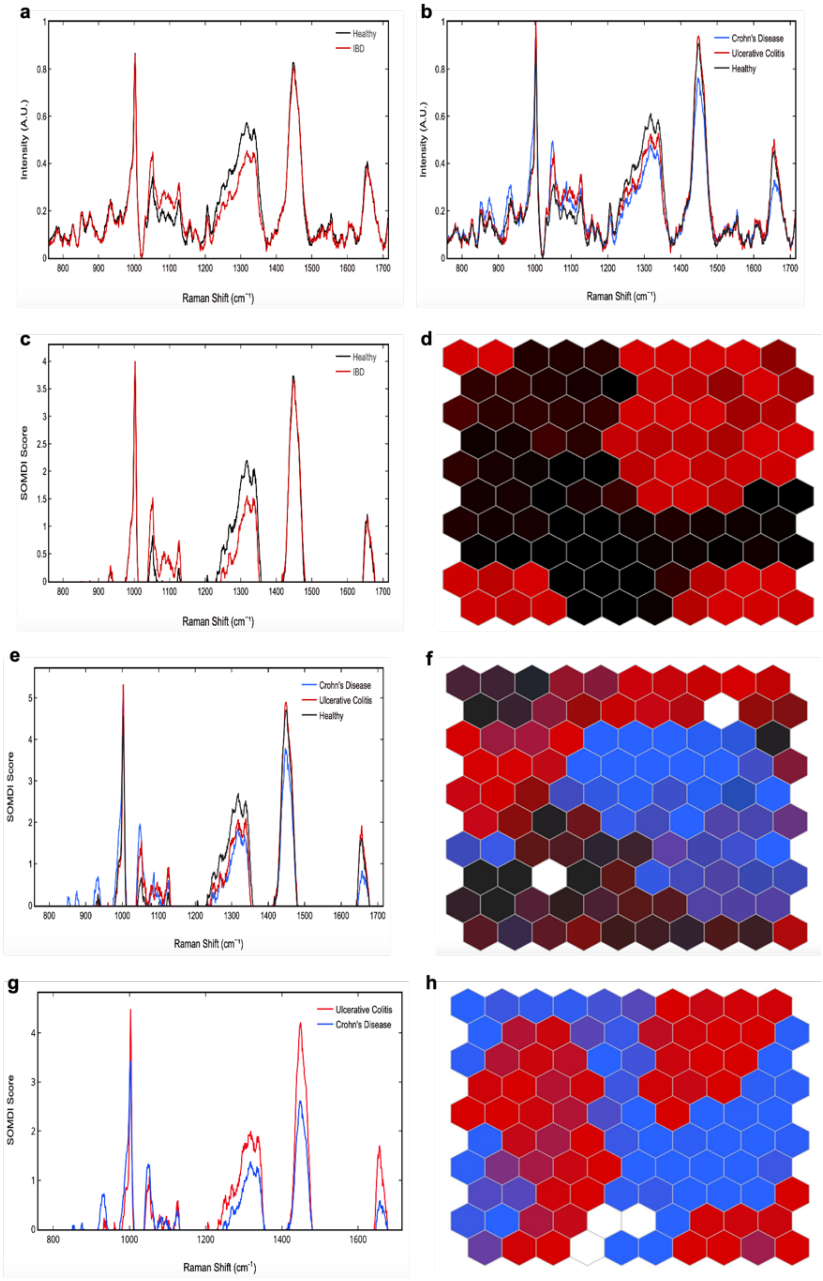

Figure 4.2. Representative mean spectra comparing **(a)** healthy and IBD participants and **(b)** healthy, CD and UC participants. SOMDI (left) and SOM (right) of **(c)** healthy and IBD, **(e)** healthy, UC and CD and **(g)** CD and UC saliva. The classification according to diseased state is represented by black hexagons for healthy, blue hexagons for CD and red hexagons for UC.

Table 4.2. Classification performance of saliva samples.

Comparison	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	Accuracy (%) (95% CI)	*p*-values
Healthy *vs*. IBD	87.5 (86.1 – 88.9)	89 (86.9 – 91.1)	89 (87.3 – 90.7)	87.3 (84.9 – 89.6)	89.9 (87.6 - 92.2)	<0.0001
Healthy *vs*. UC	91.6 (88.9-94.3)	88.1 (85.8 – 90.4)	88.5 (86.5 – 90.5)	88.5 (86.6 - 90.4)	80.1 (78.4 - 81.8)	<0.001
Healthy *vs*. CD	87.3 (84.3 – 90.3)	87.7 (85.7 – 89.7)	86.1 (83.3 – 88.8)	88.9 (87.2 – 90.7)	80.3 (79.1 - 81.5)	<0.001

Further subclassification identifying the accuracy of UC *versus* the CD and healthy controls and CD *versus* the UC and healthy controls was performed (**Table 4.3**), exhibiting overall high sensitivity and specificity in discriminating UC from CD, achieving 90.7% and 87.8%, respectively.

Table 4.3. Classification performance of ulcerative colitis and Crohn's disease sub-classification.

Comparison	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	Accuracy (%) (95% CI)	*p*-values
UC *vs*. All (CD + HC)	88 (84.8 - 91.2)	89.4 (86.9 – 91.9)	89.6 (87.7 - 91.5)	87.8 (86.1 – 89.6)	88.7 (87.2 – 90.2)	<0.0001
CD *vs*. All (UC + HC)	89.5 (87.2 – 91.2)	85.3 (82.3 – 88.3)	86.7 (85.1 – 88.6)	88.3 (86.3 – 90.3)	86.5 (85.5 – 87.5)	<0.0001
UC *vs*. CD	90.7 (88.5 – 92.9)	87.8 (86.0 – 89.6)	87.4 (85.5 – 89.3)	91.1 (88.8 – 93.4)	89.2 (87.7 – 90.7)	<0.0001

The spectral signature for both the healthy and IBD classes appears similar with primary peaks observable at Raman shifts of 853, 879, 932, 1003, 1051, 1077, 1086, 1127, 1316, 1445 and 1656 cm^{-1}. The most striking difference between the three saliva subgroups is in the intensity of the peaks at 1003, 1051 and 1445 cm^{-1}. These are found to be considerably higher in patients diagnosed with UC than in both CD and healthy cohorts. In terms of the biomolecular composition, all three spectra exhibit vibrational contributions from proteins, lipids, nucleic acids, amino acids as well as carotenoids, hypoxanthine, myoglobin and haem. The detailed list of detected peaks and the corresponding Raman assignment is summarised in **Table 4.4**.

Table 4.4. Raman peak assignments of healthy, IBD saliva and tissue biopsy.

Peak (cm^{-1})	Assignment	Reference
759	Tryptophan -ring breathing vibration	[28]–[30]
830-853	Tyrosine doublet (protein)	[29], [31]
873	($\rho(CH_2)$ in proteins)	[32]
875, 879	Phosphatidylcholine	[31]
936	Protein Backbone	[30]
1003	Phenylalanine, carotenoids	[28], [29], [31], [33], [34]
1045	Proline	[35], [36]
1124	Lipids	[36]
1160	(β-carotene)	[32]
1153	V(C-N), β-carotene, terpene	[29], [36]
1125	Myoglobin (haem core), phospholipids, proteins	[34]
1307	Phospholipids, lipids, adenine, myoglobin (haem core)	[34]
1342	Protein	[31]
1352	Nucleic acids	[37]
1368	Tryptophan, guanine, thymine, myoglobin (haem core)	[34]
1372	Lipid	[32]
1445	Lipid and protein	[31]
1450	CH_2-deformation	[33]
1468	$\delta(CH_2)$	[29]
1472	CH2 bending of lipids and proteins	[37]
1656	Lipid/Amide I	[30], [31]

1657	C=C stretching in lipids	[28]
1658	(v(C = O)), Amide I and lipids	[32]

The spectra in **Fig.4.2a-b** depict typical molecular characteristics of biological materials. For instance, the primary band at 1003 cm^{-1} is attributed to phenylalanine, 1445 cm^{-1} is the CH_2 deformation, and 1620-1680 cm^{-1} is associated with the Amide I band. Assignments for the peaks which contribute significantly to class discrimination, not limited to prominent features, include 853 cm^{-1} (tyrosine), 879 cm^{-1} (phosphatidylcholine), 932 cm^{-1} (protein backbone), 1051 cm^{-1} (Proline), 1077 cm^{-1} (DNA and phospholipids), 1095 cm^{-1} (Hypoxanthine), 1127 cm^{-1} (Myoglobin) and 1445 cm^{-1} (Proteins and lipids). Comparing UC to CD (**Fig.4.2g**), reveals more intense peaks in CD samples, related to protein backbone, tricaprylin and myoglobin whereas the UC samples exhibit stronger peaks of phenylalanine, carotenoids, Amide II, CH_2 deformation of proteins and lipids and Amide I.

4.3.2 Raman Spectroscopy of Colon Tissues

Raman analysis of 44 tissue biopsies collected from IBD patients (17 UC and 15 CD) and healthy controls (n=14) was subsequently performed. 4,400 spectra were used for characterisation and disease state classification. A comparison between the mean spectra obtained from participants within each disease category is shown in **Figs.4.3a-b** with subsequent classification *via* SKiNET (**Fig.4.3c-h**).

Similarly, to the saliva data set, results discriminating between healthy and diseased biopsy tissue types were assessed as well as those comparing UC to CD (**Table 4.5**).

Test sensitivity, specificity and accuracy were found to be high for all comparisons drawn.

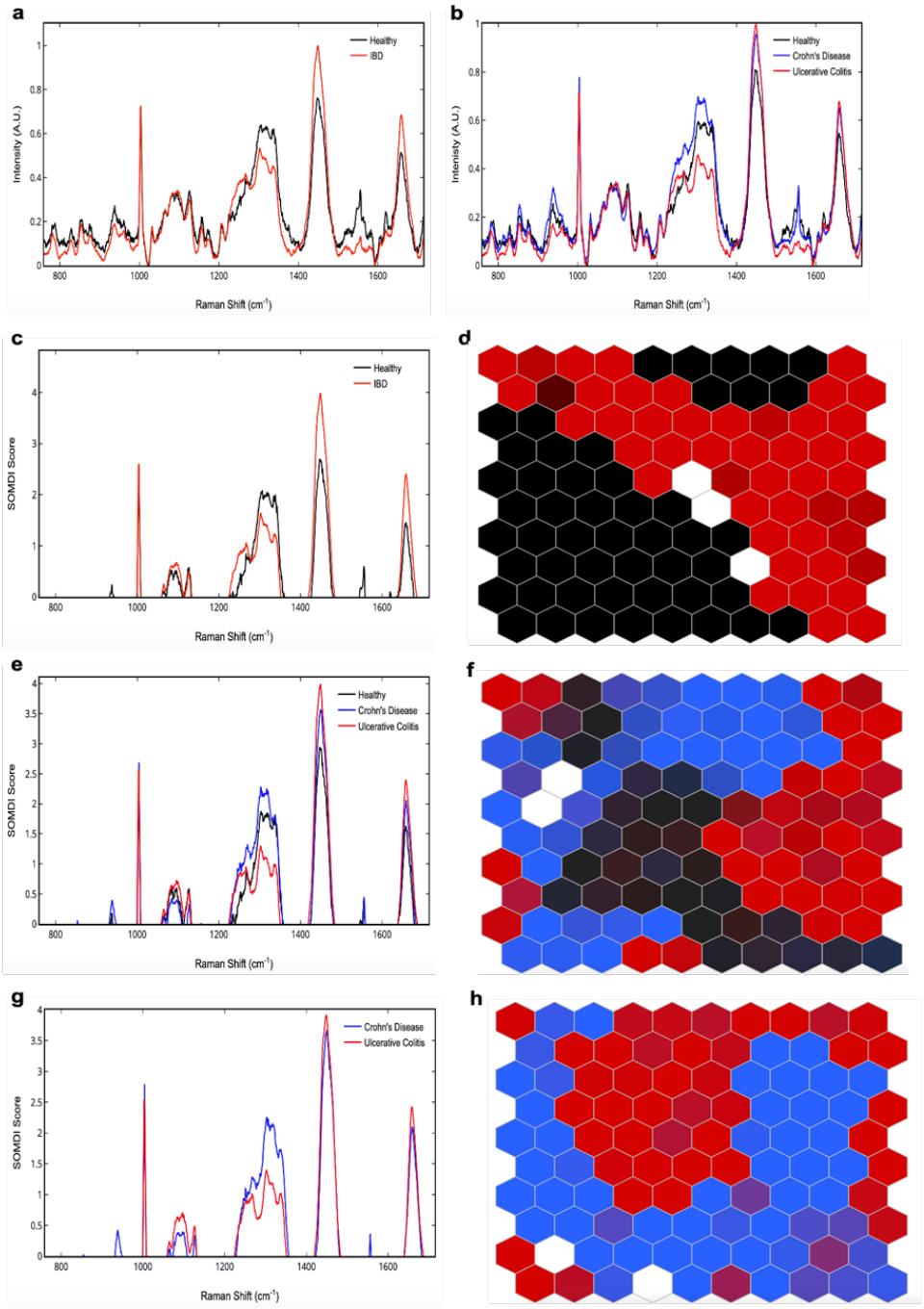

Figure 4.3. Representative mean spectra comparing **(a)** healthy and IBD participants and **(b)** healthy, CD and UC participants. SOMDI (left) and SOM (right) of **(c)** healthy and IBD, (**e)** healthy, UC and CD

and **(g)** CD and UC tissue biopsy. The classification according to disease state are represented by black hexagons for healthy, blue hexagons for CD and red hexagons for UC.

Table 4.5. Classification performance of biopsy tissue comparisons.

Comparison	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%) (95% CI)	NPV (%) (95% CI)	Accuracy (%) (95% CI)	*p*-values
Healthy v. IBD	83.9	84.6	87.6	88.2	82.5 (81.1 - 83.9)	<0.001
Healthy v. UC	94.7	90.4	90.0	95.0	90.0 (89 - 91)	<0.0001
Healthy v. CD	94.4 (92.4 – 96.4)	86.3 (84.8 – 87.8)	85 (82.7 – 87.3)	95 (92.8 – 97.2)	90.2 (89.4 - 90.8)	<0.0001
UC v. CD	91.2 (90.0 – 92.5)	91.9 (90.2 – 93.7)	91.9 (89.6 - 94.3)	91.1 (88.5 – 93.7)	97.5 (96.5 - 98.5)	<0.0001

The most prominent spectral features of tissue biopsy for the three groups are found at 780, 828, 853, 877, 936, 1003, 1110, 1127, 1305, 1339, 1445, 1548 and 1657 cm^{-1} with the greatest intensity changes arising from 936, 1003, 1305 and 1657 cm^{-1}. The peaks at 1003 cm^{-1} (phenylalanine), 1305 cm^{-1} (CH_2-twisting), 1445 cm^{-1} (CH_2-deformation) and 1657 cm^{-1} (Amide I) are typically used to characterise the protein-lipid composition of biological tissues. Other prominent features of the three tissue biopsy groups include the 780 cm^{-1} (Tryptophan), 828 cm^{-1} (DNA Phosphate group), 877 cm^{-1} (Phosphatidylcholine), 936 cm^{-1} (Protein backbone), 1110 cm^{-1} (C-C vibration of gauche bonded chain) 1127 cm^{-1} (Myoglobin), 1298 cm^{-1} (CH_2 bending of proteins and lipids) 1339 cm^{-1} (Proteins) and 1548 cm^{-1} (Amide II). The CD saliva samples (**Fig. 4.3g**) exhibit more intense peaks associated with the CH_2 bending of lipids and proteins, a stronger protein backbone peak and more prominent bands in the Amide II region as well as the phenylalanine. In contrast, the UC saliva samples exhibit stronger

intensities representative of the C-C vibration of gauche-bonded chain, of the CH_2-deformation of proteins and lipids and in the Amide I region.

4.3.3 Comparison of IBD Saliva Samples and Tissue Biopsies

The average Raman spectra comparing UC and CD saliva and tissue biopsy with the corresponding SOM clustering are shown in **Figure 4.4**. Raman analysis of 32 tissue biopsy samples (17 UC and 15 CD) and 40 saliva samples (22 UC and 18 CD) was performed, yielding a total of 7,200 spectra for characterisation and SKiNET classification.

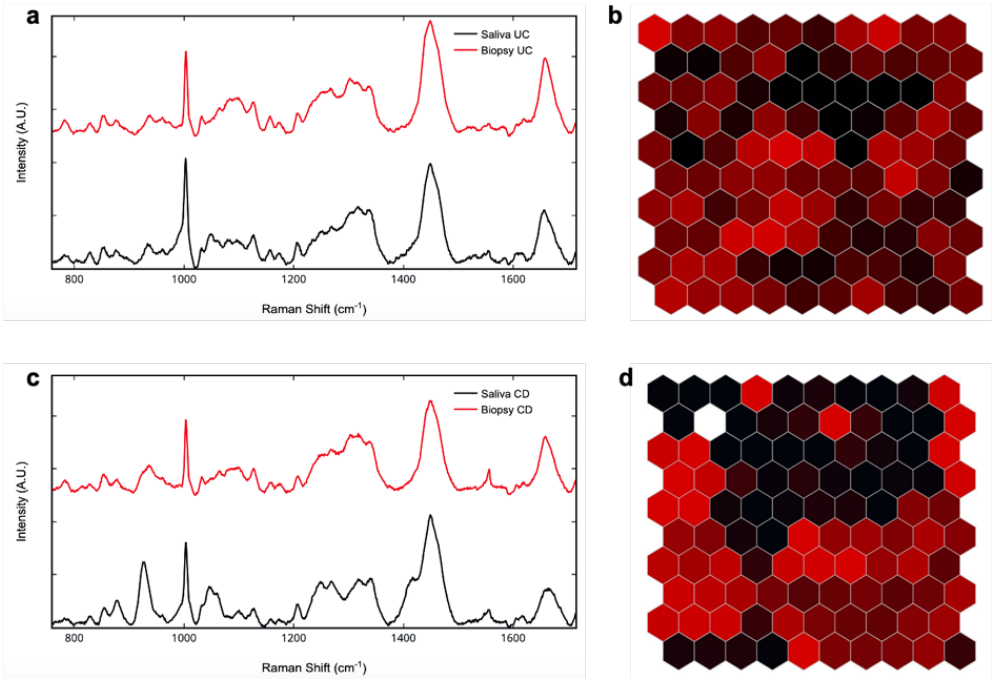

Figure 4.4. Average baseline-subtracted Raman spectra comparing **(a)** UC saliva to UC biopsy tissue and **(c)** CD saliva to CD biopsy tissue. Corresponding SOMs of **(b)** UC and **(d)** CD. Red hexagons indicate biopsy tissue and black hexagons indicate saliva.

The mean UC spectra in the fingerprint region depicted in **Fig.4.4a** is of a similar spectral fingerprint, with comparable intensities at each primary peak. This is further supported *via* the SOM **(Fig.4.4b**), exhibiting an overlap between the two spectra. Strong peaks are identified in both sample types at 1003 cm^{-1} (Phenylalanine), 1445 cm^{-1} (Lipids) and 1656 cm^{-1} (Amide I).

In contrast, CD saliva and tissue biopsy spectra (**Fig.4.4c**) indicate significant variance throughout the fingerprint region. Significant intensity changes ($p^{***}<0.0001$) are observed at shifts of 940 cm^{-1} (protein backbone), 1080 cm^{-1} (phospholipids), 1549 cm^{-1} (deoxy-myoglobin) and 1656 cm^{-1} (Amide I). Additional shoulder peaks are observed in CD saliva at 1045 cm^{-1} (proline) and 1412 cm^{-1} (plasmatic lipids). **Fig.4.4d** demonstrates the difference between the sample types with the clustering of the two groups visible in SOM. Intensity changes observed at Raman bands of 940 and 1125 cm^{-1} are most likely attributed to higher levels of acute phase proteins, such as CRP, circulating in both blood and saliva. Indeed, amongst participants within the study, CRP levels were significantly higher in CD patients than those with UC (12.6 mg/L *vs*. 3.5 mg/L)**.**

4.4 Discussion

In recent years saliva has gained a lot of attention as a non-invasive, easy and readily accessible diagnostic biofluid. This pilot study is the first to investigate the potential role of saliva in the diagnosis of IBD *via* Raman spectroscopic analysis, successfully validating and confirming that RS can accurately detect molecular fingerprints associated with IBD and further distinguish IBD from healthy controls. Building upon

the previous preliminary studies by Bielekci *et* al., Bi *et* al. and Smith *et* al., which indicated that RS could differentiate between inflamed and healthy colon tissues and CD and UC [4], [28], [38], the adopted approach is yet to overcome the obstacle of endoscopy and indeed, an urgent need in IBD - the availability of reliable, non-invasive diagnostic tools, remains unmet.

Herein, we have developed a protocol and utilised an advanced spectroscopic analysis of saliva, an often-overlooked body fluid whose composition is known to reflect the overall health state. *Via* RS, an emerging technique, which permits the detection of biochemicals down to *ultra-low* concentrations, combined with advanced AI, we have overcome the main limitations associated with the use of saliva, resolving peaks of significance within the largely overlapping spectra from saliva and colon. This indicates the ability of saliva to act as a diagnostic biofluid, paving the way for the broader exploitability of this biofluid in the wider medical fields in various clinical applications. The study further highlights the potential of Raman spectroscopic analysis in conjunction with advanced machine learning algorithms for the rapid and precise differentiation between individuals with IBD and those healthy control groups. By combining Raman spectroscopy and machine learning techniques, we have achieved a higher diagnostic performance than current biomarkers. This enables swift and accurate discrimination of IBD patients from healthy individuals. Noteworthy, this holds significant promise for medical diagnostics, particularly due to the convenience and accessibility of saliva as a sample source. Furthermore, this hybrid approach can also distinguish between subtypes of IBD with high accuracy, unlike other non-invasive tools. When comparing RS with other non-invasive biochemical tests for IBD, it is

important to note that this spectroscopic technique provides an inherent overview of the biochemical composition of the whole tissue as opposed to detecting individual molecules and thus, allowing for a more comprehensive assessment of the studied matter.

We demonstrate that Raman spectral analysis, although presenting definite similarities between UC, CD and healthy controls, allows the successful differentiation of patients according to these groups with high accuracy. Specifically, several statistically significant peaks identified from Raman spectra indicate the abundance of certain biochemical bonds, allowing the subsequent characterisation of the molecular composition of each sample. A specific Raman spectrum reflects biochemical changes attributed to a particular disease, which may precede any identifiable macroscopic tissue changes. For instance, the increased intensity of the peak at 780 cm^{-1} in healthy control biopsies is attributed to tryptophan. Tryptophan is negatively associated with inflammation and immune response, and its serum levels are significantly lower in patients with IBD than in healthy controls, with a stronger reduction in CD patients [39], corresponding with the decrease observed in IBD patients in our study. Moreover, gross comparison of Raman spectra revealed substantial differences between spectra from healthy, UC and CD patients, with the greatest intensity change observed at 1003 cm^{-1} associated with the phenylalanine, which was detected at a greater abundance in both UC and CD saliva samples relative to the healthy counterparts. Increased levels of phenylalanine have previously been associated with immune activation and inflammation and therefore, in the diseased state, there is a greater infiltration of inflammatory cells, leading to increased phenylalanine levels (**Fig. 4.2b**) [40], [41].

We recognise the potential limitations of this study. While the connection between specific chemical bonds and their corresponding spectra is well understood, interpreting the biological implications of changes in the abundance or reduction of certain molecules can be more complex and may only sometimes have a straightforward interpretation. Yet, the differences in molecular abundance as detected *via* RS can allow for the identification of new disease biomarkers. Most of the relevant peaks detected in this study correspond to the previously established spectroscopic changes [5], [28], reassuring the reproducibility of this technique. However, it is worth noting that this is a pilot exploratory study and further larger clinical trials will be needed to validate the results with larger cohorts. Furthermore, the quantitative association between RS and inflammation is yet to be established. At present, our methodology differentiates dichotomous disease from healthy cohorts, however, it is reasonable to speculate that with further analysis, this will allow the grading of severity of inflammation from salivary spectra, expanding the potential use of saliva not only as a diagnostic biomarker but also, as one capable of monitoring disease activity as well as patient response to current and newly developed therapeutics.

Overall, remarkably, RS rapidly and accurately captures intensity changes at specific spectral bands linked to variations in acute phase protein levels, particularly the CRP. These changes are indicative of inflammatory responses, such as those seen in IBD patients. Hence, the link between Raman spectral changes and biomolecular alterations paves the way for grading the disease severity. Long-term applications of saliva as a Raman diagnostic tool include early disease detection and intervention, the

ability to monitor disease progression and patient response to treatment and integrating the methodology into point-of-care diagnostics.

4.5 Conclusions

New technologies capable of detecting biomolecular changes in a timely manner as well as discriminating between diseased states of individuals, are essential for independent disease identification and non-invasive, biopsy-free monitoring. Herein, Raman spectroscopy has been established as an efficient technique for diagnosing IBD, including both UC and CD. The diseased state of the patients as well as sub-IBD differentiation have been discerned from Raman spectral fingerprints combined with advanced SKiNET classification from clinical samples. Our technique successfully discriminated between UC and CD subtypes of IBD with high accuracy in saliva and tissue biopsies. Overall, Raman-based diagnosis of IBD and its associated pathologies *via* saliva detection is emerging as a promising tool for rapid, non-invasive disease diagnostics. Future research will implement a Raman-based methodology for timely, non-invasive diagnosis of IBD combined with the ability to grade the disease severity and monitor patient response to therapeutics.

4.6 Methods

4.6.1 Patients

Patients with an established diagnosis of UC or CD (aged ≥ 18 years), irrespective of diseased state or treatment history, were recruited from the IBD clinic of Queen Elizabeth Hospital Birmingham, UK. Healthy controls, those with no known concomitant or previous health issues, aged 18 ≥ years, were also recruited. The study

was approved by Ethics Committee (RRK6508) with written consent obtained from all participants.

4.6.2 Saliva Collection and Preparation

All participants were instructed not to consume any food or drink 45 minutes before sample collection, to only drink water at least 1 hour prior to collection and to avoid eating any foods on the day of collection that contain strong dyes. Using the passive drool method, approximately 5 mL of unstimulated saliva was collected from each participant in a 50 mL Falcon tube (Thermo Fisher, UK). 1 mL of collected saliva was transferred to an Eppendorf tube (Thermo Fisher, UK) and centrifuged for 30 minutes at 10,000 x g to remove unwanted larger cellular debris. 5 µL of the supernatant was pipetted onto an aluminium foil-covered glass microscope slide and air dried in an airtight container for 30 minutes. Raman measurements of the dried saliva were carried out immediately.

4.6.3 Tissue Biopsy Collection and Sample Preparation

Colonoscopies and flexible sigmoidoscopy were performed as standard of care and endoscopic activity was measured using Mayo Endoscopic score (MES), Ulcerative Colitis Endoscopic Index of Severity (UCEIS) for UC and Simple Endoscopic Score (SES) for CD (21–23). Tissue biopsies were taken for both histological assessment and Raman spectroscopic analyses. Fresh tissue samples were transferred to an aluminium foil-covered glass microscope slide for RS and immediately analysed. No pre-treatment of the tissue sections was performed to keep the conditions as similar as possible to the *in-vivo* environment. The histological severity of UC was graded

using validated indices such as Robart histopathological Index (RHI), Nancy's histopathological Index (NHI) and the newly developed PICaSSO Histological Remission Index (PHRI) (24–27). The histological assessment of CD samples was graded using RHI. Histological remission was defined as NHI ≤ 1, RHI ≤ 3 without neutrophils in the epithelium or lamina propria, or PHRI = 0 (absence of neutrophils from superficial epithelium and lamina propria). Pathologists were blinded to clinical and endoscopic data. All tissue samples were snap frozen in liquid nitrogen and stored at -80°C after analysis.

4.6 4 Raman Spectroscopy

Raman spectra were acquired using a Renishaw InVia Qontor confocal Raman system equipped with a Leica DMLM (Renishaw PLC, UK). 785 nm excitation laser was used in the analysis and focused using x50 objective. The average output laser power at the sample was 7 mW. All Raman spectra were acquired in the main fingerprint region of 700-1700 cm^{-1}. Spectral maps were obtained over 50 x 50 μm^2 areas for each sample using a 5 μm step size between map points, 10 accumulations of 1-second acquisition per spectra. Overall, 1 map with 100 spectra were collected per sample for saliva and tissue biopsy and used for data processing and analysis.

4.6.5 Data Acquisition and Analysis

Raman data was acquired using WiRE 5.1 (Renishaw PLC, UK) and underwent a polynomial background subtraction cosmic ray removal. Spectra were normalised using the standard normal variate (SNV) using Python (Python 3.7). The signal-to-noise ratio was improved using 10 averaged spectra at each map location. Multivariate

analysis was performed using the SKiNET alongside the Raman Toolkit web interface to build the self-organizing map (SOM) and self-organizing map discriminant index (SOMDI) models. Data sets were grouped and split into test and training data (20:80). 10-fold cross-validation was performed on the training data with optimisation of grid size, initial learning rate and number of training steps. The end model used to classify the test data consisted of a 10 x 10 grid of neurons, 46,080 training steps (4 epochs of the data) and an initial learning rate of 0.1.

4.6.6 Statistical Analysis

Given that the present study was a pilot study, and since utilising SKiNET [14] in the context of IBD was novel, a sample size calculation could not be performed. To analyze the distinctions between the classes, we calculated the average sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy, along with their corresponding standard deviations. SKiNET was used to analyse the datasets, with the samples labelled accordingly (IBD/UC/CD or healthy). The algorithm was evaluated by a 10-fold cross-validation to assess the prediction ability of the model. Data analysis with SKiNET was repeated 10 times, determining the average sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy (95% confidence interval).

4.7 References

[1] Crohn's and Colitis UK, "New research shows over 1 in 123 people in UK living with Crohn's or Colitis," Crohn's and Colitis UK. [Online]. Available: https://crohnsandcolitis.org.uk/news-stories/news-items/new-research-shows-over-1-in-123-people-in-uk-living-with-crohn-s-or-colitis

[2] N. Ghosh and P. Premchand, "A UK cost of care model for inflammatory bowel disease," *Frontline Gastroenterol*, vol. 6, no. 3, pp. 169–174, 2015, doi: 10.1136/flgastro-2014-100514.

[3] I. J. Pence *et al.*, "Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy," *Biomed Opt Express*, vol. 8, no. 2, p. 524, Feb. 2017, doi: 10.1364/BOE.8.000524.

[4] C. Bielecki *et al.*, "Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells," *J Biomed Opt*, vol. 17, no. 7, p. 0760301, Jul. 2012, doi: 10.1117/1.JBO.17.7.076030.

[5] C. Morasso *et al.*, "Raman Analysis Reveals Biochemical Differences in Plasma of Crohn's Disease Patients," *J Crohns Colitis*, vol. 14, no. 11, pp. 1572–1580, Nov. 2020, doi: 10.1093/ecco-jcc/jjaa080.

[6] G. Calado, I. Behl, H. J. Byrne, and F. M. Lyng, "Raman spectroscopic characterisation of non stimulated and stimulated human whole saliva," *Clinical Spectroscopy*, vol. 3, no. February, p. 100010, Dec. 2021, doi: 10.1016/j.clispe.2021.100010.

[7] K. J. I. Ember *et al.*, "Raman spectroscopy and regenerative medicine: a review," *Regen. Med.*, vol. 2, p. 12, 2017, doi: 10.1038/s41536-017-0014-3.

[8] A. Daniel *et al.*, "Comparative study of oral dysplasia by conventional and surface enhanced Raman spectroscopy of whole saliva," in *BIOMEDICAL VIBRATIONAL SPECTROSCOPY 2020: ADVANCES IN RESEARCH AND INDUSTRY*, Petrich, W and Huang, Z, Ed., in Proceedings of SPIE, vol. 11236. 2020. doi: 10.1117/12.2539545.

[9] K. Ember *et al.*, "Saliva-based detection of COVID-19 infection in a real-world setting using reagent-free Raman spectroscopy and machine learning," *J Biomed Opt*, vol. 27, no. 02, 2022, doi: 10.1117/1.jbo.27.2.025002.

[10] R. Akçan, M. Ş. Yildirim, H. Ilhan, B. Güven, U. Tamer, and N. Sağlam, "Surface enhanced Raman spectroscopy as a novel tool for rapid quantification of heroin and metabolites in saliva.," *Turk J Med Sci*, vol. 50, no. 5, pp. 1470–1479, Aug. 2020, doi: 10.3906/sag-1912-196.

[11] S. Desai *et al.*, "Raman spectroscopy-based detection ofRNAviruses in saliva: A preliminary report," *J Biophotonics*, vol. 13, no. 10, Oct. 2020, doi: 10.1002/jbio.202000189.

[12] N. M. Ralbovsky, L. Halámková, K. Wall, C. Anderson-Hanley, and I. K. Lednev, "Screening for Alzheimer's Disease Using Saliva: A New Approach Based on Machine Learning and Raman Hyperspectroscopy," *Journal of Alzheimer's Disease*, vol. 71, no. 4, pp. 1351–1359, 2019, doi: 10.3233/JAD-190675.

[13] Dr. G. Calado *et al.*, "Raman Spectral Study of Saliva: A New Tool for Detection of Malignant and Premalignant Oral Lesions," *Oral Surg Oral Med Oral Pathol Oral Radiol*, vol. 128, no. 1, pp. e90–e90, 2019, doi: 10.1016/j.oooo.2019.02.230.

[14] A. Hole *et al.*, “Salivary Raman Spectroscopy: Standardization of Sampling Protocols and Stratification of Healthy and Oral Cancer Subjects,” *Appl Spectrosc*, vol. 75, no. 5, pp. 581–588, May 2021, doi: 10.1177/0003702820973260.

[15] D. P. Lima, D. G. Diniz, S. A. S. Moimaz, D. H. Sumida, and A. C. Okamoto, “Saliva: reflection of the body,” *International Journal of Infectious Diseases*, vol. 14, no. 3, pp. e184–e188, Mar. 2010, doi: 10.1016/j.ijid.2009.04.022.

[16] C.-Z. Zhang *et al.*, “Saliva in the diagnosis of diseases,” *Int J Oral Sci*, vol. 8, no. 3, pp. 133–137, Sep. 2016, doi: 10.1038/ijos.2016.38.

[17] M. Hardy, L. Kelleher, P. de Carvalho Gomes, E. Buchan, H. O. M. Chu, and P. Goldberg Oppenheimer, “Methods in Raman spectroscopy for saliva studies–a review,” *Applied Spectroscopy Reviews*, vol. 57, no. 3. Taylor and Francis Ltd., pp. 177–233, 2022. doi: 10.1080/05704928.2021.1969944.

[18] K. Hanna, E. Krzoska, A. M. Shaaban, D. Muirhead, R. Abu-Eid, and V. Speirs, “Raman spectroscopy: current applications in breast cancer diagnosis, challenges and future prospects,” *Br J Cancer*, vol. 126, no. 8, pp. 1125–1139, 2022, doi: 10.1038/s41416-021-01659-5.

[19] J. M. Yoshizawa, C. A. Schafer, J. J. Schafer, J. J. Farrell, B. J. Paster, and D. T. W. Wong, “Salivary Biomarkers: Toward Future Clinical and Diagnostic Utilities,” *Clin Microbiol Rev*, vol. 26, no. 4, pp. 781–791, Oct. 2013, doi: 10.1128/CMR.00021-13.

[20] G. Eom *et al.*, “Diagnosis of tamiflu-resistant influenza virus in human nasal fluid and saliva using surface-enhanced raman scattering,” *ACS Sens*, vol. 4, no. 9, pp. 2282–2287, Sep. 2019, doi: 10.1021/acssensors.9b00697.

[21] S. P. L. Travis *et al.*, “Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS),” *Gut*, vol. 61, no. 4, pp. 535–542, Apr. 2012, doi: 10.1136/gutjnl-2011-300486.

[22] E. Koutroumpakis and K. Katsanos, “Implementation of the simple endoscopic activity score in crohn’s disease,” *Saudi Journal of Gastroenterology*, vol. 22, no. 3, p. 183, 2016, doi: 10.4103/1319-3767.182455.

[23] W. Xu *et al.*, “The Mayo Endoscopic Score Is a Novel Predictive Indicator for Malignant Transformation in Ulcerative Colitis: A Long-Term Follow-Up Multicenter Study,” *Front Surg*, vol. 9, Mar. 2022, doi: 10.3389/fsurg.2022.832219.

[24] R. K. Pai *et al.*, “Definitions of response and remission for the Robarts Histopathology Index,” *Gut*, vol. 68, no. 11, pp. 2101–2102, Nov. 2019, doi: 10.1136/gutjnl-2018-317547.

[25] A. Marchal-Bressenot, A. Scherl, J. Salleron, and L. Peyrin-Biroulet, “A practical guide to assess the Nancy histological index for UC,” *Gut*, vol. 65, no. 11, pp. 1919.2-1920, Nov. 2016, doi: 10.1136/gutjnl-2016-312722.

[26] M. Iacucci *et al.*, “Development and reliability of the new endoscopic virtual chromoendoscopy score: the PICaSSO (Paddington International Virtual ChromoendoScopy ScOre) in ulcerative colitis,” *Gastrointest Endosc*, vol. 86, no. 6, pp. 1118-1127.e5, Dec. 2017, doi: 10.1016/j.gie.2017.03.012.

[27] M. Iacucci *et al.*, “An International Multicenter Real-Life Prospective Study of Electronic Chromoendoscopy Score PICaSSO in Ulcerative Colitis,”

Gastroenterology, vol. 160, no. 5, pp. 1558-1569.e8, Apr. 2021, doi: 10.1053/j.gastro.2020.12.024.

[28] S. C. L. Smith *et al.*, “Raman spectroscopy accurately differentiates mucosal healing from non-healing and biochemical changes following biological therapy in inflammatory bowel disease,” *PLoS One*, vol. 16, no. 6 June, pp. 1–16, 2021, doi: 10.1371/journal.pone.0252210.

[29] G. Acri *et al.*, “Raman spectroscopy as noninvasive method of diagnosis of pediatric onset inflammatory bowel disease,” *Applied Sciences (Switzerland)*, vol. 10, no. 19, pp. 1–14, 2020, doi: 10.3390/app10196974.

[30] C. Morasso *et al.*, “Raman analysis reveals biochemical differences in plasma of crohn’s disease patients,” *J Crohns Colitis*, vol. 14, no. 11, pp. 1572–1580, 2020, doi: 10.1093/ecco-jcc/jjaa080.

[31] H. Ding *et al.*, “In vivo analysis of mucosal lipids reveals histological disease activity in ulcerative colitis using endoscope-coupled Raman spectroscopy,” *Biomed Opt Express*, vol. 8, no. 7, p. 3426, 2017, doi: 10.1364/boe.8.003426.

[32] I. J. Pence *et al.*, “Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy,” *Biomed Opt Express*, vol. 8, no. 2, p. 524, 2017, doi: 10.1364/boe.8.000524.

[33] C. Bielecki *et al.*, “Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells,” *J Biomed Opt*, vol. 17, no. 7, p. 0760301, 2012, doi: 10.1117/1.jbo.17.7.076030.

[34] J. Addis, N. Mohammed, O. Rotimi, D. Magee, A. Jha, and V. Subramanian, “Raman spectroscopy of endoscopic colonic biopsies from patients with ulcerative colitis to identify mucosal inflammation and healing,” *Biomed Opt Express*, vol. 7, no. 5, p. 2022, 2016, doi: 10.1364/boe.7.002022.

[35] C. Tefas and M. Tanțău, “Clinical applications of raman spectroscopy in inflammatory bowel diseases. A review,” *Journal of Gastrointestinal and Liver Diseases*, vol. 27, no. 4, pp. 433–438, 2018, doi: 10.15403/jgld.2014.1121.274.app.

[36] C. Tefas *et al.*, “Surface-enhanced Raman scattering for the diagnosis of ulcerative colitis: will it change the rules of the game?,” *Anal Bioanal Chem*, vol. 413, no. 3, pp. 827–838, 2021, doi: 10.1007/s00216-020-03036-2.

[37] B. Li *et al.*, “Non-invasive diagnosis of Crohn’s disease based on SERS combined with PCA-SVM,” *Analytical Methods*, vol. 13, no. 44, pp. 5264–5273, 2021, doi: 10.1039/d1ay01377g.

[38] X. Bi, A. Walsh, A. Mahadevan-Jansen, and A. Herline, “Development of Spectral Markers for the Discrimination of Ulcerative Colitis and Crohn’s Disease Using Raman Spectroscopy,” *Dis Colon Rectum*, vol. 54, no. 1, pp. 48–53, Jan. 2011, doi: 10.1007/DCR.0b013e3181fcf68d.

[39] S. Nikolaus and S. Schreiber, “Diagnostics of Inflammatory Bowel Disease,” *Gastroenterology*, vol. 133, no. 5, pp. 1670–1689, Nov. 2007, doi: 10.1053/j.gastro.2007.09.001.

[40] C. Murr, T. B. Grammer, A. Meinitzer, M. E. Kleber, W. März, and D. Fuchs, “Immune Activation and Inflammation in Patients with Cardiovascular Disease Are Associated with Higher Phenylalanine to Tyrosine Ratios: The Ludwigshafen Risk and Cardiovascular Health Study,” *J Amino Acids*, vol. 2014, pp. 1–6, Feb. 2014, doi: 10.1155/2014/783730.

[41] M. H. Mosli *et al.*, “C-Reactive Protein, Fecal Calprotectin, and Stool Lactoferrin for Detection of Endoscopic Activity in Symptomatic Inflammatory Bowel Disease Patients: A Systematic Review and Meta-Analysis,” *American Journal of Gastroenterology*, vol. 110, no. 6, pp. 802–819, Jun. 2015, doi: 10.1038/ajg.2015.120.

CHAPTER 5

RAMAN SPECTROSCOPIC MOLECULAR FINGERPRINTING OF BIOMARKERS FOR INFLAMMATORY BOWEL DISEASE

The work in this results chapter is based on a previously published paper;

1. **Buchan E.,** Rickard J.J.S., Goldberg Oppenheimer P. [2023], 'Raman spectroscopic molecular fingerprinting of biomarkers for inflammatory bowel disease.' *Clinical and Translational Medicine, DOI: 10.1002/ctm2.1345*

Author contributions – **E Buchan:** Conceptualisation, Investigation, Methodolgy, Formal analysis, Writing – original draft, review and editing. **J Rickard**: Validation, Writing – review and editing. **P Goldberg Oppenheimer**: Conceptualisation, Writing – review and editing, Resources, Funding acquisition, Supervision.

Acknowledgement – Jinglei Yu for help with mass spectrometry analysis.

5.1 Abstract

Inflammatory bowel disease (IBD), including ulcerative colitis (UC) and Crohn's disease (CD) is a chronic condition that affects the gastrointestinal tract. IBDs are increasingly difficult to diagnose and differentiate despite serological testing, imaging, and endoscopic and histopathological assessments. At present there are no known biomarker to predict IBD and as such many patients with non-specific abdominal pain may have to undergo needless invasive procedures. The aim of this study is to develop a Raman spectral library of markers of interest of IBD and systematically explore their potential in disease diagnosis. Raman spectroscopy (RS) is a powerful analytical technique that can be used to identify biomarkers of interest due to its high sensitivity, specificity, non-invasiveness, and portability. This study investigated a panel of biomarkers to include proteins and cytokines. When assessing the potential of molecules as biomarkers of disease we investigate the specificity, sensitivity, accessibility, and association. Of the biomarkers studied C-reactive protein and M2-Pyruvate Kinase combined were powerful indicators of IBD as well as a group of cytokines; IL-4, IL-8, IL-12, and IL-17. Our results indicate the ability of RS as a specific tool for biomarker discovery with the ability to significantly contribute to the development of new diagnostic and therapeutic approaches for IBD.

5.2 Introduction

Inflammatory bowel diseases (IBD), predominantly represented by ulcerative colitis (UC) and Crohn's disease (CD), are chronic inflammatory conditions affecting around 2.5-3 Million individuals in Europe with the incidences continue rising worldwide [1].

CD can affect any portion of the digestive tract, from mouth to anus, and the inflammation can penetrate deeply into the layers of the intestinal walls [2]. UC, on the other hand, primarily affects the colon (large intestine) and the rectum [3], with the resulting inflammation typically limited to the innermost lining of the colon. As these complex conditions involve both acute and chronic disease states, individuals will undergo phases of inflammatory flare-ups, remission, and relapse [4]. At present there is no cure for IBD; nevertheless, treatments and management strategies including prescription medications, surgery and lifestyle modifications can help alleviate symptoms and improve quality of life.

Clinically, a typical diagnosis of IBD involves a combination of medical history, physical examination as well as blood and stool tests and a subsequent endoscopy [5-7]. A range of techniques has been explored to develop new tools for timely and rapid diagnosis of IBD including for instance, computed tomography, magnetic resonance imaging, video capsule endoscopy and infrared and Raman spectroscopy [8-12]. Further ongoing developments of new methodologies for IBD diagnosis have examined the use of microbial biomarkers, such as DNA or RNA, to accurately distinguish between individuals with IBD and healthy controls as well as investigated volatile organic compounds, emitted by living organisms, and can be subsequently detected in breath or other bodily fluids [13-15].

Collection of biofluids, *e.g.,* blood / urine, is a standard procedure in routine clinical practice and hence, is ideal for a point-of-care or bedside settings. Biofluids derived biomarkers have many useful applications in healthcare including disease prevention

and detection, determination of individual's risk and disease monitoring [16]. Currently, however, there are no validated blood biomarkers to accurately and reliably predict IBD. The complexity of interactions between host factors and the dynamic fluctuation of the gut microbiota during IBD hinders the identification of consistent changes in microbial composition and thus, the universal biomarkers for disease prediction. [17] Hence a large proportion of patients with non-specific abdominal pain often undergo unnecessary painful endoscopy or colonoscopy to rule out IBD [18]. Whilst effective, these are invasive and painful and thus, less unsuitable for routine use for timely and rapid IBD diagnosis. Not only are these procedures invasive but also, a confirmatory IBD diagnosis requires further clinical examinations, imaging and pathological investigations, which can be often inconclusive. A collection of diagnostic and/or prognostic biomarkers of IBD would thus, significantly decrease or eliminate the necessity for invasive or nonspecific procedures. This would greatly enhance early-stage diagnosis, management and therapeutic interventions for IBDs. The ability to determine the type, severity and patient response to therapeutics has long been a priority of clinical researchers. Potential IBD indicative biomarkers currently under investigation include C-reactive protein (CRP), faecal calprotectin, anti-*Saccharomyces cerevisiae* antibodies (ASCA), anti-neutrophil cytoplasmic antibodies (ANCA) and serum amyloid A [19-23]. Cytokines, proteins produced by immune cells, are also known to play an active role in inflammation and thus, are additional candidate biomarkers in IBD [24]. Idiopathic IBDs typically occur in clinically immunocompetent individuals with cytokine-driven inflammation of the gut giving rise to the characteristic signs and symptoms in individuals. CD is most often associated with increased levels of IL-12/IL-23 and IFN-γ/IL-17, whereas UC is associated with excess IL-13 production

[25]. Research studying the pathogenesis of IBD suggests that IBD arises due to a dysfunctional interaction between the bacterial microflora of the gut and the mucosal immune system and thus, cytokines are well known to play a key-role in controlling intestinal inflammation and the associated diseases [26]. Cytokine testing therefore, has the potential to support the IBD diagnosis due to its low cost and lack of invasiveness compared to routine IBD diagnostics. Although demonstrating a correlation with the severity of IBD and offering potential diagnostic, prognostic, and stratification value, these biomarkers cannot be detected unless subjected to analysis in specialised laboratories. Analysis typically involves techniques such as ELISA, enzyme immunoassay, chemiluminescent assay, high-performance liquid chromatography (HPLC), or mass spectrometry [27-30]. Although reliable, these techniques have time-to-results in the range of days, require complex equipment and trained personnel and do not meet the need for detection of ultra-low levels of target analytes in complex biological samples using rapid and direct readouts. ELISAs are relatively cost-effective but, despite advancements, suffer from poor stability, not rapid enough, require expensive labels for each biomarker and result in unacceptable rates of false negatives and false positives [31]. While the challenges of accurate diagnosis and monitoring of IBD have created an urgent need for rapid detection of the indicative biochemical changes, currently no validate biomarkers and suitable sensing techniques exist that can address such a critical need for the 21st century.

Herein, we present Raman spectroscopy (RS) as a systematic approach for profiling and categorising potential biomarkers indicative of inflammatory bowel disease (IBD) [32-35] towards establishing important biochemical ‘fingerprints’ as characteristic

barcodes that can be used in current and future studies for diagnostic and prognostic applications in IBD. We systematically evaluate Raman spectra of multiple potential IBD biomarkers and compare these with spectral fingerprints of IBD saliva and tissue. We further investigate the cytokine profiles of those to gain insight into the underlying immune and inflammatory processes involved in IBD. Through close examination of these, we identify the specific cytokines that are either over- or underproduced, providing important clues to the underpinning mechanism of disease and future therapeutic targets. The acquired spectral data is classified using our new artificial neural networks algorithm, self-optimising Kohonen index network (SKiNET) as a decision support tool, based on separation of data classes in a self-organising map (SOM) with a classification *via* the self-organising map discriminant index (SOMDI). The overall established spectroscopic workflow (**Fig.1**), from quantitative spectral acquisition, through multivariate analyses and onto fingerprint barcoding and AI classification, puts forward several potential molecular biomarkers, whilst demonstrating relevant applications for the clinical and biomedical fields including biomarker discovery and patient stratification thus, laying the platform for future studies needed to define their functional relevance in the complex aetiology of IBD.

5.3 Results and Discussion

A panel of potential IBD-biomarkers including cytokines IL-4, IL-8, IL-10, IL-12, IL-17, IL-23, Ferritin, M2-pyruvate kinase (PKM2), Calprotectin, CRP and *α*-1-antitrypsin (Fig.2), chosen based on their specificity for IBD and known physiological response in patients (**Fig.5.1a**), was assessed *via* hybrid Raman spectroscopy (**Fig.5.2**) and subsequently, classified *via* optimised AI algorithm with inherent self-organised map

discriminant index (SOMDI) spectroscopic fingerprints analysis of saliva and tissue biopsies, assigning the dominant spectral signatures and yielding the characteristic barcodes.

a

Biomarker	Type	Specificity in IBD	Physiological Response in IBD Patients	References
a-1-antitrypsin	Protein	Indicator of disease activity	Identified 2.4-fold upregulation	[1-2]
Calprotectin	Protein	Indicator of clinical and endoscopic disease activity	6-fold increase in levels observed in IBD patients	[3-4]
C-Reactive Protein (CRP)	Protein	Used as an assessment of acute inflammation	Increase in levels from 10-40 mg/L up to 200-250 mg/L	[5-6]
Ferritin	Protein	Indirect estimate of body iron stores	Decrease in levels from 100 µg/L to 28 µg/L detected	[7-8]
M2 Pyruvate Kinase (PKM2)	Enzyme	Mediator of the inflammatory process with a crucial role in immune cell recruitment	6-fold increase in levels observed in IBD patients	[9-10]
Interleukin-4 (IL-4)	Cytokine	Critical driver of immune activation and inflammation. Exerts anti-inflammatory and pro-healing response	Decrease in plasma IL-4 levels with IL-4 observed in 81.2% healthy controls, 77.8% in UC patients and 0% of CD patients	[11-12]
Interleukin-8 (IL-8)	Cytokine	Pro-inflammatory cytokine with a central role in neutrophil activation associated with the progression of disease	Higher levels observed in patients with active IBD compared to those with inactive disease. After treatment IL-8 levels found to decrease from 37 to 20 pg/mL	[13-14]
Interleukin-10(IL-10)	Cytokine	Regulatory cytokine which inhibits antigen presentation and pro-inflammatory cytokine release	Decreased levels observed with IL-10 detected in 91.7% of healthy controls but only 30% of IBD patients	[15-16]
Interleukin-17 (IL-17)	Cytokine	Pro-anti-inflammatory cytokine associated with altered immune and inflammatory response in intestinal mucosa	Levels significantly higher in IBD v healthy (949 pg/mL v 484.8 pg/mL)	[17-18]
Interleukin-23 (IL-23)	Cytokine	Critical promoter of chronic intestinal inflammation	IL-23 levels found to increase significantly from 371.5 pg/mL to 1365.1 pg/mL)	[19-20]

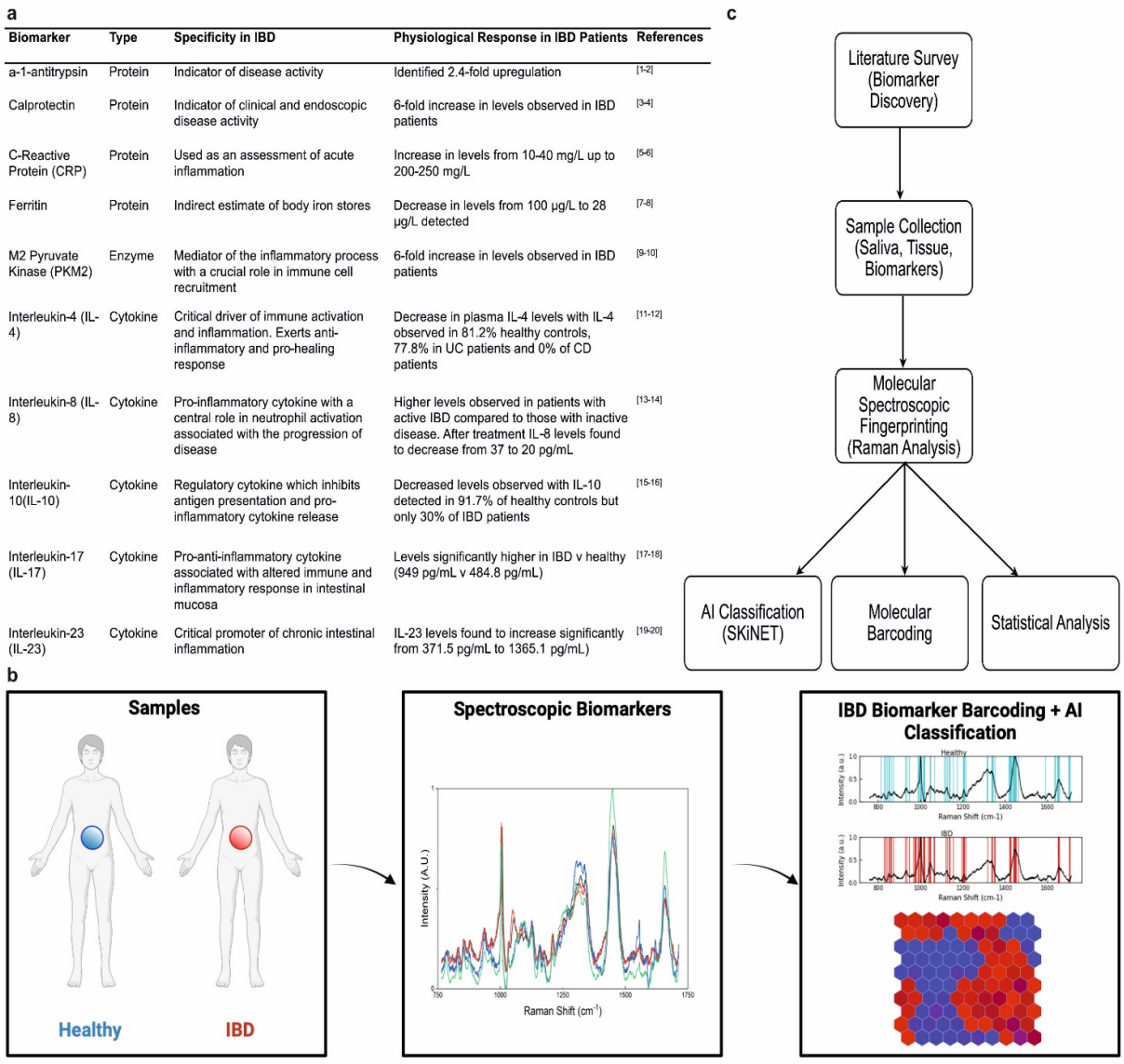

Figure 5.1. The established spectroscopic workflow, from quantitative spectral acquisition, through multivariate analyses and onto fingerprint barcoding and AI classification including the (a) overview and rationale behind the selected candidate biomarkers, describing their specificity in IBD alongside the recognised physiological response in IBD patients. Cytokines, known to play an active role in inflammation were chosen for their ability to contribute to the inflammatory status of the intestine, CRP as an overall indicator of inflammation, M2-Pyruvate Kinase as a mediator of the inflammatory process,

calprotectin allowing evaluation of gut inflammation and α-1-antitrypsin as an indicator of disease progression. Each biomarker identified in the literature has a known physiological response in IBD patients with a statistically significant increase or decrease observed such as, a 6-fold increase in PKM2 levels, 2.4-fold upregulation in α-1-antitrypsin and a decrease in IL-8 levels from 37 to 20 pg/mL. In this workflow (b) biofluid samples were obtained from patients, analysed to identify unique spectroscopic signatures using a comprehensive multiplex Raman analysis with subsequent barcoding and advanced Self-optimising Kohonen index NETwork (SKiNET) machine learning for validation and classification. (c) The overall workflow describing the individual components and their corelation within the developed Raman profiling and spectroscopic fingerprinting towards candidate IBD biomarkers discovery (Supplementary information-S5).

We have identified a panel of potential IBD biomarkers, previously described in literature [46-48], including a range of cytokines, *i.e.,* IL-4, IL-8, IL-10, IL-12, IL-17, IL-23 (**Fig.5.1a**) and Ferritin, M2-pyruvate kinase (PKM2), calprotectin, C-reactive protein and α-1-antitrypsin (**Fig.5.2b**). Due to conflicting reports on the well-known anti-Saccharomyces cerevisiae antibody (ASCA) and perinuclear antineutrophil cytoplasmic antibodies (pANCAs) and due to their poor sensitivity of only 55% [49], these potential markers were excluded from our study. We have subsequently assessed their relevance through a direct comparison of spectral fingerprints of both saliva and tissue biopsy (**Fig.5.2, c-d**).

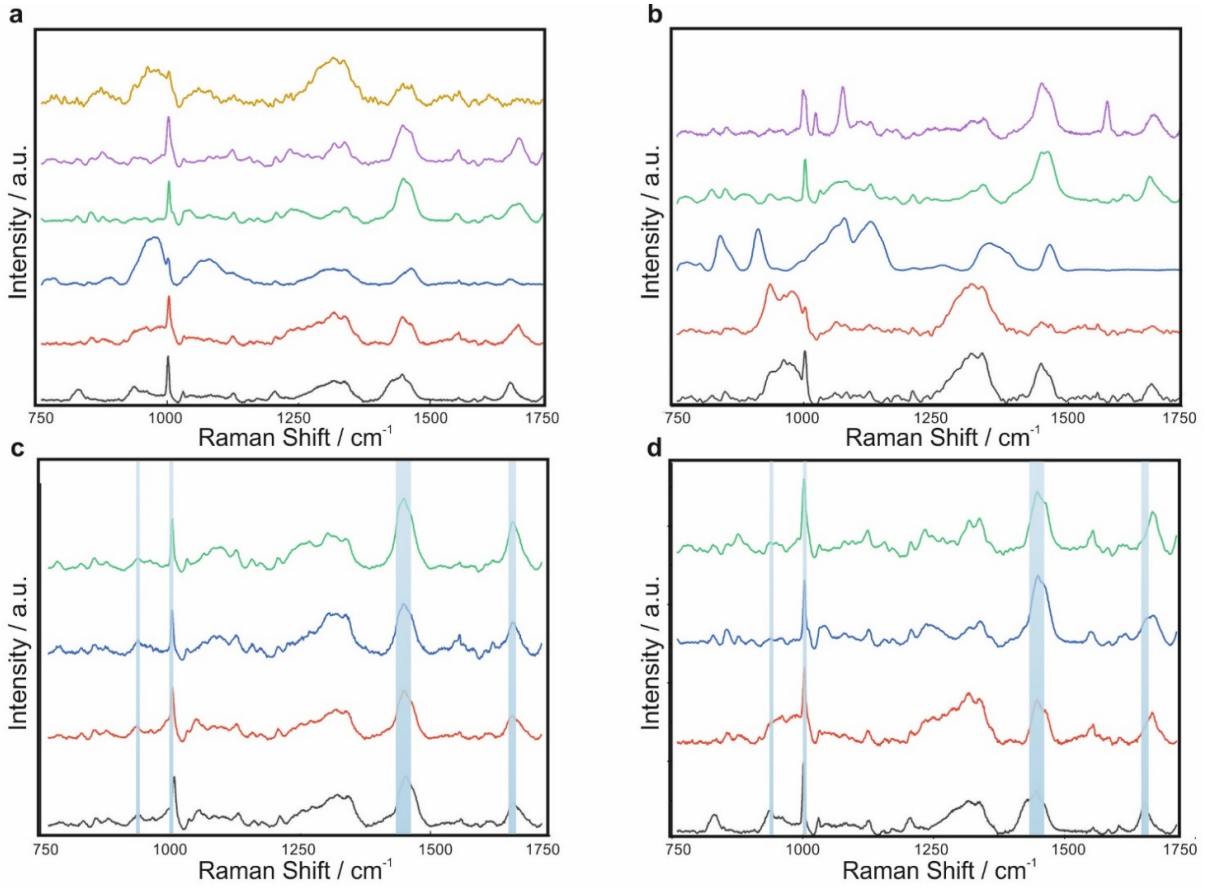

Figure 5.2. Average Raman spectra of candidate IBD biomarkers including **(a)** cytokine profiling of IL-4 (black), IL-8 (red), IL-10 (blue), IL-12 (green), IL-17 (purple), IL-23 (light brown) and **(b)** a-1-antitrypsin (black), calprotectin (red), C-reactive protein (blue), ferritin (green) and M2-Pyruvate kinase (purple). Mean representative spectra of **(c)** saliva of healthy (black) and IBD (red) and biopsy tissue of healthy (blue) and IBD (green) patients and of the identified IBD biomarkers **(d)** IL-4, IL-8, IL-12, and IL-17 highlighting the peaks of interest at 936, 1003, 1340, 1445 and 1656cm^{-1}. The detected primary peaks at 936cm^{-1} (protein), 1003cm^{-1} (phenylalanine), 1340cm^{-1} (C-H deformation in proteins) 1445cm^{-1} ($\delta(CH_2$ deformation of proteins) and 1656cm^{-1} (Amide-I) provide the fingerprints of both IBD saliva and tissue biopsy. Notably, the peak at 936cm^{-1} exhibits a decreased intensity in the IL-4 due to C-C stretch of the amino acids in the protein backbone.

A trio of characteristic bands at 821, 851, 924cm^{-1} were identified for PKM2, emerging as a significant mediator of inflammatory processes with serum levels 6-fold higher in IBD *versus* healthy patients [1], posing PKM2 as a putative IBD-biomarker (**Fig.2b/S3**).

Prior research by Czub *et* al. identified increased fecal PKM2 levels 100% of active CD patients and 94.3% of active UC patients with associated enzyme reactivity higher in all IBD patients than those healthy controls. Here we identified PKM2 at increased levels in both saliva and tissue at each of the three characteristic Raman bands, thus indicating its potential as an identifier of IBD in both tissue and saliva as opposed to its current use in faecal samples. In addition, whilst CRP exhibited increased intensities at 841, 912, 1076 and 1126 cm^{-1} ($p^{***}<0.0001$), this marker alone was found to be non-specific due to the elevation in systemic inflammatory disease other than those identified within the intestinal tract, for example in cancers of the body or infections, CRP levels can also be significantly elevated. However, when combined with other potential markers of IBD, such as PKM2, IL-4, IL-8, IL-12 and IL-23, CRP provides invaluable insight into the level of inflammation in the gut. Increased levels of CRP for example, can help differentiate mucosal active disease from quiescent IBD as well as acting as a predictor as to the need for colectomy through the reflection of severe ongoing and uncontrolled inflammation of the gut [49]. Cytokines IL-4, IL-8, IL-12, IL-17 exhibited specific Raman bands, providing identification and discrimination of IBD (**Fig.5.2a**), where the detected peaks yield fingerprints of saliva and tissue biopsies (**Fig.5.2c-d/Supplementary Information-SC.3**).

We have subsequently compared the relative peak intensities of cytokines with those in IBD saliva and biopsies, identified *via* SKiNET and SOMDI, selecting those with the greatest influence on the classifier (**Fig.5.3/Supplementary Information-Table SC.2**). This yielded the detection of several Raman key-bands, which were characteristic in patients with IBD including the 936cm^{-1} (C-C stretching-amino acids,

protein backbone), 1003cm^{-1} (Phenylalanine), 1340cm^{-1} (C-H protein deformation), 1445cm^{-1} ($\delta(CH_2)$ deformation of proteins and lipids) and 1656cm^{-1} (C=O proteins stretch, Amide-I). We further discovered considerably downregulated 936, 1340cm^{-1} and upregulated 1003, 1445, 1656cm^{-1} spectral signatures in IBD patients (**Supplementary Information-SC.3**).

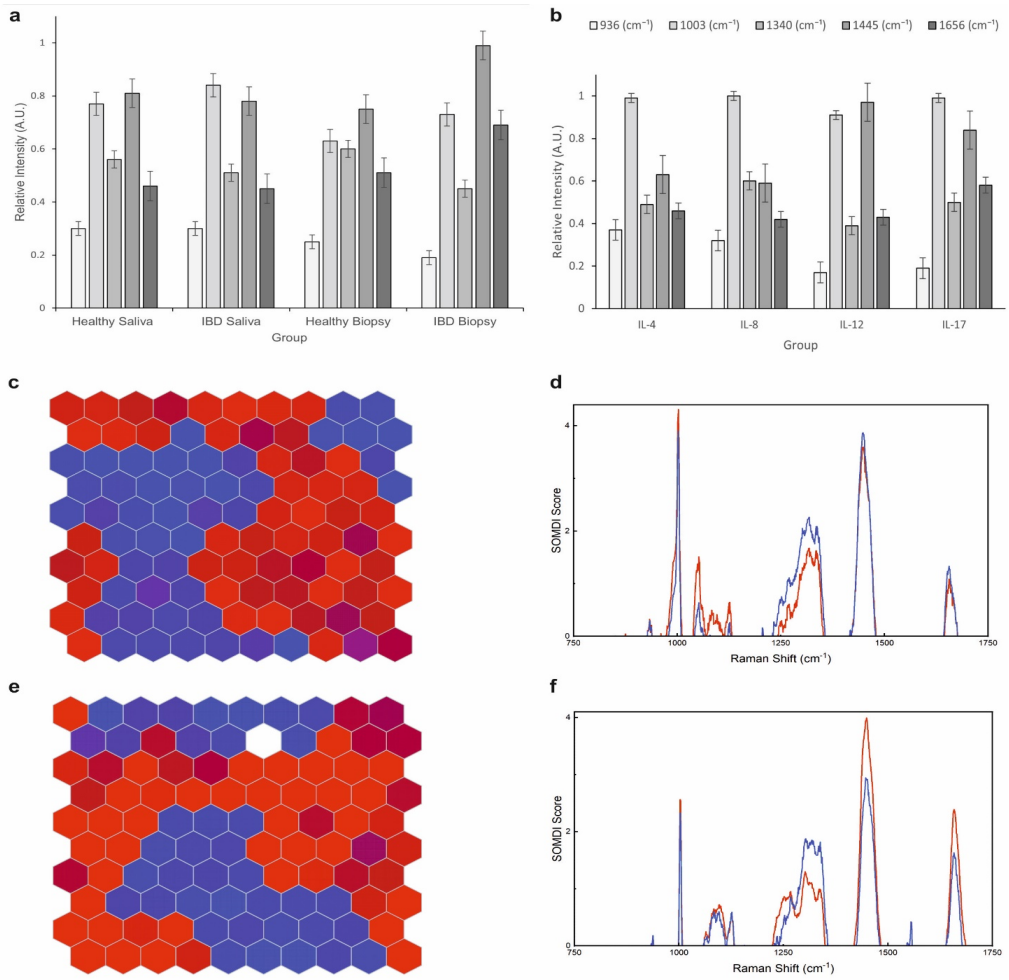

Figure 5.3. Histograms of the relative intensities of the dominant Raman peaks identified in **(a)** healthy and IBD saliva and tissue biopsies and **(b)** potential IBD biomarkers (Error bars indicate the standard error). Spectral data is normalised using standard-normal-variate (SNV) normalisation prior to machine learning to enable enhanced and accurate comparison and analysis *via* subtraction of each spectrum

from its own mean and a subsequent division by the standard deviation. Representative self-organising map (SOM) classification according to disease state of IBD (red) and healthy (blue) of **(c)** saliva and **(e)** tissue biopsy with the corresponding self-organising map discriminant index (SOMDI) extracted features **(d** and **f)** from SOMs, highlighting the most influential Raman bands responsible for the clustering observed in SOM **(c** and **e)**, classifying the patient samples according to diseased state with an accuracy of 88.5% and 86.8%, respectively. Comparison of saliva and tissue biopsies reveals distinct spectral changes with dominant bands at 936cm^{-1} (protein) , 1003cm^{-1} (phenylalanine), 1110cm^{-1} (C-C bond) , 1340cm^{-1} (C-H deformation), 1445cm^{-1} (δCH_2 deformation of proteins/lipids) and 1656cm^{-1} (Amide-I) identified *via* SOMs with significant spectral features extracted *via* SOMDI, distinguishing between control and IBD cohorts. Decreased intensity of the 936cm^{-1} mode in IBD saliva and tissue biopsy for IL-8, IL-12 and IL-17 is associated with the C-C stretching in amino acids with the increased number of hydrogen bonds leading to its weakening.

Levels of IL-4 were decreased for healthy individuals when compared with their IBD counterparts and increased for IL-8, IL-12 and IL-17 as reflected *via* the SOMDI-identified significant Raman shifts (**Fig.5.3d-f)**. Given their widely accepted role in IBD, *via* detecting the changes discriminating between diseased and healthy individuals, we highlight the significance of these markers in the pathogenesis of IBD (**Fig.5.3/Supplementary Information-SC.3**). The decreased intensity of IL-4 at 936cm^{-1} correlates with the known changes in pro-inflammatory cytokines, which mediate interaction between immune and non-immune cells, previously shown contributing to the inflammatory status of the intestine, where IBD-related IL-4-mediated downregulated activation has been impaired [50]. In UC, the T-helper 2 response consists of IL-4 and IL-13. IL-4 is an important contra inflammatory cytokine, which limits monocyte and macrophage activation. IL-4-mediated downregulation of activation has been shown by Ruckert *et* al. to be impaired in IBD [51]. Hoving *et* al.

have also indicated that mice deficient in IL-4, IL-13 and IL-4 receptor-alpha on all cells develop an exacerbated IBD phenotype [52.] A further study by Xiong *et* al. identified the effects of IL-4 and IL-10 gene therapy on TNBS-induced murine colitis, shown to significantly inhibit the TNBS-induced colon tissue damage and disease activity index as well as a marked block in expression of IFN-γ and TNF-α, highlighting the significance of monitoring IL-4 levels in IBD patients as well as the potential of IL-4 as a first line of defence in the treatment of IBD [53].

The ~27% upregulation in the 1003cm^{-1}, detected for each of the cytokines, is indicative of an increased inflammatory response due to a greater influx of inflammatory cells in the diseased state. In IBD, neutrophils are known to be important cellular mediators with the IL-8 being a powerful chemoattractant found in increased quantities in mucosa. Marked increased levels of IL-8 at 1003 and 1656cm^{-1} are indicative of protein confirmational changes, suggesting IL-8 as a significant indicator of disease with a potential to monitor mucosal healing and response to therapeutics.

The statistically significant detected (p***<0.0001) increases of these bands combined with the decrease at 1445cm^{-1} are indicative of IL-12, known to play a key-role in activation and regulation of multiple cytotoxic immune cells including macrophages, natural killer and T cells [54-56], highlighting its importance in following the pathogenesis and progression of IBD. IL-12 is rarely detected in normal intestinal mucosa and therefore, acts as an appealing potential marker of IBD. Recently, IL-12, in combination with other interleukins such as IL-23, has been emerging as a target molecule for treatment in IBD patients, for example, the use of Ustekinumab, an anti-

IL-12/23p40 antibody has been approved for CD. Almradi *et* al. have also indicated that in early stages of IBD, IL-12 is the dominant p40-containing cytokine, driving inflammation in response to intestinal barrier disruption [14].

A further ability of IL-17, exerting a strong inflammatory response, to act as a multiplexed IBD biomarker is established with evident increases at 1003 and 1656cm^{-1} [57]. IL-17 is known to exert a strong proinflammatory response. Its secretion has been reported to be limited to T lymphocytes with major indications of it being a strong mediator of inflammatory response in various tissues [58-61] as well as enhancing the proinflammatory response induced by IL-1β and TNF-α. Previous studies have shown that IL-17 expression has been detected in inflamed mucosa of active UC and CD patients *via* immunohistochemical techniques and serum ELISA [62]. Levels of IL-17 have been also previously found to increase significantly from 371.5 pg/mL to 1365.1 pg/mL in patients with IBD [17].

The fact that the same set of peaks are visible in both the healthy and diseased sample types would thus indicate that their biochemical composition is similar, which is consistent with our general understanding of inflammation. Inflammation does not introduce new metabolites into the system but instead, leads to an overproduction or overuse of the existing metabolites. Therefore, it is expected to see the same molecular changes as identified *via* intensity changes of the peaks in the same region in both tissue types as was evidenced in our study, supporting that these are due to the inflammatory response, where the variation in peak intensities between the two

types are thus due to the detected differences in the concentration of the biomolecules. Furthermore, the significant increase in intensity at 1003cm^{-1} is an additional indication of inflammation due to IBD, since in previous studies, this peak has been associated with higher levels of phenylalanine, due to the immune activation and inflammation. Smith *et* al. observed a significant reduction in intensity due to mucosal healing and Addis et al. recognised a similar effect, following biological therapy [55]. An additional finding relating to the peak changes at 1656cm^{-1} (Amide I) could indicate conformational changes induced by factors evolving from the inflammatory state. Although, it is possible that factors stemming from age, dehydration or radiological conditions, could also be responsible for the changes detected, these do not affect the data reported in our study as the recruited subjects were of a highly similar age, with no known health conditions thus predominantly, the biochemical changes detected are indicative of the inflammatory response associated with IBD.

Our study subsequently, has proceeded to identify which biomarkers associated with IBD could be responsible for the intensity changes found between healthy and IBD patient samples within these regions. Changes in response to IBD, have shown downregulation at 936cm^{-1}, (representative of C-C stretching associated with amino acids such as glutamine and arginine which influence the progression of IBD in individuals) with the main statically significant differences arising from amino acid composition ($p^{***}<0.001$). Our results of decreased intensity at 936cm^{-1} in IBD patients is further consistent with previous studies, where changes in pro-inflammatory cytokines, which mediate interaction between immune cells and non-immune cells, have been shown [63-69] to contribute to the inflammatory status of the intestine.

From the spectroscopic profiling and the corresponding relative peak analyses, the most dominant spectral changes, determined to be statistically significant for detecting the different biomarkers, arising from bands at 936, 1003, 1340, 1445 and 1656cm^{-1} (p***<0.0001), establish these for barcode fingerprints (**Fig.5.4**). An accurate diagnostic and prognostic biomarkers pool for IBD remains an unmet clinical need. The derived barcodes combined with the identified biochemical changes and their close match with notable biomarker characteristics, establish a group of potential first-line screening biomarkers of IBD.

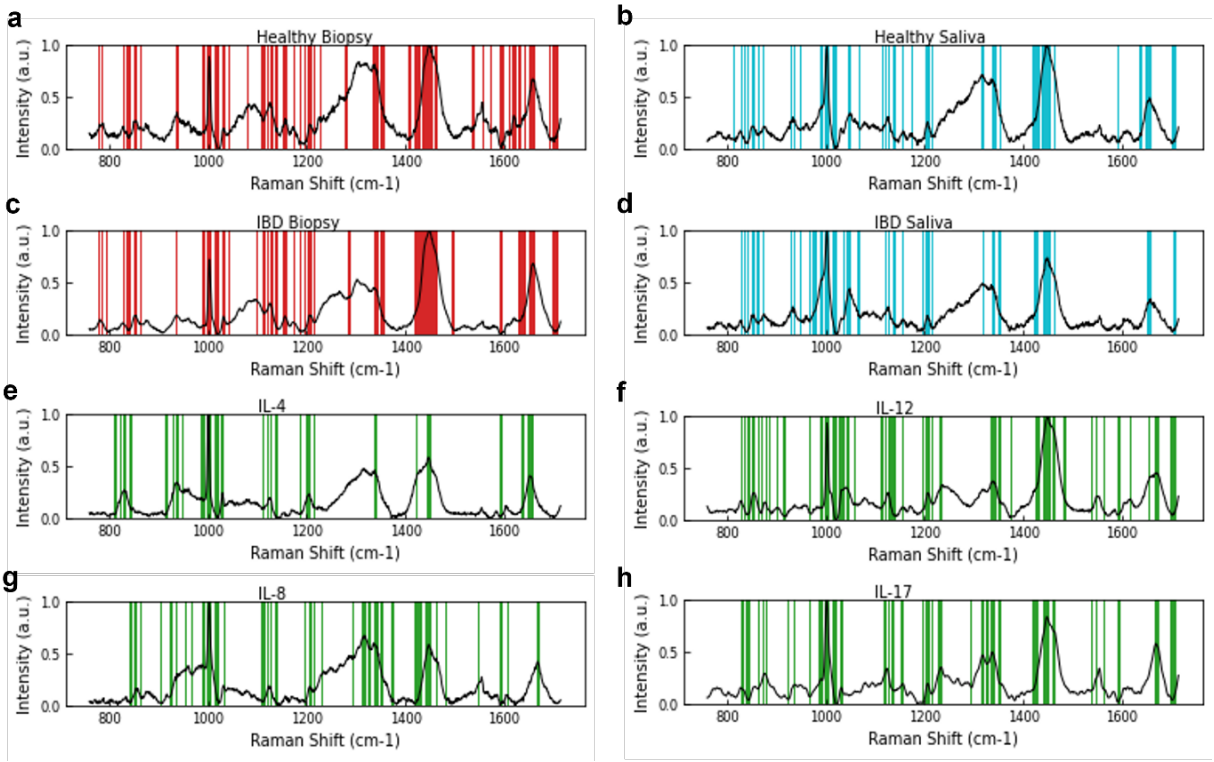

Figure 5.4. Barcodes derived from **(a)** and **(c)** tissue biopsy, **(b)** and **(d)** saliva and potential IBD biomarkers **(e)** IL-4, **(f)** IL-12 **(g)** IL-8 and **(h)** IL-17, highlighting the dominant, statistically significant peaks. Overall, the hybrid Raman spectroscopy, combined with advanced AI algorithm of SKiNET-SOMDI supervised machine learning, used to establish the complete linked workflow from projection, though classification to the development of molecular fingerprinting, introduces the spectroscopic approach for profiling and classification of IBD-indicative biomarkers, enabling the assessment of biomarkers for inflammatory bowel disease. The derived spectroscopic molecular barcodes **(a-h)**

combined with the identified biochemical changes, closely matching notable biomarker characteristics, establish a group of potential first-line screening of IBD-biomarkers.

Table 8.1. Relative Raman intensity of main statistically significant and dominant spectral peaks identified.

Marker	**Raman Shift / cm^{-1}**				
	936	**1003**	**1340**	**1445**	**1656**
Healthy Saliva	0.29	0.77	0.56	0.81	0.46
IBD Saliva	0.28	0.84	0.51	0.78	0.45
Healthy Biopsy	0.25	0.63	0.60	0.75	0.51
IBD Biopsy	0.19	0.73	0.45	0.99	0.69
IL-4	0.37	0.99	0.49	0.63	0.45
IL-8	0.28	1.00	0.60	0.59	0.42
IL-12	0.17	0.91	0.39	0.97	0.43
IL-17	0.19	0.99	0.50	0.84	0.58

Mass spectrometry was used to analyse both healthy and IBD saliva samples (**Fig.5.5)**. Relative abundance of the majority of detected ions varied between the samples with the primary observed proteins and their corresponding coverage, as summarised in **Table 5.2**.

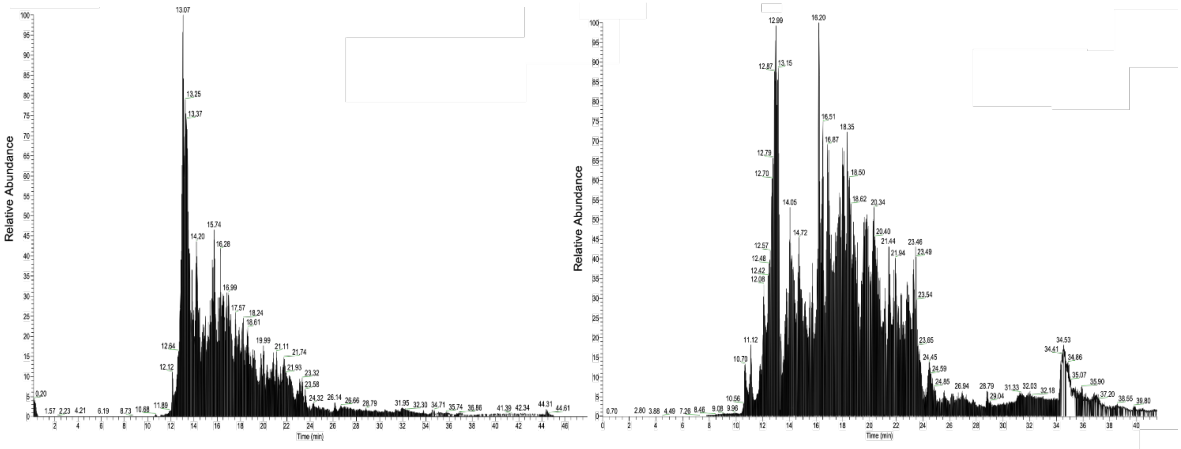

Figure 5.5. Mass spectrometry of the total ion chromatographs for **(a)** healthy and **(b)** unhealthy-IBD saliva, indicating the relative abundancies of all ions in each mass spectrum as highlighted in the corresponding time stamps.

The majority of observed proteins varied significantly in their coverage between healthy and IBD samples particularly in those proteins involved in inflammation and the inflammatory cascade. Of a particular importance was the identification of IL-4, IL-8 and IL-17 receptors which were identified in IBD saliva with only IL-8 detected in healthy saliva. Also of note were the differing levels of S100-A9 and S100-A8 with both found at significantly higher levels in IBD saliva. These proteins are both known to have a role in inflammation with involvement in inducing secretion of pro-inflammatory cytokines including TNF-α and IL-1β, with these mediators understood to exaggerate and sustain inflammatory conditions. Increasing levels of S100A8 and A9 has also been shown to significantly increase protein release of Il-6 and IL-8 [68]. In addition, there was a significant decrease in the levels of both alpha-amylase and mucin-5 in IBD patient sample compared with healthy indicative of differing salivary secretion rates between the groups. Pyruvate kinase was also identified in IBD patient saliva but was notably absent in healthy patient saliva, thus indicative of elevated levels in

patients with IBD. Fecal pyruvate kinase is readily applied in the diagnosis of disease however, its presence in saliva observed with both Raman spectroscopy and mass spectrometry indicates its ability to act as a non-invasive salivary biomarker of IBD.

Table 5.2. Mass spectrometry protein assignment and coverage for healthy and IBD saliva samples.

Protein	Coverage (%)		Molecular Weight (kDa)
	IBD	Healthy	
S100-A9	36	20	13.2
S100-A8	32	18	10.8
Annexin A1	18	65	38.7
Lipocalin-1	6	-	19.2
Interleukin-8	4	1	8.4
Matrix Metalloproteinase-9	1	1	78.4
Pyruvate Kinase	3	-	57.9
Interleukin-17	2	-	35.0
Alpha-amylase	47	71	57.7
Mucin-5	4	14	596
Cornulin	7	42	53.5
Interleukin-1 receptor	11	8	6.19
Trefoil factor-3	12	-	10.2
Interleukin-4 receptor	2	-	14.9

5.4 Conclusions

Raman spectroscopy has been successfully demonstrated as a powerful, specific, and non-invasive technique for detection and multiplexed profiling of IBD-indicative candidate biomarkers, establishing spectral fingerprints of many biomolecules of interest. Rapid differentiation of biomolecular composition has been enabled from the acquired spectral profiles, establishing unique molecular barcodes for candidate biomarkers of IBD as well as saliva and tissue biopsies. The subsequent classification

of the molecular spectroscopic profiles *via* SKiNET ANN algorithm provided identification accuracy of >88% and average sub-classification of 82%. Whilst each of MPK2 and CRP on their own have been identified as indicators of disease, together they present a powerful combination as specific potential markers of inflammatory bowel disease with significant associated spectroscopic changes established at 936, 1003, 1340, 1445 and 1656 cm^{-1} Raman bands. Rapid spectral profiling has further successfully identified a range of inflammatory cytokines as key indicators of IBD including the IL-4, IL-8, IL-12, and IL-17.

Overall, through a combination of simultaneously detecting a panel of identified biomarkers and cytokines at quantifiable levels, threshold of which will need to be established, multiplex Raman spectroscopic profiling and barcoding along with AI classification network lays the platform for the possibility of rapidly detecting the inflammation associated with IBD, discriminating between the disease subtypes as well as monitoring the therapeutic response to various treatments for a timely, non-invasive disease diagnosis. Building upon demonstrating relevant applications for the clinical and biomedical fields including biomarker discovery and patient stratification with a potential to contribute significantly to the development of new diagnostic methodologies and therapeutic monitoring, further studies validating these findings will lay the platform for defining their functional relevance in the complex aetiology of IBD as well as cementing Raman spectroscopy as a powerful technique for discovery of biomarkers in other diseases and biological samples with many ramifications.

5.5 Methods

5.5.1 Sample Collection and Preparation

Saliva was collected from 101 participants (fifty one IBD and fifty healthy) at Queen Elizabeth Hospital Birmingham, UK. All healthy volunteers had no previous health issues or known conditions with participants. 5 mL of unstimulated saliva was collected from each participant in a 50 mL Falcon tube (Thermo Fisher) *via* the passive drool method. Subsequently, 5μL of raw saliva was pipetted onto aluminium slide and air dried in an airtight container for 30 minutes. Colon tissue biopsies were collected from 44 participants (32 IBD and 12 healthy) in the outpatient endoscopy into specimen containers (Thermo Fisher) containing 2mL PBS and analysed immediately. All candidate markers were purchased directly from Sigma-Aldrich and Miltenyi Biotec Cologne) and subsequently diluted in PBS (pH 8) to a working concentration of 1μg/mL. 5μL of each was then pipetted onto an aluminium foil glass slide and dried in an airtight container for 30 minutes.

5.5.2 Raman Spectroscopy

Raman spectra were acquired using a Renishaw *InVia* Qontor confocal Raman microscope equipped with a microscope Leica DMLM and 785 nm laser (Renishaw PLC). Laser light was focussed using a x50 objective lens with a laser power of 5mW. Spectral maps were acquired with an area of 50x50μm^2 in the fingerprint region of 700-1700 cm^{-1} with 5μm step size, 10 accumulations and 1s exposure time per spectrum. Overall, 100 spectra per sample were collected and used for data processing and analyses.

5.5.3 Data Acquisition and Analysis

The data was collected using WiRE 5.1 further employed for the polynomial background subtraction and the removal of cosmic rays. Normalisation was done using the standard normal variate (Python 3.7), which was used to derive the multi-chemical barcoding of the data. Savitzky-Golay filter was applied to calculate the second derivative of each spectrum. Smoothing window was set to 21 with a polynomial order of 2. Maximum peak heights with absolute values over 40% were assigned a value of 1 with values below 40% assigned a value of 0. Values were overlaid on the averaged spectra with reference to the main peaks identified for the barcode generation. Statistical analysis was performed using a one-tailed students *t*-test, peaks identified as having a $p<0.001$ were determined as statistically significant.

5.5.4 Data Processing and Multivariate Analysis

Multi-variate analysis was performed using the self-optimising Kohonen index network (SKiNET) based on self-organising maps (SOMs) with the accompanying Raman Toolkit web interface to build SOM models using training data and perform predictions against test data. SKiNET models were optimised by performing 10-fold cross validation on the training data, and tuning the number of neurons, initial learning rate and number of training steps. The final model used a 10x10 grid of neurons, 38,200 training steps (6 epochs of the data), with an initial learning rate of 0.3. The optimised model was subsequently used to classify the previously unused test data. Classification using the test data were repeated ten times from separate SOM initialisations. To achieve higher accuracy the SOM size, learning rate and number of epochs was empirically tested, with classification accuracy determined using a 10-fold

cross validation. Stability of the model was further verified by running a repeat initialisation of the classification six times.

5.5.5 Mass Spectrometry

5.5.5.1 Trypsin digestion

For trypsin digestion, 10 µL of sample, containing up to 10 µg of protein, was combined with 40 µL of 100 mM ammonium bicarbonate (pH 8). Subsequently, 50 µL of 10 mM dithiothreitol (DTT) was added, with sample incubated at 56ºC for 30 minutes. After cooling to room temperature, cysteines were alkylated with the addition of 50 µL of 50 mM iodoacetamide, followed by mixing and further room temperature incubation for 30 minutes. 50 µL trypsin gold (Promega, Southampton, Hampshire, UK, 6 ng/nL) was added to the samples and incubated overnight at 37ºC.

5.5.5.2 Desalt samples

Samples containing a mix of peptides were desalted using Millipore C18 ZipTips. Tips are prepared by pre-wetting with 100% acetonitrile and rinsed in 2 x 10 µL of 0.1% formic acid. The samples were pipetted through the tip 10 times then washed with 3 x 10 µL 0.1% formic acid to remove excess salts prior to elution of the peptides with 10 µL of 50% acetonitrile/water/0.1% formic acid. Samples were dried down to remove acetonitrile, and then re-suspended in 0.1% formic acid solution in water.

5.5.5.3 LC-MS/MS Experiment

The UltiMate® 3000 HPLC series (Dionex, Sunnyvale, CA USA) was used for peptide concentration and separation. Samples were trapped on precolumn, Thermo Scientific Acclaim PepMap 100 C18 HPLC Columns, 3µm particle size, 2cm length, 75 µm I.D.,

(Dionex, Sunnyvale, CA USA) and separated in Nano Series™ Standard Columns 75 µm I.D. x 15 cm, packed with C18 PepMap100, 3 µm, 100Å (Dionex, Sunnyvale, CA USA). A full MS scan was performed (m/z 360-1600) with further HCD MS/MS scans of the 20 most abundant ions with dynamic exclusion setting 15S. Spectra were acquired for 56 mins.

Uniprot database using Protein Discovery 2.2 software, Sequest HT algorithm (Thermo Fisher) was used to search against the MS and MS/MS scans. Deamidation (N and Q), oxidation (M), phosphorylation (S, T and Y) and acetylation (K) comprised the variable modifications. Two missed cleavages were permitted with data filtered using a false discovery rate (FDR) set to 0.01.

5.6 References

[1] S. Alatab *et al.*, "The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *Lancet Gastroenterol Hepatol*, vol. 5, no. 1, pp. 17–30, Jan. 2020, doi: 10.1016/S2468-1253(19)30333-4.

[2] Y. Ye, Z. Pang, W. Chen, S. Ju, and C. Zhou, "The epidemiology and risk factors of inflammatory bowel disease.," *Int J Clin Exp Med*, vol. 8, no. 12, pp. 22529–42, 2015.

[3] Y.-Z. Zhang, "Inflammatory bowel disease: Pathogenesis," *World J Gastroenterol*, vol. 20, no. 1, p. 91, 2014, doi: 10.3748/wjg.v20.i1.91.

[4] I. J. Pence *et al.*, "Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy," *Biomed Opt Express*, vol. 8, no. 2, p. 524, 2017, doi: 10.1364/boe.8.000524.

[5] T. Kirchberger-Tolstik *et al.*, "Towards an Interpretable Classifier for Characterization of Endoscopic Mayo Scores in Ulcerative Colitis Using Raman Spectroscopy," *Anal Chem*, vol. 92, no. 20, pp. 13776–13784, 2020, doi: 10.1021/acs.analchem.0c02163.

[6] S. Nikolaus and S. Schreiber, "Diagnostics of Inflammatory Bowel Disease," *Gastroenterology*, vol. 133, no. 5, pp. 1670–1689, Nov. 2007, doi: 10.1053/j.gastro.2007.09.001.

[7] S. Saeid Seyedian, F. Nokhostin, and M. Dargahi Malamir, “A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease,” *J Med Life*, vol. 12, no. 2, pp. 113–122, Apr. 2019, doi: 10.25122/jml-2018-0075.

[8] H. Goldberg, R. Gore, A. Margulis, A. Moss, and E. Baker, “Computed tomography in the evaluation of Crohn disease,” *American Journal of Roentgenology*, vol. 140, no. 2, pp. 277–282, Feb. 1983, doi: 10.2214/ajr.140.2.277.

[9] D. H. Bruining *et al.*, “Consensus Recommendations for Evaluation, Interpretation, and Utilization of Computed Tomography and Magnetic Resonance Enterography in Patients With Small Bowel Crohn’s Disease,” *Gastroenterology*, vol. 154, no. 4, pp. 1172–1194, Mar. 2018, doi: 10.1053/j.gastro.2017.11.274.

[10] J. P. Shoenut, R. C. Semelka, R. Silverman, C. S. Yaffe, and A. B. Micflikier, “Magnetic Resonance Imaging in Inflammatory Bowel Disease,” *J Clin Gastroenterol*, vol. 17, no. 1, pp. 73–78, Jul. 1993, doi: 10.1097/00004836-199307000-00018.

[11] A. Kornbluth, P. Legnani, and B. S. Lewis, “Video Capsule Endoscopy in Inflammatory Bowel Disease,” *Inflamm Bowel Dis*, vol. 10, no. 3, pp. 278–285, May 2004, doi: 10.1097/00054725-200405000-00017.

[12] R. Eliakim *et al.*, “Evaluation of a new pan-enteric video capsule endoscopy system in patients with suspected or established inflammatory bowel disease – feasibility study,” *Endosc Int Open*, vol. 06, no. 10, pp. E1235–E1246, Oct. 2018, doi: 10.1055/a-0677-170.

[13] S. Liu, Y. Zhu, P. Wu, and H. Xiong, “Highly Sensitive D–A–D-Type Near-Infrared Fluorescent Probe for Nitric Oxide Real-Time Imaging in Inflammatory Bowel Disease,” *Anal Chem*, vol. 93, no. 11, pp. 4975–4983, Mar. 2021, doi: 10.1021/acs.analchem.1c00281.

[14] I. J. Pence *et al.*, “Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy,” *Biomed Opt Express*, vol. 8, no. 2, p. 524, Feb. 2017, doi: 10.1364/BOE.8.000524.

[15] C. Bielecki *et al.*, “Classification of inflammatory bowel diseases by means of Raman spectroscopic imaging of epithelium cells,” *J Biomed Opt*, vol. 17, no. 7, p. 0760301, Jul. 2012, doi: 10.1117/1.JBO.17.7.076030.

[16] G. Acri *et al.*, “Raman Spectroscopy as Noninvasive Method of Diagnosis of Pediatric Onset Inflammatory Bowel Disease,” *Applied Sciences*, vol. 10, no. 19, p. 6974, Oct. 2020, doi: 10.3390/app10196974.

[17] F. Casellas *et al.*, “Fecal Excretion of Deoxyribonucleic Acid in Long-term Follow-up of Patients with Inactive Ulcerative Colitis,” *Inflamm Bowel Dis*, vol. 13, no. 4, pp. 386–390, Apr. 2007, doi: 10.1002/ibd.20042.

[18] F. Casellas *et al.*, “Fecal excretion of human deoxyribonucleic acid as an index of inflammatory activity in ulcerative colitis,” *Clinical Gastroenterology and Hepatology*, vol. 2, no. 8, pp. 683–689, Aug. 2004, doi: 10.1016/S1542-3565(04)00291-5.

[19] P. S. Karatzas, G. J. Mantzaris, M. Safioleas, and M. Gazouli, “DNA Methylation Profile of Genes Involved in Inflammation and Autoimmunity in Inflammatory Bowel Disease,” *Medicine*, vol. 93, no. 28, p. e309, Dec. 2014, doi: 10.1097/MD.0000000000000309.

[20] N. Patel *et al.*, “Metabolomic analysis of breath volatile organic compounds reveals unique breathprints in children with inflammatory bowel disease: a pilot study,” *Aliment Pharmacol Ther*, p. n/a-n/a, Jul. 2014, doi: 10.1111/apt.12861.

[21] L. C. Hicks *et al.*, “Analysis of Exhaled Breath Volatile Organic Compounds in Inflammatory Bowel Disease: A Pilot Study,” *J Crohns Colitis*, vol. 9, no. 9, pp. 731–737, Sep. 2015, doi: 10.1093/ecco-jcc/jjv102.

[22] A. A. Soubières, “Emerging role of novel biomarkers in the diagnosis of inflammatory bowel disease,” *World J Gastrointest Pharmacol Ther*, vol. 7, no. 1, p. 41, 2016, doi: 10.4292/wjgpt.v7.i1.41.

[23] A. Wang, C. Wang, M. Tu, and D. Wong, “Oral Biofluid Biomarker Research: Current Status and Emerging Frontiers,” *Diagnostics*, vol. 6, no. 4, p. 45, Dec. 2016, doi: 10.3390/diagnostics6040045.

[24] I. Khan *et al.*, “Alteration of Gut Microbiota in Inflammatory Bowel Disease (IBD): Cause or Consequence? IBD Treatment Targeting the Gut Microbiome,” *Pathogens*, vol. 8, no. 3, p. 126, Aug. 2019, doi: 10.3390/pathogens8030126.

[25] C. Tefas *et al.*, “Surface-enhanced Raman scattering for the diagnosis of ulcerative colitis: will it change the rules of the game?,” *Anal Bioanal Chem*, vol. 413, no. 3, pp. 827–838, Jan. 2021, doi: 10.1007/s00216-020-03036-2.

[26] M. Wakai *et al.*, “Serum amyloid A is a better predictive biomarker of mucosal healing than C-reactive protein in ulcerative colitis in clinical remission,” *BMC Gastroenterol*, vol. 20, no. 1, p. 85, Dec. 2020, doi: 10.1186/s12876-020-01229-8.

[27] C. ROOZENDAAL and C. G. M. KALLENBERG, “Are anti-neutrophil cytoplasmic antibodies (ANCA) clinically useful in inflammatory bowel disease (IBD)?,” *Clin Exp Immunol*, vol. 116, no. 2, pp. 206–213, Dec. 2001, doi: 10.1046/j.1365-2249.1999.00905.x.

[28] E. Israeli, “Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease,” *Gut*, vol. 54, no. 9, pp. 1232–1236, Sep. 2005, doi: 10.1136/gut.2004.060228.

[29] R. Sherwood and N. Walsham, “Fecal calprotectin in inflammatory bowel disease,” *Clin Exp Gastroenterol*, p. 21, Jan. 2016, doi: 10.2147/CEG.S51902.

[30] S. Vermeire, G. van Assche, and P. Rutgeerts, “C-Reactive Protein as a Marker for Inflammatory Bowel Disease,” *Inflamm Bowel Dis*, vol. 10, no. 5, pp. 661–665, Sep. 2004, doi: 10.1097/00054725-200409000-00026.

[31] G. Rogler and T. Andus, “Cytokines in Inflammatory Bowel Disease,” *World J Surg*, vol. 22, no. 4, pp. 382–389, Apr. 1998, doi: 10.1007/s002689900401.

[32] W. Strober, I. Fuss, and P. Mannon, “The fundamental basis of inflammatory bowel disease,” *Journal of Clinical Investigation*, vol. 117, no. 3, pp. 514–521, Mar. 2007, doi: 10.1172/JCI30587.

[33] Q. Guan and J. Zhang, “Recent Advances: The Imbalance of Cytokines in the Pathogenesis of Inflammatory Bowel Disease,” *Mediators Inflamm*, vol. 2017, pp. 1–8, 2017, doi: 10.1155/2017/4810258.

[34] F. Fava, “Intestinal microbiota in inflammatory bowel disease: Friend of foe?,” *World J Gastroenterol*, vol. 17, no. 5, p. 557, 2011, doi: 10.3748/wjg.v17.i5.557.

[35] M. Mahler *et al.*, “PR3-ANCA: A promising biomarker for ulcerative colitis with extensive disease,” *Clinica Chimica Acta*, vol. 424, pp. 267–273, Sep. 2013, doi: 10.1016/j.cca.2013.06.005.

[36] J. Osredkar, T. Kurent, T. Fabjan, K. Kumer, E. Božnar Alič, and D. Drobne, “The comparison of the three assays for determination of fecal calprotectin in inflammatory bowel disease,” *Biochem Med (Zagreb)*, vol. 31, no. 2, pp. 295–301, Jun. 2021, doi: 10.11613/BM.2021.020707.

[37] G. Cangemi, A. Barabino, S. Barco, A. Parodi, S. Arrigo, and G. Melioli, “A Validated HPLC Method for the Monitoring of Thiopurine Metabolites in Whole Blood in Paediatric Patients with Inflammatory Bowel Disease,” *Int J Immunopathol Pharmacol*, vol. 25, no. 2, pp. 435–444, Apr. 2012, doi: 10.1177/039463201202500213.

[38] J. Piestansky *et al.*, “Profiling of Amino Acids in Urine Samples of Patients Suffering from Inflammatory Bowel Disease by Capillary Electrophoresis-Mass Spectrometry,” *Molecules*, vol. 24, no. 18, p. 3345, Sep. 2019, doi: 10.3390/molecules24183345.

[39] M. Cappello and G. C. Morreale, “The Role of Laboratory Tests in Crohn’s Disease,” *Clin Med Insights Gastroenterol*, vol. 9, p. CGast.S38203, Jan. 2016, doi: 10.4137/CGast.S38203.

[40] S. Hosseini, P. Vázquez-Villegas, M. Rito-Palomares, and S. O. Martinez-Chapa, “Advantages, Disadvantages and Modifications of Conventional ELISA,” 2018, pp. 67–115. doi: 10.1007/978-981-10-6766-2_5.

[41] C. Morasso *et al.*, “Raman Analysis Reveals Biochemical Differences in Plasma of Crohn’s Disease Patients,” *J Crohns Colitis*, vol. 14, no. 11, pp. 1572–1580, Nov. 2020, doi: 10.1093/ecco-jcc/jjaa080.

[42] M. Veenstra *et al.*, “Raman Spectroscopy in the Diagnosis of Ulcerative Colitis,” *European Journal of Pediatric Surgery*, vol. 25, no. 01, pp. 56–59, Aug. 2014, doi: 10.1055/s-0034-1387951.

[43] I. J. Pence *et al.*, “Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy,” *Biomed Opt Express*, vol. 8, no. 2, p. 524, Feb. 2017, doi: 10.1364/BOE.8.000524.

[44] Y. Wu *et al.*, “The Specific Changes of Urine Raman Spectra Can Serve as Novel Diagnostic Tools for Disease Characteristics in Patients with Crohn’s Disease,” *J Inflamm Res*, vol. Volume 15, pp. 897–910, Feb. 2022, doi: 10.2147/JIR.S341871.

[45] X. Bi, A. Walsh, A. Mahadevan-Jansen, and A. Herline, “Development of Spectral Markers for the Discrimination of Ulcerative Colitis and Crohn’s Disease Using Raman Spectroscopy,” *Dis Colon Rectum*, vol. 54, no. 1, pp. 48–53, Jan. 2011, doi: 10.1007/DCR.0b013e3181fcf68d.

[46] N. Stone and P. Matousek, “Advanced Transmission Raman Spectroscopy: A Promising Tool for Breast Disease Diagnosis,” *Cancer Res*, vol. 68, no. 11, pp. 4424–4430, Jun. 2008, doi: 10.1158/0008-5472.CAN-07-6557.

[47] L. Wang *et al.*, “Raman spectroscopy, a potential tool in diagnosis and prognosis of castration-resistant prostate cancer,” *J Biomed Opt*, vol. 18, no. 8, p. 087001, Aug. 2013, doi: 10.1117/1.JBO.18.8.087001.

[48] C. Kendall *et al.*, “Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett’s oesophagus,” *J Pathol*, vol. 200, no. 5, pp. 602–609, Aug. 2003, doi: 10.1002/path.1376.

[49] M. Sbroscia *et al.*, “Thyroid cancer diagnosis by Raman spectroscopy,” *Sci Rep*, vol. 10, no. 1, p. 13342, Aug. 2020, doi: 10.1038/s41598-020-70165-0.

[50] E. Ryzhikova *et al.*, “Raman spectroscopy of blood serum for Alzheimer’s disease diagnostics: specificity relative to other types of dementia,” *J Biophotonics*, vol. 8, no. 7, pp. 584–596, Jul. 2015, doi: 10.1002/jbio.201400060.

[51] A. A. Bunaciu, H. Y. Aboul-Enein, and V. D. Hoang, “Raman Spectroscopy for Protein Analysis,” *Appl Spectrosc Rev*, vol. 50, no. 5, pp. 377–386, May 2015, doi: 10.1080/05704928.2014.990463.

[52] N. Kuhar, S. Sil, and S. Umapathy, “Potential of Raman spectroscopic techniques to study proteins,” *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 258, p. 119712, Sep. 2021, doi: 10.1016/j.saa.2021.119712.

[53] J. R. Beattie, S. E. J. Bell, and B. W. Moss, “A critical evaluation of Raman spectroscopy for the analysis of lipids: Fatty acid methyl esters,” *Lipids*, vol. 39, no. 5, pp. 407–419, May 2004, doi: 10.1007/s11745-004-1245-z.

[54] G. Zhu, X. Zhu, Q. Fan, and X. Wan, “Raman spectra of amino acids and their aqueous solutions,” *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 78, no. 3, pp. 1187–1195, Mar. 2011, doi: 10.1016/j.saa.2010.12.079.

[55] I. Maitra, C. L. M. Morais, K. M. G. Lima, K. M. Ashton, R. S. Date, and F. L. Martin, “Raman spectral discrimination in human liquid biopsies of oesophageal transformation to adenocarcinoma,” *J Biophotonics*, vol. 13, no. 3, Mar. 2020, doi: 10.1002/jbio.201960132.

[56] C. Banbury *et al.*, “Development of the Self Optimising Kohonen Index Network (SKiNET) for Raman Spectroscopy Based Detection of Anatomical Eye Tissue,” *Sci Rep*, vol. 9, no. 1, p. 10812, Dec. 2019, doi: 10.1038/s41598-019-47205-5.

[57] P. Chen *et al.*, “Serum Biomarkers for Inflammatory Bowel Disease,” *Front Med (Lausanne)*, vol. 7, Apr. 2020, doi: 10.3389/fmed.2020.00123.

[58] D. Liu, V. Saikam, K. A. Skrada, D. Merlin, and S. S. Iyer, “Inflammatory bowel disease biomarkers,” *Med Res Rev*, vol. 42, no. 5, pp. 1856–1887, Sep. 2022, doi: 10.1002/med.21893.

[59] Y. Fengming and W. Jianbing, “Biomarkers of Inflammatory Bowel Disease,” *Dis Markers*, vol. 2014, pp. 1–11, 2014, doi: 10.1155/2014/710915.

[60] B. E. Sands, “Biomarkers of Inflammation in Inflammatory Bowel Disease,” *Gastroenterology*, vol. 149, no. 5, pp. 1275-1285.e2, Oct. 2015, doi: 10.1053/j.gastro.2015.07.003.

[61] A. A. Almousa, M. Morris, S. Fowler, J. Jones, and J. Alcorn, “Elevation of serum pyruvate kinase M2 (PKM2) in IBD and its relationship to IBD indices,” *Clin Biochem*, vol. 53, pp. 19–24, Mar. 2018, doi: 10.1016/j.clinbiochem.2017.12.007.

[62] S. Tilakaratne, D. A. Lemberg, S. T. Leach, and A. S. Day, “C-Reactive Protein and Disease Activity in Children with Crohn’s Disease,” *Dig Dis Sci*, vol. 55, no. 1, pp. 131–136, Jan. 2010, doi: 10.1007/s10620-009-1017-8.

[63] C. A. Solem, E. v Loftus, W. J. Tremaine, W. S. Harmsen, A. R. Zinsmeister, and W. J. Sandborn, “Correlation of C-Reactive Protein With Clinical, Endoscopic, Histologic, and Radiographic Activity in Inflammatory Bowel Disease,” *Inflamm Bowel Dis*, vol. 11, no. 8, pp. 707–712, Aug. 2005, doi: 10.1097/01.MIB.0000173271.18319.53.

[64] S. Karoui *et al.*, “Correlation between levels of C-reactive protein and clinical activity in Crohn’s disease,” *Digestive and Liver Disease*, vol. 39, no. 11, pp. 1006–1010, Nov. 2007, doi: 10.1016/j.dld.2007.06.015.

[65] T. Nozoe, T. Matsumata, M. Kitamura, and K. Sugimachi, “Significance of preoperative elevation of serum C-reactive protein as an indicator for prognosis in colorectal cancer,” *The American Journal of Surgery*, vol. 176, no. 4, pp. 335–338, Oct. 1998, doi: 10.1016/S0002-9610(98)00204-9.

[66] R. Sierra *et al.*, “C-reactive protein used as an early indicator of infection in patients with systemic inflammatory response syndrome,” *Intensive Care Med*, vol. 30, no. 11, pp. 2038–2045, Nov. 2004, doi: 10.1007/s00134-004-2434-y.

[67] J. Cooper, Y. Pastorello, and M. Slevin, “A meta-analysis investigating the relationship between inflammation in autoimmune disease, elevated CRP, and the risk of dementia,” *Front Immunol*, vol. 14, Jan. 2023, doi: 10.3389/fimmu.2023.1087571.

[68] R. J. Bisoendial, S. M. Boekholdt, M. Vergeer, E. S. G. Stroes, and J. J. P. Kastelein, “C-reactive protein is a mediator of cardiovascular disease,” *Eur Heart J*, vol. 31, no. 17, pp. 2087–2091, Sep. 2010, doi: 10.1093/eurheartj/ehq238.

CHAPTER 6

RAMAN SPECTROSCOPY FOR THE MOLECULAR CHARACTERISATION, DIAGNOSIS, MONITORING AND BIOMARKER CLASSIFICATION OF CARDIOVASCULAR DISEASE

6.1 Abstract

Cardiovascular diseases (CVD) remain a major global health concern and a leading cause of morbidity and mortality worldwide. Risk factors such as high blood pressure, elevated cholesterol levels and obesity contribute to the development of CVDs with the current prevention and management strategies including lifestyle modifications, medication and surgical interventions. Early diagnosis and prompt medical attention are crucial in managing CVDs and reducing their overall impact on health and wellbeing and therefore, necessitating the development of innovative diagnostic and monitoring tools which transcend traditional methodologies. Raman spectroscopy has the ability to provide molecular fingerprinting and structural information thus offering insights into chemical composition, molecular structure, and even dynamic processes in various biomaterials. Herein, the integration of RS with advanced machine learning is established as a powerful clinical adjunct for point of care detection of cardiovascular diseases. A non-invasive, label-free miniaturised RS platform coupled with neural network algorithm of self-organising Kohonen-index network, *i.e.,* 'SKiNET', has been developed to accurately detect and assess molecular changes within blood plasma of CVD patients compared to healthy cohorts, enabling rapid CVD diagnosis and *in-situ* longer term monitoring, where the RS real-time capabilities provide dynamic assessment of disease progression, aligning treatment strategies with evolving diseased states. Subsequently, CVDs have been successfully detected from blood plasma and classified *via* SkiNET with 88.6% accuracy, 92.9% specificity and 85.1% sensitivity distinguishing these from healthy controls and with accuracy of 83.8%, sensitivity of 83.3% and specificity of 84.3% from the follow-up (3 months post cardiac event) blood plasma samples. Furthermore, the hybrid RS-SkiNET molecularly specific

detection signposted a comprehensive panel of CVD-indicative biomarkers, including SIL-6, IL-9, LpA, ApoB, PCSK9 and NT-ProBNP, offering important insights into disease mechanisms and risk stratification. Ultimately, this multidimensional approach holds significant potential for both improved patientcare and healthcare management for CVDs and other detrimental diseases.

6.2 Introduction

Cardiovascular disease (CVD) describes a group of disorders of the heart and blood vessels and are one of the leading causes of death globally. More than half a billion people around the world continue to be affected by CVD, with 20.5 million deaths associated with the disease in 2021, representing 32% of all deaths worldwide [1]. CVDs are caused by a combination of socioeconomic, metabolomic, behavioural and environmental risk factors and up to 80% of the premature heart attacks and strokes could be prevented by addressing the major contributors such as high blood pressure, obesity, tobacco use and physical inactivity [1], [2]. Addressing risk factors directly associated with the development of atherosclerosis is of a vital importance since these are central in the pathophysiology of CVD. CVD, arising from various aetiologies, can for instance, lead to ischemic stroke in patients with atrial fibrillation with an embolus. Often referred to as a “silent killer,” CVD is challenging to detect and manage in a timely manner and its often remains undiagnosed until it reaches advanced stages or causes significant health issues [3]. CVD, unlike many other health conditions lacks the apparent warning signs. Symptoms such as angina or shortness of breath often only present in the latter stages of the disease and frequently, a potential fatal cardiac

event, *e.g.,* heart attack or stroke, is the primary indicator of the underlying cardiovascular disease.

Whilst clinicians can assess a patient's risk in the primary care setting through evaluation of body mass index (BMI) and blood pressure, to obtain more detailed information and make an informed CVD diagnosis, broad blood panels of cholesterol and plasma ceramide levels are required accompanied by echocardiogram, echocardiogram and chest x-ray analyses [3], [4], [5]. Such evaluation for the diagnosis of CVD requires specialised instrumentation and skilled medical professionals and can only be made in hospital environments or dedicated medical centres. Other diagnostic methods including the angiography and cardiac catheterisation are invasive and painful procedures associated with a risk of infection and discomfort for patients. Whilst plasma myoglobin (Mb) levels are commonly used to indicate the presence of myocardial damage and are therefore used to ubiquitously diagnose heart attacks or myocardial infarction (MI), presently applied methods of Mb detection are broadly unaffordable in and suffer from poor time to results delays [6]. Furthermore, an evident health disparity is presenting a significant challenge in underprivileged or remote regions where there is the unmet need for advanced facilities and support. With four in every five CVD deaths occurring in low- or middle-income countries (LMIC) and staggering inequalities in CVD outcomes in LMICs, the global progress toward achieving equitable access to healthcare and wellbeing, distribution of prevention, diagnosis and treatment of CVDs is significantly lagging behind [1]. Collectively, with the growing burden of CVDs on society, there is an urgent unmet medical need for developing targeted techniques for CVD diagnosing and pre-disease evaluation

worldwide, which are non-invasive, reliable, sensitive, specific and rapid - integral for improving the overall patient care and clinical outcomes *via* pre-hospital timely diagnosis.

Despite the knowledge available to mitigate the harm to cardiovascular health, early-stage identification of disease remains a significant challenge. Indicative biomarkers can provide a powerful approach in the screening, diagnosis and monitoring of CVDs. Although disease-specific biomarkers such as troponin and D-dimer are currently used in the clinical laboratory, they rely on assays which take >60 minutes to provide outputs [6], [7], [8]. Furthermore, troponin is associated with a range of conditions such as heart inflammation, a blocked coronary artery and necrosis when found at elevated levels in the blood [9] and therefore, further testing is subsequently required to determine the underlying cause of illness. Concurrently, Raman scattering (RS) has been emerging as a non-invasive and label-free spectroscopy for clinical diagnosis of various cancer types, neurodegenerative disorders as well as for chemical detection of analytes [7], [10], [11], [12], [13], [14], [15], [16], positioning as a valuable technique for non-destructive, highly specific and rapid analysis of biofluids, offering detailed molecular information about the biochemical composition and structural molecular properties [17], [18], [19]. Raman spectroscopy, generating unique fingerprints of target molecules *via* an inelastic light scattering [20], [21] requires little to no sample preparation, allowing the measurement of biofluids in their native state as well as being unaffected by the presence of water or phosphate buffered saline, essential for cell/tissue analysis [20], [22]. These, combined with the remarkable ability of *real-time* measurements further enhanced by the added value of the portable and easy-to-use

medical device, cement its uses in clinical diagnosis, and could be incorporated into pre-hospital and hospital equipment lists and capability.

Here, we develop a hybrid RS-SKiNET approach and demonstrate the ability to detect early-stage biochemical changes indicative of CVD from blood plasma as a potential diagnostic adjunct. By applying the fused Raman-AI approach to characterise the presentation of CVD in a diverse patient population we successfully demonstrate disease discrimination *via* bimolecular spectral fingerprints of healthy and CVD patients. Further, we examined the ability of this technique to monitor therapeutic response of patients to treatments through meticulous evaluation of acute CVD clinical samples immediately following a cardiac event *versus* same patients' samples obtained three months post treatment for CVD. The acquired data is classified *via* the advanced artificial neural network algorithm, 'SKiNET,' as a decision support tool, demonstrating accuracy of 88.6±0.9%. With SKiNET (self-organising Kohonen index network), self-optimising map (SOM) is defined as a 2D map of neurons, organised as a grid of hexagons. A weight vector is randomly assigned to each individual neuron in the map and determines which neuron is activated for each sample. Neighbouring neurons having similar weightings lead to observable spatial clustering in each of the SOMs. Subsequently, self-organised discriminant index (SOMDI), by assigning class vectors as labels for each spectrum with corresponding weight vectors for each neuron, allows the identification of the data type responsible for activating a particular neuron. This is used to inspect weights across all neurons and allow the extraction of prominent spectral features belonging to each class. Importantly, *via* Raman-SkiNET we have profiled and identified a new panel of potential biomarkers of CVDs,

establishing unique characteristic barcode fingerprint for each, laying the platform for early CVD diagnosis at the point-of-care. This demonstrated capability of RS as a powerful analytical tool in recognising changes within complex biological mixtures [23], [24], [25], [26] is of particular benefit in CVD diagnostics since the differences between healthy and diseased samples reveal spectral biomarkers associated with biochemical changes in the early onset of CVD. Additionally, the inherent multiplexing ability enhances its utility in further CVD research and diagnostics by offering a comprehensive and efficient approach for rapidly detecting and characterising multiple biomarkers simultaneously. It can also also complement in disease subtyping by identifying distinct biomarker profiles associated with different subtypes of CVD *e.g.,* MI, heart failure or atherosclerosis and analyse numerous samples with high throughput, all collectively of a great value for large-scale biomarker screening. The overall developed biomolecular profiling methodology opens new avenues for disease discrimination, monitoring and biomarker discovery, towards enabling timely diagnosis and advanced understanding of biomolecular changes in context of the complex aetiology of CVDs.

6.3 Results and Discussion

A total of 60 CVD and 40 healthy control samples were collected with the main characteristics of each group summarised in **Table 6.1.** Subsequently**,** Raman spectral fingerprints were acquired to study the biochemical composition of the plasma samples to differentiate and classify CVD from healthy patients. The overall workflow protocol of RS-SkiNET process included blood collection from patients within first hour of a known cardiac event, sample pre-processing *via* centrifugation and deposition onto

aluminium slide, *in-situ* spectroscopic profiling followed by SkiNET classification with further molecular characterisation *via* a combination of multivariate and statistical analyses (**Fig 6.1**).

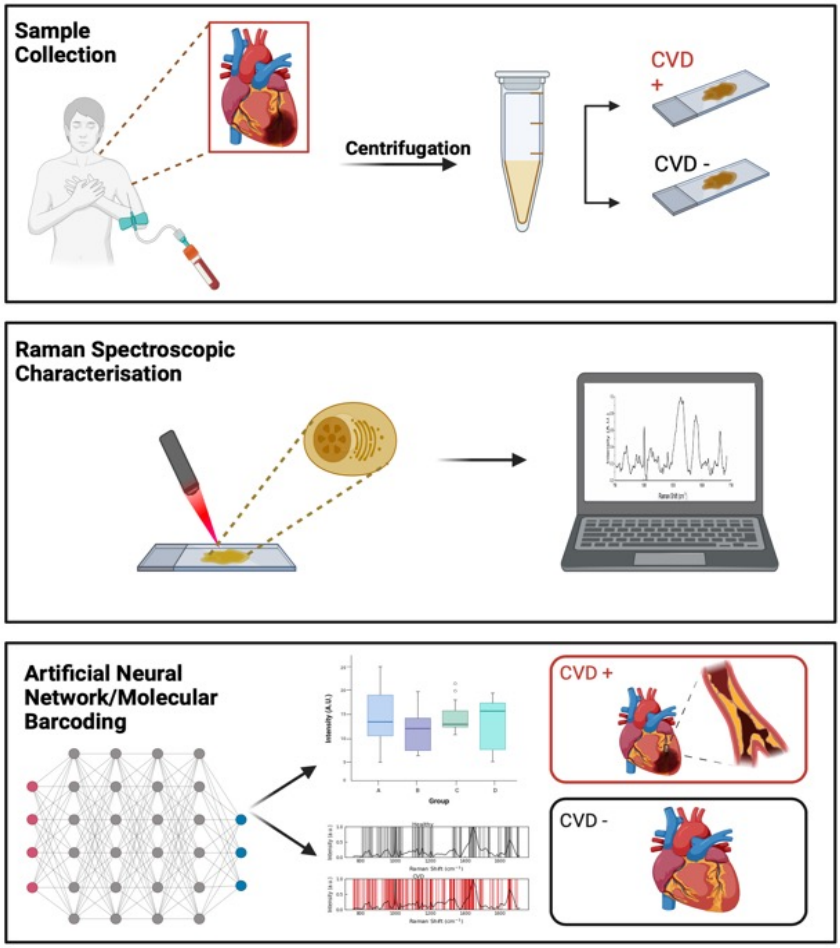

Figure 6.1 - Schematic representation illustrating workflow of the analysis. After collection of patient blood samples, the sample is sequentially centrifuged to obtain blood plasma where it is then deposited on an aluminium foil covered glass slide. Plasma is then measured using Raman spectroscopy and subsequently classified using an artificial neural network, SKiNET with further molecular characterisation via statistical methods including Student's t-test. (Figure created in Biorendor.com)

Since RS provides a unique spectral fingerprint, each spectrum is molecularly specific and contains characteristic bands allowing an unambiguous identification. To effectively detect representative changes within biofluids, it is essential to initially

acquire the spectral signatures of the baseline healthy samples. This subsequently, enables the recognition and attribution of the significant bands found within the sample, which are then be used to identify and discriminate disease within patient samples. Since fluorescence is known to interfere with Raman measurements of biological samples, to obtain high signal to noise spectra, the acquisition parameters were optimised and established as 785 nm excitation laser, 50x objective, 100 map scan spectra per each clinical sample within a 10x10 grid with 5 μm increments. The average Raman spectra acquired in the fingerprint region of 700-1750 cm^{-1} of CVD (red) and healthy (black) blood plasma exhibited similar molecular characteristics typical of biological specimens including for instance, phenylalanine (1003 cm^{-1}), Amide III (1230-1285 cm^{-1}) and CH deformation of proteins and lipids (1445 cm^{-1}) bands (**Fig 6.2**). Additional primary peaks were detected at Raman shifts of 960, 1156, 1206, 1315, 1520, 1615 and 1655cm^{-1} with each band being significantly different (p^{**}<0.001, p^{***}<0.0001) between the studied sample types (**Table 6.1**). Although similar patterns were observed with the main bands present in both groups, there were clear differences between their intensities, reflecting variations between the biochemical composition of blood plasma in patients immediately after a cardiac episode and healthy controls.

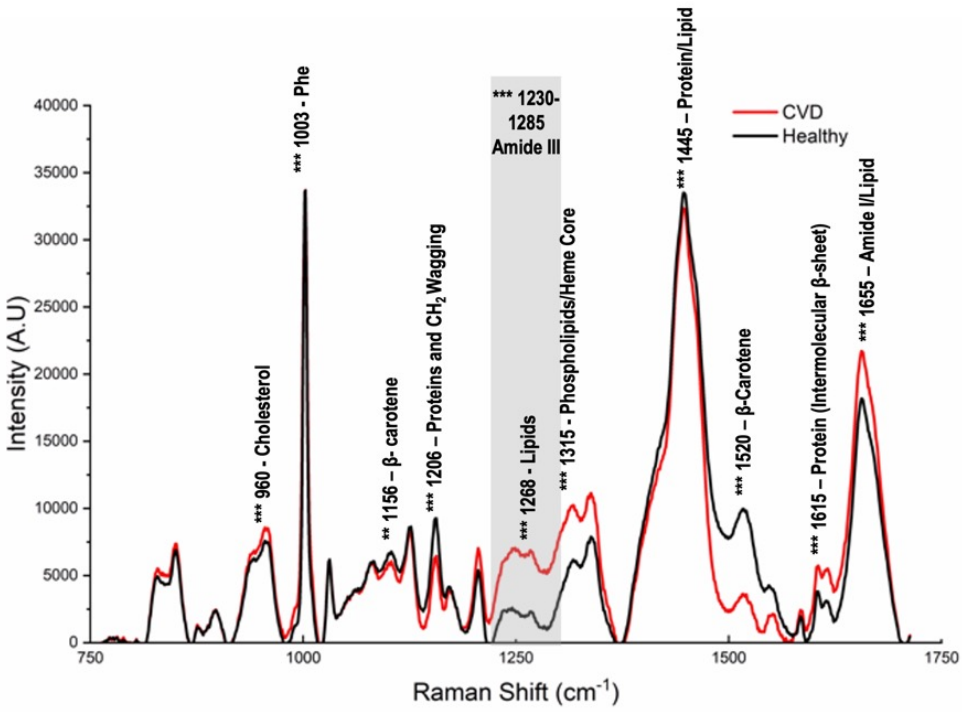

Figure 6.2. Representative average Raman spectra of CVD (red) and healthy (black) blood plasma with the designated significant identified changes (p**<0.001, p***<0.0001).

Bands at 960 cm^{-1} (cholesterol), 1230-1285 cm^{-1} (Amide III/lipids), 1315 cm^{-1} (Phospholipids/Heme Core), 1615 cm^{-1} (Protein/tyrosine Intermolecular β-sheet) and 1655 cm^{-1} (Amide I/Lipid) were identified at increased intensities in CVD blood plasma while bands at 1156/1520 cm^{-1} (β-Carotene), 1206 cm^{-1} (Proteins/tryptophan and CH_2 wagging) and 1445 cm^{-1} (Protein) were observed at increased intensities in healthy plasma cohorts.

Table 6.1. Tentative Raman assignment of significant peaks of interest between CVD and healthy patient samples.

Shift (cm^{-1})	Vibrational Mode(s)	Assignment	Tendency (Disease v	*p*-value

			Healthy Control)	
960	C-C backbone vibration	Cholesterol	Increased	<0.0001
1003	Aromatic δ ring mode	Phenylalanine	Increased	<0.0001
1156	Bond stretching (ν) C-C	β-Carotene	Decreased	<0.001
1206	Amide II, CH_2 wagging vibrations	Proteins	Decreased	<0.0001
1268	=C-H in plane deformation vibrations	Lipids	Increased	<0.0001
1315	τ(CH2), ν21-δas(CmH)	Phospholipids/Heme Core	Increased	<0.0001
1340	Aromatic ring mode/CH deformation	Tryptophan	Increased	<0.0001
1445	CH_2 bend	Protein/Lipid	Decreased	<0.0001
1520	NH_3-sym bend	β-Carotene	Decreased	<0.0001
1615	Aromatic ring mode	Protein (Intermolecular β-Sheet)	Increased	<0.0001
1665	Amide I, C=C stretch	Amide I/Lipid	Increased	<0.0001

RS further provides a comprehensive overview of the overall plasma composition with the main spectral variations related to lipid/cholesterol peaks. Lipids and lipoproteins are known to have a major impact on the genesis and progression of CVD within their cellular assembly, degradation, synthesis and plasma concentrations. Classified according to their function as either storage lipids, including fatty acids and triglycerides or structural lipids such as phospholipids, they are essential for various physiological processes. High plasma concentrations of lipids and lipoproteins promote thrombus

and plaque formation. Lipids and lipoproteins, together with their transport and metabolism, are significant contributing factors of CVD since they regulate plasma cholesterol concentration, which enhances the uptake of cholesterol by macrophages and subsequently, leads to foam cell formation, ultimately resulting in the formation of plaques and inflammation [27]. The observable variation in the Raman spectra also correlates with the patient demographics whereby ~40% of CVD patients had known hypercholesterolaemia prior to their cardiac episode. Due to the involvement of lipid and lipoprotein imbalances in the development and advancement of CVD, there is a pressing need to classify them according to their mechanistic impact on the disease in the pursuit of novel approaches to effectively predict and treat CVDs [28]. Through correlation of the identified intensity ratio changes at 960/1315 bands, which was found to be ×8 higher in CVD than in healthy blood plasma, representative of cholesterol and total lipids, it is possible to rapidly detect increased levels of both cholesterol and lipids with potential to assess CVD risk.

Further to the molecular characterisation, which revealed significant differences between the spectral fingerprints of healthy and CVD plasma, an in-depth spectral classification was performed *via* artificial neural network (SKiNET) used as a decision support tool to automatically differentiate CVD and healthy controls. This rapid classification provides a robust, accurate, high-throughput and simple approach to categorising samples into their pre-defined classes from their characteristic spectra as well as allowing quicker (<2 mins) profiling and screening of clinical samples. This hybrid approach therefore makes it highly suitable for *real-time* medical diagnostics and decision making, leading to immediate interventions or treatments, of a particular

importance in a time-sensitive diagnosis such as the CVD. A total of 6,000 CVD and 4,000 healthy control spectra were spectroscopically profiled and analysed *via* SKiNET with the derived weight vectors characteristic to each class shown in **Fig 6.3a.** Spectral differences of healthy (black) and CVD (red) blood plasma were found to vary significantly with SOMDI scores indicating the intensity changes responsible for the observed spectral variations and subsequent classification and clustering in SOM (**Fig 6.3b**). Main differences were detected at shifts of 850, 960, 1003, 1156, 1206, 1230, 1268, 1315, 1355, 1445, 1520 and 1655 cm^{-1}. CVD plasma had an increased response in the Amide III (1230-1285 cm^{-1}), Amide I (1655 cm^{-1}), total lipids (1315 cm^{-1}) while healthy plasma had an increased response in β-carotene (1156/1520 cm^{-1}), phenylalanine (1003 cm^{-1}) and CH_2 deformation band (1445 cm^{-1}). Overall, SKiNET differentiated healthy and CVD plasma with an accuracy of 88.6±0.9%, sensitivity of 92.9±1.2%, specificity of 85.1±0.8%, positive predictive value (PPV) of 83.6±1.3% and negative predictive value of 93.6±1.2%.

Protein, lipid and amino acid attributed peaks were found to represent the main changes in the spectra with significant peaks identified in both the Amide I and Amide III regions. The amide band I observed between 1600 and 1700 cm^{-1} is attributed to the C=O stretching vibration and is directly related to the backbone conformation. Amide I is the most intense absorption band in proteins and the exact position of this band is determined by both the backbone confirmation and the pattern of hydrogen bonding. An increased Amide I in CVD patients indicates an increase in α-helix proteins. Recent studies have indicated that chronic inflammation and oxidative stress- both associated with the process of atherogenesis- are mechanistically connected to

conversion of proteins into *α*-helix forms [29]. Moreover, fibrin, a blood protein with *α*-helix structures plays a central role in blood clotting, forming at sites of vascular injury. Fibrin also serves as the framework for pathological obstructive thrombi, which occlude blood vessels, impede blood flow and result in myocardial infarctions and ischemic strokes [30]. The increase of fibrin present in the blood plasma after the cardiac event most likely contributes to this observed intensity change.

Amide III, however, is a rather complex band and results from a mixture of several different coordinate displacements. It is dependent on the nature of the side chains as well as the hydrogen bonding present. Within these banded regions, the presence of specific amino acid side chains can be recognised. These side chain vibrations are located between 1400 and 1800 cm^{-1} (amide I and amide II region). Of the 20 amino acids, 9 are particularly dominant in this region, aspartic acid (1574 cm^{-1}), glutamic acid (1560 cm^{-1}), arginine (1673 cm^{-1}), lysine (1629 cm^{-1}), asparagine (1678 cm^{-1}), glutamine (1670 cm^{-1}), tyrosine (1518 cm^{-1}), histidine (1596 cm^{-1}) and phenylalanine (1598 cm^{-1}). In CVD, localised inflammation as well as increased cell-to-cell communications could potentially affect the polarisation and proliferation of immune cells *via* alterations within the metabolism of amino acids [31]. Glutamate and glutamine have associated with atherosclerosis development as well as other cardio-metabolomic disorders [31], [32]. For example, in patients with cardiac catheterisation, these amino acids were used as the most significant discriminator between coronary artery disease (CAD) and non-CAD patients [33]. High glutamate levels were also associated with the incidence of coronary artery disease and was linked to the clinical

onset of atherosclerosis due to its increased risk of plaque development and increased intima-media thickness [34], [35], [36].

Furthermore, methionine paired with its metabolic intermediate homocysteine has also been found to promote atherosclerosis in both human and animal models with methionine-induced hyperhomocysteinemia, accelerating the initial development of plaques and enhancing further plaque fibrosis *via* numerous mechanisms, including increased lipid peroxidation and impaired antioxidant activity [37]. In addition, a study by Jauhiainen *et* al. investigated the significance of 9 different amino acids (alanine, glutamine, glycine, histidine, isoleucine, leucine, phenylalanine, tyrosine, and valine) as risk factors for CVD events [38]. They determined that phenylalanine and tyrosine were significantly associated with an increased risk of CVD while alanine was nominally associated with CAD events. Tyrosine, a nonessential amino acid, and a precursor of the catecholamines has previously been linked with an increase of carotid intima-media thickness while alanine has been associated with elevated blood pressure, insulin resistance and type 2 diabetes, all of which are widely accepted risk factors for cardiometabolic diseases and CVD [39], [40], [41].

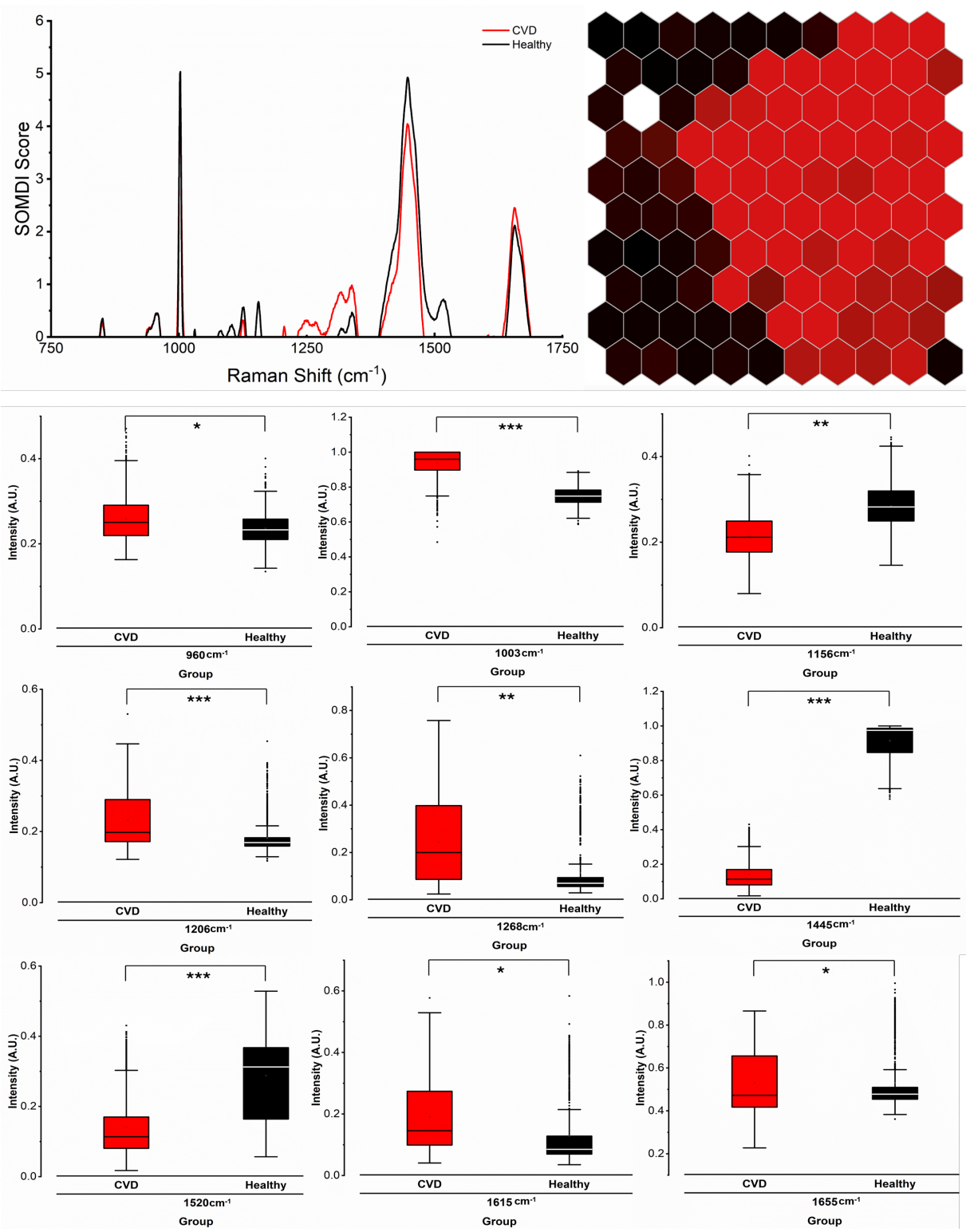

Figure 6.3. (a) Characteristic bands of importance identified *via* SOMDI SKiNET analysis (left) highlighting the most influential Raman bands responsible for the clustering observed in SOM (right) successfully classifying blood plasma according to the correct class of CVD (red) or healthy (black) with an accuracy of 88.6±0.9%. (b). Box and whisker plots represent the minima, maxima, interquartile

ranges, whiskers and the median in peak intensities identified via SOMDI as having the greatest effect on the classifier, *i.e.,* the spectral changes of greatest importance with the largest changes observed at 960, 1003, 1156, 1206, 1268, 1445, 1520, 1615 and 1655 cm^{-1} as identified through molecular characterisation and SKiNET classification with a 10-fold cross validation.

Further, there was a decrease in intensity of the peaks assigned to β-carotene in CVD patients. Widely available in plant-based foods, the efficacy of β-carotene in CVD risk is currently controversial. Many studies e.g., by Rapola et al. and Omen et al. reported no benefit of β-carotene and only a slight adverse effect on the incidence of CVD, whereas Lee et al. and Cook et al. reported no overall effects of β-carotene on CVD events amongst apparently healthy women and those at high risk for CVD [42], [43], [44], [45].

To identify specific RS peaks correlating with the different stages of CVD, we further studied the spectral changes after three months following therapeutic interventions including aspirin, clopidogrel and ticagrelor and compared these to patient samples immediately after a cardiac event. 2000 CVD, matched with further 2000 follow-up and 2500 healthy control cohort spectra were acquired and abolished (**Fig 6.4a**). Visually observable primary peaks of lipids arising from the stretching of various CH_2 and CH_3 groups (1150-1268 cm^{-1}) and of proteins including contributions from the amide groups of the secondary protein structures (α-helix and β-sheets), aromatic and branched amino acids such as valine, leucine, phenylalanine, tryptophan and tyrosine (1003, 1340 and 1600 - 1650 cm^{-1}) and stretching or deformation of CH, C=C and NH_3 (1600-1650 cm^{-1}) were identified. Spectral analysis revealed that across the majority of the follow-up patients' plasma after three months of therapeutic interventions returned to

similar levels observed in the average spectra from healthy control cohort plasma. Interestingly, peaks attributed to cholesterol (960 cm^{-1}) and lipids (1268/1315 cm^{-1}) exhibited significant decreased in intensity in line with those obtained from healthy control plasma. Differences between healthy and follow-up plasma were largely confined to CH_2 bending of proteins between Raman shifts of 1450-1500cm^{-1}. To further confirm the significance of these spectral markers in the definition of the healthy, diseased and therapeutic response stages of CVD as well as to define the sensitivity and efficiency of the approach, hybrid RS-SKINET (**Fig 6.4b-g**) revealed 77.3±1.9% accuracy differentiating healthy, CVD and follow-up plasma, 65.9±1.5% accuracy differentiating healthy and follow-up plasma and 83.8±1.2% accuracy distinguishing CVD from follow-up plasma (**Table 6.2**). RS-SKiNET classification indicated a higher degree of sample crossover between healthy and follow-up patients, achieving a relatively poor classification accuracy and thus, indicating a higher degree of similarity between these cohorts. Whereas a comparison between patient matched CVD and follow-up plasma yielded a higher classification accuracy, revealing clearer spectral clustering with significantly lower sample crossover.

Furthermore, to provide a comprehensive assessment of the diagnostic accuracy of the classification we used receiver operating characteristic (ROC) curves to assess the differentiation of healthy and CVD blood plasma by calculating the area under the curve (AUC), plotting the true positive versus the false negative rates (**Fig 6.5**) alongside bow and whisker plots determining the significance of the most influential peaks highlighted via SKiNET and SOMDI classification (**Fig 6.5c-d)**. From the ROC curves we determined the intrinsic distinguishing between individuals with CVD and

those follow-up patients, with AUC of 0.94 (95% CI 0.86-1.03), indicating high levels of discrimination between the sample types. The ROC's steep ascent with the consistent trajectory toward the upper-left corner, indicating a minimal rate of false positives while maximising true positive predictions of the classification. This is critical when assessing diagnostic or predictive models, due to the indication of the ability to accurately detect relevant disease cases and states. This high discriminative power underscores the utility of the developed classifier in clinical practice, where the rapid measurement of patient response to therapeutics is of paramount importance. Contrastingly, when investigating the performance of the classifier to distinguish between healthy plasma and of CVD patients three months post therapeutic interventions, the ROC curve indicated limited discrimination ability. With an AUC of only 0.54 (95% CI 0.37-0.70), the classifier's ability to discriminate between the two groups is insignificant, indicating a considerably higher degree of similarity and crossover between healthy and patients on therapeutics post cardiac event. These findings may indicate the high effectiveness of the medication in mitigating the molecular or biochemical changes associated with CVD and therefore, the spectra closely match those of healthy controls. It further can be an indication of biomarkers' stabilisation, where the therapeutics stabilised, or reversed disease-specific biomarkers or molecular alterations originally present in CVD plasma. In such cases, disease remission indicating that CVD patient's health has improved to a level comparable to healthy individual. There is also a possibility of limited disease progression, leading to a preservation of patient molecular profiles similar to healthy subjects. Further research and clinical assessment would be required to fully understand the clinical implication of these findings.

Table 6.2. CVD differentiation *via* classification performance of healthy and CVD patient plasma compared to healthy controls including sensitivity, specificity, accuracy, positive predictive values and negative predictive values.

Comparison	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
Healthy *vs.* Follow-up	61.1 ± 2.1	56.6 ± 1.3	54.0 ± 1.6	63.4 ±1.8	65.9 ± 1.5
CVD *vs.* Follow-up	83.3 ± 0.8	84.3 ± 0.9	84.5 ± 1.3	83.1 ± 0.9	83.8 ± 1.2

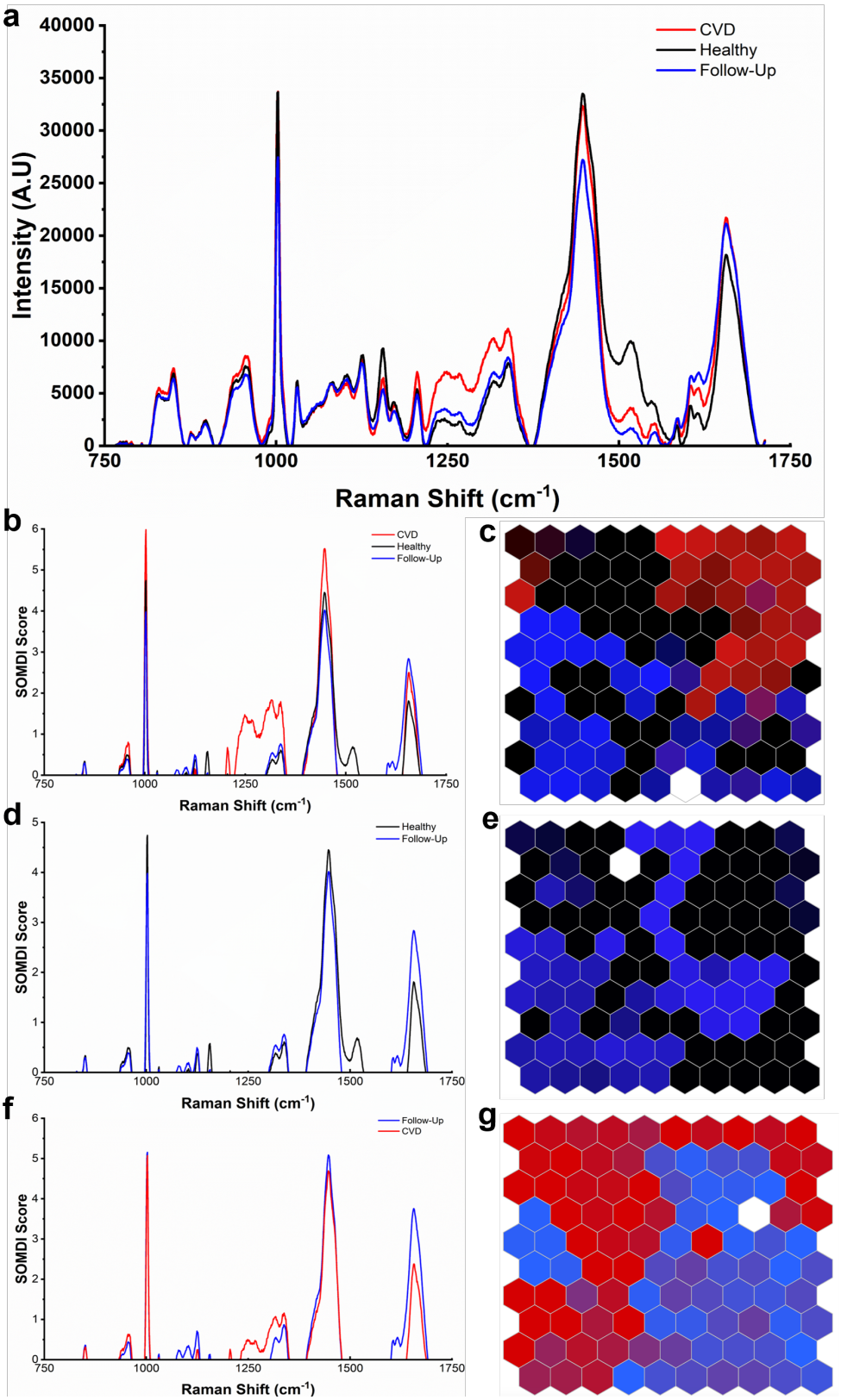

Figure 6.4. (a). Representative average Raman spectra of healthy control (black), CVD (red) and patient matched three-month post therapy follow-up (blue) blood plasma. **(b)** SOMDI and **(c)** SOM of healthy,

CVD and follow-up blood plasma, **(d)** SOMDI and **(e)** SOM of healthy and follow-up blood plasma and **(f)** SOMDI and **(g)** SOM of CVD and follow-up blood plasma, classifying the samples according to diseased state with an accuracy of 77.3±1.9%, 65.9±1.5% and 83.8±1.2% respectively.

This suggests patient plasma, three months post cardiac event has significantly lower levels of cholesterol and lipids, most likely due to the positive effects of therapeutics. The main peaks of significance were identified in the Amide I band region (1600-1655cm^{-1}) indicating differences within secondary protein structures.

Interestingly, classifying the significant spectral shifts with the greatest effect on the classifier *via* box and whisker plots analysis, revealed the previously identified cholesterol (960 cm^{-1}) and lipid peaks (1200 -1340 cm^{-1}) as having the greatest influence on sample classification and subsequent greatest variation amongst CVD and healthy plasma, three month post therapeutic follow-up samples compared with healthy plasma were now found to be significant with only little effect on the classifier (**Fig 6.5c-d**).

Collectively, these findings suggest that whilst medication is highly beneficial in managing CVD and preventing further complications, complete reversal of all structural changes may not be achievable. Phenylalanine has long been associated with immune activation and inflammation. Previous studies, *e.g.,* by Chen *et* al., have shown that increased plasma concentrations of phenylalanine predict mortality in critical patients such as in those with heart failure [46]. Higher phenylalanine levels correlate with substantial inflammation and changes in cytokines, suggesting some level of immunodepression and malnutrition. The decrease in intensity of phenylalanine at

1003 cm^{-1} in follow-up plasma compared with CVD patients, therefore, indicates the ability of the therapeutics to manage certain aspects of CVD, such as reducing inflammation evidenced through a decreased in phenylalanine. However, this combined with the significant difference in secondary protein structure, also suggests these interventions do not reverse all molecular and structural changes which have already taken place at the time of the cardiac event.

Characterising complex biofluids *via* RS coupled with advanced AI provides a powerful tool for point-of-care detection of clinical diseases. Machine learning is particularly effective for categorising otherwise uninterpretable, patient-plasma derived spectra which contain a variety of low concentration, disease-specific molecular biomarkers within a hugely populated and cumbersome spectral background of biological molecules. Therefore, SKiNET in conjunction with RS molecular profiling has been applied to classify a panel of potential CVD-indicative biomarkers (**Table 6.3**). These candidate biomarkers were systematically evaluated and compared with the obtained spectral fingerprints of healthy and CVD blood plasma (**Fig 6.6**). The identified panel included apolipoprotein B (ApoB), lipoprotein A (LpA), proprotein convertase subtilisin/kexin type 9 (PSCK9), C-reactive protein (CRP), *d*-dimer, interleukin-9 (IL-9), myeloperoxidase (MPO), tumour necrosis factor-alpha (TNF-*α*), copeptin, galectin-3, growth and differentiation factor-15 (GDF-15), matrix metalloproteinase 9 (MMP9), N-terminal pro-*B*-type natriuretic peptide (NT-proBNP), suppression of tumorigenicity 2 (ST2), soluble interleukin-6 (SIL-6) and atrial natriuretic peptide (ANP), all chosen based on their specificity for CVD and known physiological response in patients.

Fig 6.6 shows the obtained mean spectral fingerprints for each biomarker signposting the most significant peaks previously determined *via* SKiNET in the classification of CVD, follow-up and healthy control plasma. The candidate biomarkers were split into three subgroups including cardiac lipids, those involved in inflammation and thrombosis and those associated with cardiac injury with each marker exhibiting specific Raman bands. When compared with the spectral fingerprints of healthy and CVD plasma all biomarkers, excluding copeptin and ST2, were identified as statistically significant with an increased tendency in CVD plasma. Atherosclerosis affecting the coronary arteries is primarily influenced by cholesterol, specifically low-density lipoprotein cholesterol (LDL-C). This is underscored by the success of LDL-C lowering medications such as, statins as well as Mendelian randomisation studies, which clearly demonstrate the causative role of LDL-C's in CVD [47], [48]. Recent studies have identified the significant role of PCSK9 in regulating LDL-C with levels of soluble PCSK9 emerging as a novel indicator of CVD risk [49]. While previously thought to play a protective role against CVD risk, high-density lipoprotein cholesterol (HDL-C) has recently been brought into question with clinical drugs trials aimed at raising HDL-C levels casting doubt on this assumption with evidence on the casual role, if any, of HDL-C on CVD risk remaining uncertain [47], [48]. Aside from their involvement in the pathogenesis of CVD, lipids and lipoprotein particles also influence inflammatory processes as well as the functions of vascular and cardiac cells, thus impacting both the vessels and heart [49], [50]. As inflammation plays a central role in the pathophysiology of CVDs, mediators of inflammation are therefore of interest as potential risk markers. Lipid-driven inflammation of the arterial wall characterised by the innate and adaptive immune response is a hallmark of coronary artery

atherosclerosis. The early development of an atherosclerotic plaque is dependent upon interactions between damaged smooth muscle cells, smooth muscle cells in the arterial wall, and circulating inflammatory cells [51], [52]. Following the rupture of an atherosclerotic plaque, complex communications between immune cells and platelets drive the formation of a thrombus and the onset of widespread inflammation. The developed thrombus subsequently can obstruct a coronary artery, potentially leading to myocardial ischaemia and thus, the necrotic cell death triggers a pro-inflammatory response of cytokines, including IL-6, IL-1β and TNF-*α* and CRP *via* reactive oxygen species (ROS) production, complement cascade activation and the release of damage-associated molecular patterns (DAMPs) such as high mobility group box 1 (HMGB1) and heat shock proteins [52], [53], resulting in elevated levels of these proteins due to systemic inflammation [54], [55]. Anti-inflammatory cytokines together with TGF-*β*, work to suppress the initial inflammatory response. Also released by neutrophils and macrophages are matrix metalloproteinase enzymes such as MMP9, known to breakdown components of the extracellular matrix [56], [57]. Tissue remodelling due to increased MMP activity is known to play a central role in inflammation and injury. MMPs in CVD are known to mediate the rupture of atherosclerotic plaques, the outcome of which may result in a chronic disease progressing to an acute MI. Interestingly, previous studies have indicated that patients who have suffered an MI or AF and present with elevated levels of GDF-15 may suffer from poorer outcomes, with this cytokine having been positively correlated with increased risk of cardiovascular events and as an independent predictor of mortality in patients with coronary artery disease and heart failure [56], [58].

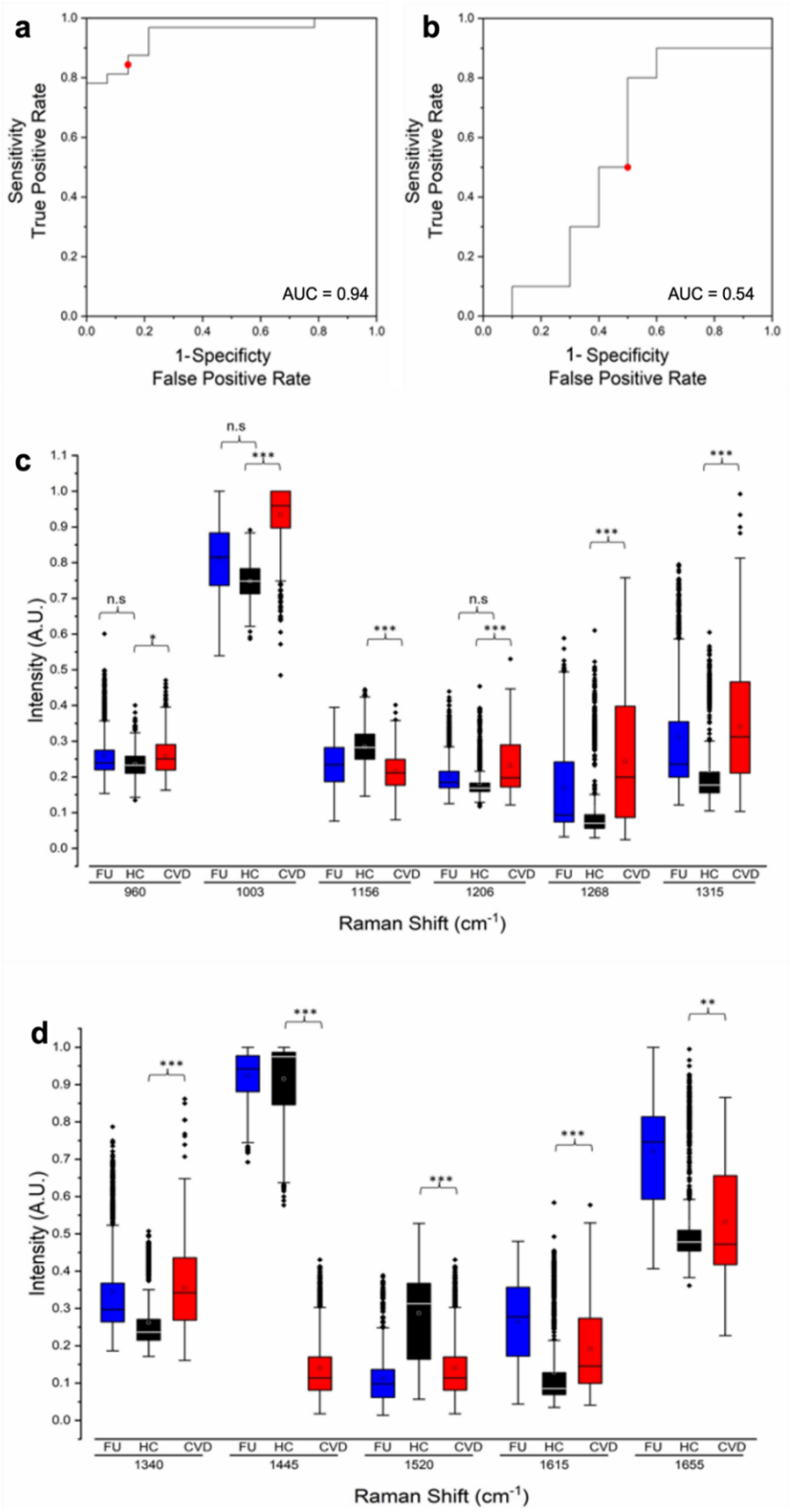

Figure 6.5. ROC plots representing the sensitivity *versus* 1-specificity derived from post probability assignment applied to determine the success for classifying the healthy and CVD-derived blood plasma

for **(a)** CVD and follow-up and **(b)** healthy and follow-up patients. The AUC indicates that the change in the protein and lipid levels of plasma are valuable markers to discriminate healthy from CVD-indicative blood plasma showing an excellent performance with AUC of 0.94 (95% CI 0.86-1.03) and only 0.54 (95% CI 0.37-0.70) for healthy compared to 3 month post cardiac event follow-up blood plasma, respectively. **(c-d).** Box and whisker plots represent the minima, maxima, interquartile ranges, whiskers and the median in peak intensities identified *via* SOMDI as having the greatest effect on the classifier, *i.e.,* the spectral changes of greatest importance with the largest changes observed at 960, 1003, 1156, 1206, 1268, 1315, 1340, 1445, 1520, 1615 and 1655cm^{-1}, which were further found to be statistically significant *via* the *t-test* evaluation (***p*<0.001, ****p*<0.0001).

Table 6.3. Potential biomarkers of CVD with their associated roles and tendency in CVD patients.

Biomarker	**Category**	**Physiological Range (ng/mL)**	**Released in Response To/Modified By**	**Role in CVD**	**Tendency (Disease v Healthy Control)**	**p-value**
Apolipoprotein B (48,49)	Lipid	5 x 10^5 to 15 x 10^5	Inflammation	Key structural protein of all major atherogenic lipoproteins; plays multiple roles in regulating lipid metabolism	Increased	<0.001
Lipoprotein A (50,51)	Lipid	5 x 10^5 to 15 x 10^5	Inflammation	Involved in foam cell formation, smooth muscle cell proliferation, plaque inflammation and is the main carrier of oxidised phospholipids considered proinflammatory and proatherogenic	Increased	<0.001
Proprotein convertase subtilisin/kexin type 9	Lipid	170 - 220	Influenced by use of statins	Circulating regulator of low-density lipoprotein cholesterol (LDL-C) via its ability to	Increased	<0.001

(PCSK9) (52,53)				induce degradation of the LDL receptor (LDLR) in the lysosome of hepatocytes		
C-Reactive Protein (54,55)	Protein	100 - 1000	Inflammation	Produced in response to circulating cytokines such as, interleukin-6 by hepatocytes. Robust marker of inflammation, however non-specific for CVD	Increased	<0.05
D-Dimer (56,57)	Fibrin degradation product	<250	Thrombosis, Inflammation	Indicator of thrombosis and resultant fibrin degradation	Increased	<0.001
Interleukin-9 (IL-9) (58,59)	Cytokine	0 – 0.044	Inflammation	Pro-inflammatory cytokine released in response to chronic inflammatory conditions	Increased	<0.001
Myeloperoxidase (MPO) (60,61)	Leukocyte-derived Enzyme	0 – 65.66	Inflammation	Catalyses the formation of H_2O_2 to reactive oxygen species, is an integral component of the innate immune response and contributes to tissue damage during inflammation	Increased	<0.05
Tumour necrosis factor-alpha (TNF-α) (62,63)	Cytokine	0.0247	Inflammation	Transducer of CVD, implicated in inflammation, cell survival, growth differentiation and apoptosis. Functions within cytokine cascade, inducing or supressing genes	Increased	<0.05
Copeptin (64,65)	Glycopeptide (C-terminal	0.24 – 1.58	Systemic inflammation, hemodyn	Co-released with arginine vasopressin (AVP) and acts as a marker of	-	N.S.

	portion of arginine-vasopressin)		amic stress	vasopressin secretion. Vasopressin is released in response to plasma osmolality		
Galectin-3 (66,67)	B-galactoside-binding protein	5.3-18.0	Inflammation, fibrosis, and angiogenesis	Pleiotropic regulatory role with several cellular functions such as cellular growth, proliferation, and tissue repair. Implicated in several diseases including organ fibrosis, cancer, and atherosclerosis	Increased	<0.01
Growth and differentiation factor 15 (GDF-15) (68,69)	Cytokine (Belongs to transforming growth factor (TGF-β) superfamily of proteins)	0.537 – 1.847	Inflammation, oxidative stress, hypoxia	Identified as an inflammatory marker with a cardioprotective role. Highly induced in cardiomyocytes after ischemia and in the heart after an MI	Increased	<0.05
Matrix metalloproteinase 9 (MMP9) (70,71)	Endopeptidase	3 - 41	Inflammation, fibrosis	Degrades extracellular matrix (ECM) proteins and activates cytokines and chemokines to regulate tissue remodelling	Increased	<0.01
N-terminal pro-B-type natriuretic peptide (NT-proBNP) (72,73)	Peptide (amino terminal fragment of prohormone BNP)	0.025 – 0.2	Myocardial stretch	Inactive N-terminal fragment of the precursor for BNP. Identified marker for release of BNP which is known to induce diuresis	Increased	<0.001

Suppression of tumorigenicity 2 (ST2) (74,75)	Member of IL-1 receptor family	Male 8.6 – 49.3 Female 7.2 – 33.5	Cellular stress, infection	Ligand of ST2 is IL-33 which is involved in reducing fibrosis and hypertrophy in mechanically strained tissues	-	N.S.
Soluble Interleukin-6 (SIL-6) (76,77)	Cytokine	0 – 0.044	Inflammation	Induces Th 17 differentiation, stimulates polarisation of macrophages	Increased	<0.001
Atrial Natriuretic Peptide (ANP) (78,79)	Hormone	0.004 – 0.006	Vascular remodelling, energy metabolism	Regulates salt-water balance and blood pressure by promoting renal sodium and water excretion and stimulating vasodilation. Also has anti-hypertrophic function in the heart	Increased	<0.05

P-value <0.05 was deemed statistically significant.

We have subsequently compared the relative peak intensities which were determined statistically significant (***p<0.0001) and successfully classified *via* SkiNET, yielding Raman key-bands, characteristic of patients with CVD (**Fig 6.7a**) and established the individual molecular barcodes building upon the biochemical fingerprints of both plasma and the CVD-indicative markers (**Fig 6.7b**). Machine learning classification combined with student's *t-test* revealed ApoB, LpA, IL-9, NT-ProBNP, PCSK9 and SIL-6 yielded spectral signatures of the previously identified key-bands, which are characteristic of CVD plasma. Of a particular interest are the Raman bands at 1003 and 1315 cm^{-1}, which were assigned to phenylalanine and lipids. Examining the spectral signatures of the CVD-indicative biomarkers, the SIL-6 yielded an increased intensity at 1003 cm^{-1} in-line with that observed in CVD plasma. Similarly, LpA, ApoB and PCSK9 demonstrated a considerable increase in intensity at 1315 cm^{-1} and at

1200-1340 cm^{-1} also principally assigned to lipids. Furthermore, there was an increase in Amide 1 at 1600-1655 cm^{-1}, found to be upregulated in both NT-ProBNP (~21%), IL-9 (~18%) and LpA (~28%). Collectively, these indicate the significant role of lipids and cytokines in the onset of CVD and thus, demonstrates their potential as significant indicators of disease.

The conventional understanding of atherosclerosis suggests that the amount of cholesterol contained within various types of lipoprotein particles (such as low-density lipoprotein, chylomicron and LpA particles) in the plasma plays a crucial role in determining the total mass of cholesterol deposited within the arterial walls and therefore contributing to the development of atherosclerosis. However, it is important to note that each of these particles contain one molecule of ApoB and there is now significant evidence to suggest that ApoB is a more accurate indicator of the risk of atherosclerosis associated with lipoproteins compared to either low-density lipoprotein cholesterol or non-high-density lipoprotein cholesterol, corroborating our findings where ApoB is recognised as a significant contributor and indicator of CVD [59]. Although an increase in cholesterol at 960 cm^{-1} has been observed, this effect is considerably lower than the increase associated with ApoB and LpA. Cholesterol can only penetrate the arterial wall when carried by ApoB particles and the amount of cholesterol carried per ApoB particle is known to vary. Consequently, the quantity of cholesterol deposited in the arterial wall is dependent on how many ApoB particles become trapped within it [91]. Moreover, once inside the arterial wall, smaller ApoB particles with less cholesterol content are more likely to become trapped than larger ApoB particles with higher cholesterol content since they exhibit stronger binding

affinities to the glycosaminoglycans found in the subintimal space of the arterial wall. Therefore, ApoB particles are highly atherogenic and serve to amplify and simplify the information provided by traditional lipid markers when assessing atherogenic risk attributable to the ApoB lipoproteins with the ApoB levels indicating the atherogenic particle concentration independent of the cholesterol content [92]. Furthermore, the identification of LpA as a strong indicator of CVD is supported with evidence from a multitude of mechanical, observational and genetic studies that provide substantial evidence for that LpA plays a causative role in the onset of CVD such as coronary artery disease, peripheral disease and most likely ischemic stroke [62]. High LpA concentrations are present in 10-20% of the population and have long been associated with an increased risk of ischemic cardiovascular disease however, at present they are not currently targeted in the prevention of CVD [27], [61]. In the instance of MI, high LpA concentrations predicted 2-to-3-fold increases in CVD risk. Similarly, NT-ProBNP is strongly associated with mortality in patients with CVD. A recent study by Echouffo-Tcheugui *et* al. determined that amongst individuals without prior CVD, elevated NT-ProBNP was associated with significantly higher risk of CVD mortality, with findings similar across all subgroups defined by age, sex, race, ethnicity and body mass index therefore, indicating NT-ProBNP may be useful for monitoring risk in the general population [93].

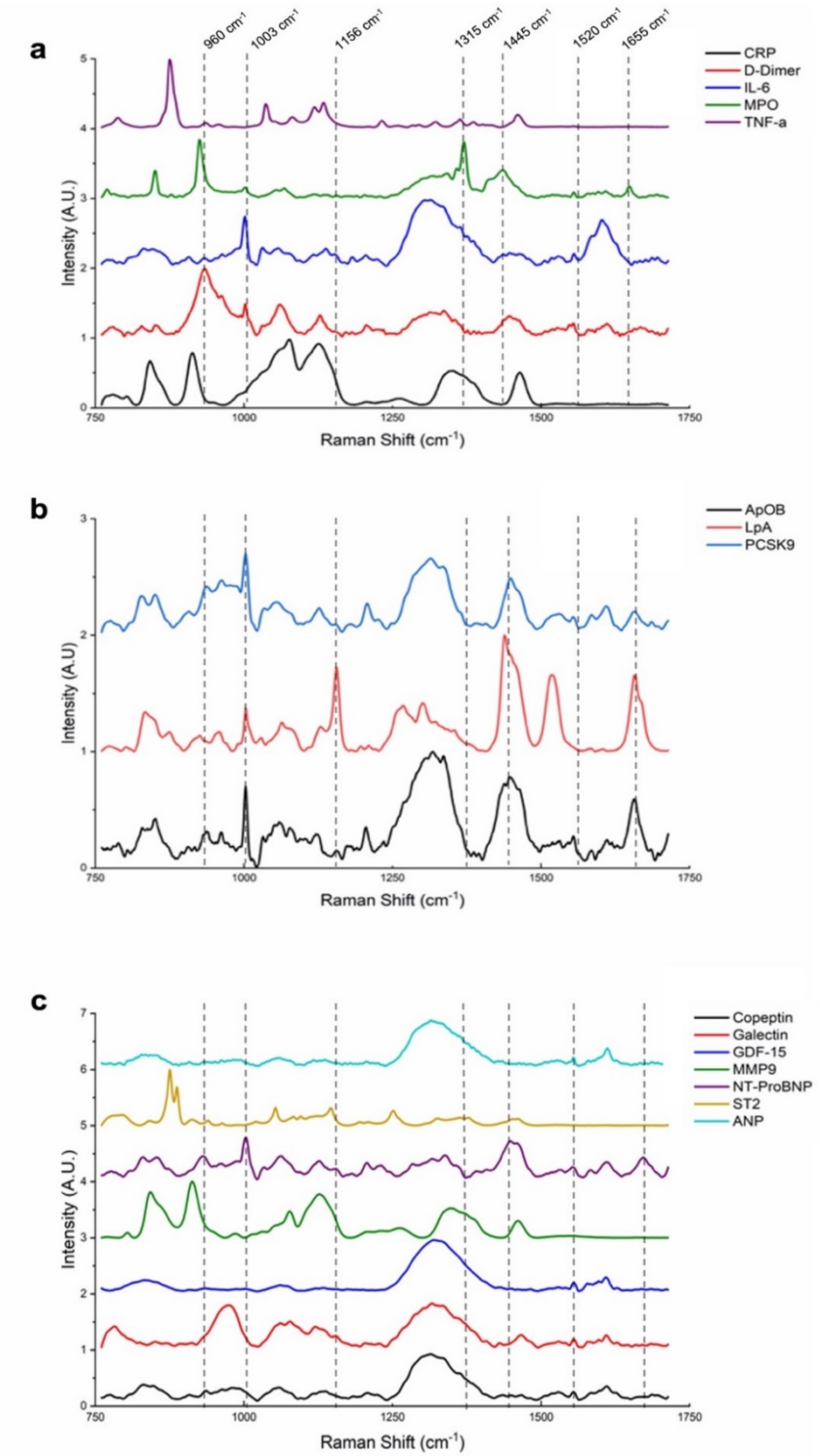

Figure 6.6 Average Raman spectra of candidate CVD biomarkers indicative of **(a)** thrombosis and inflammation, **(b)** cardiac lipids, **(c)** cardiac injury with (highlighted) SKiNET identified peaks of significance.

Currently, cardiac blood biomarkers include creatinine kinase (CK), CK-MB (a subtype of CK), cardiac troponin and myoglobin [94], [95], [96], [97]. Of these, cardiac troponin is the most used biomarker as it has the highest known sensitivity and enters the blood soon after a heart attack. Two forms of troponin may be measured, troponin T and troponin I, with troponin I highly specific to the heart and remaining in the blood longer than CK-MB. The challenge traditionally associated with CVD biomarkers is one of specificity and rapid clearance. CK for instance, may increase in conditions other than CVD, such as trauma and inflammation, which is also the case for CRP, detected at high levels in several inflammatory conditions [95]. CK-MB, while more sensitive to damage caused by a cardiac event, generally peaks 4-6 hours after a heart attack with levels returning to normal within 24-48 hours [96]. Many of these biomarkers are also unable to detect minor myocardial damage, therefore, being less suitable in patients with a delayed presentation of MI or other associated CVDs. From our further the spectroscopic profiling and the corresponding relative peak analyses, the most dominant spectral changes, determined to be statistically significant for detecting the different biomarkers ($p^{***}<0.0001$), yielded the biomolecular barcodes (**Fig 6.7b**). An accurate diagnostic and prognostic biomarkers pool for CVD remains an unmet clinical need. These barcodes provide valuable information regarding molecular composition, structure and interactions within the sample posing important advantages in healthcare setting applications including disease diagnosis, where the identified spectral features associated with disease would allow for accurate and early diagnosis and disease monitoring, where changes in spectroscopic barcodes over time could indicate whether a disease is advancing, regressing or remains stable. Moreover, molecular spectroscopy is essential in pharmacological studies to assess how medications

interact with biological molecules and tissues thus, advancing the general understanding of therapeutics mechanism and evaluating potential new candidates. The derived barcodes combined with the identified biochemical changes and their close match with notable biomarker characteristics, establish a group of potential first-line screening biomarkers of CVDs.

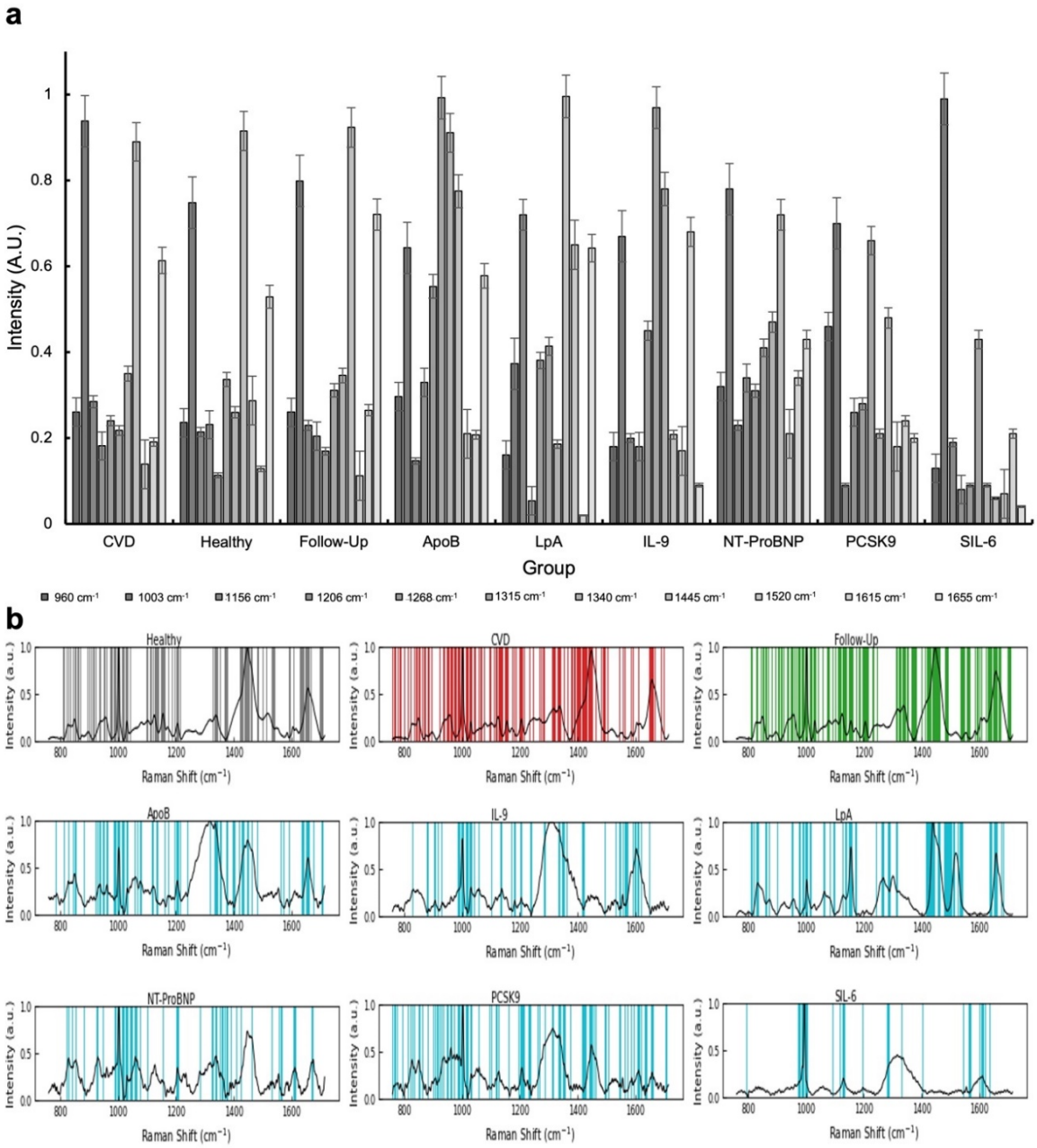

Figure 6.7. (a) Histograms of the relative intensities of the dominant Raman peaks identified in CVD, healthy, follow-up patient plasma and ApoB, LpA, IL-6, PCSK9, NT-ProBNP and SIL-6 CVD-indicative biomarkers. (b). Molecular barcodes derived from healthy, CVD and follow-up patient plasma and candidate CVD biomarkers including the ApoB, IL-6, LpA, NT-ProBNP, PCSK9 and SIL-6, highlighting the dominant, statically significant peaks.

Since many individuals with CVD are asymptomatic in the early stages of disease resulting in a delayed diagnosis and often when symptoms do present, they are non-specific, varied and attributed to a number of other conditions, a potential new spectroscopically determined biomarker panel could enhance disease diagnosis, treatment planning and monitoring in both at home and at healthcare settings. It can further provide an in-depth molecular-level understanding of disease, revealing changes in the underlying biochemistry *prior* to macroscopic tissue changes, enabling more precise and personalised approaches to patientcare. In addition, RS-SkiNET holds the potential to streamline and simplify CVD diagnostics, leading to quicker and more accurate results and improved long-term prognosis.

6.4 Conclusions

The presented findings underscore the potential of Raman spectroscopy as a powerful diagnostic and monitoring tool in the field of cardiology. We have demonstrated the ability of RS to accurately (>88%) detect molecular changes associated with CVD, where the high specificity allows for the detection of subtle alterations in the molecular composition of tissues and biomolecules, enabling the detection of cardiovascular disease at early stage when the clinical symptoms are absent or absent. This approach derives molecular barcodes and combined with advanced machine learning classification, correlates this to CVD-indicative biomarkers *via* the spectral fingerprints. Furthermore, the spectroscopic-SkiNET ability to classify healthy, CVD and follow-up patient plasma positions this as a pioneering technology for not only early diagnostics but also for longer term patient monitoring and prognosis. The *real-time* data acquisition, coupled with the non-invasive nature, renders RS-SkiNET a powerful asset

for *in-situ* tracking of CVDs, which is particularly crucial in chronic and progressive conditions, where regular monitoring can influence patient outcomes. Through the provision of dynamic molecular insights, this technique offers the potential to guide treatment adjustments, ensuring that interventions remain aligned with the evolving disease states. Furthermore, the remarkable role of RS-SkiNET in biomarker discovery led to the identification of six candidate biomarkers of CVDs including the ApoB, LpA, PCSK9, IL-9, SIL-6 and NT-ProBNP. These biomarkers hold promise not only for enhancing diagnostic accuracy, but also for providing subtle insights into the underlying pathophysiology.

In-depth analyses of CVD plasma spectra with those of healthy controls revealed further noticeable differences indicating CVD is heavily associated with systemic metabolomic alterations, offering potential diagnostic utility. Of a particular interest were the marked differences in cholesterol/lipid signatures as well as the overall protein profile of blood plasma. This differentiation consolidates the capability of Raman as a screening method in CVDs. Since patients often experience no symptoms of an underlying CVD, the ability of RS-SkiNET to detect molecular changes associated with it, may further enable the identification of risk factors including for instance, increased lipid or amino acid levels *prior* to the manifestation of clinical symptoms or a potentially fatal cardiac event. The non-invasive nature of RS allows for the assessment of cardiovascular health without the need for invasive procedures or contrast agents, reducing patient risk and discomfort. It also provides multi-parametric analysis with multiple molecular components analysed *simultaneously*, collectively providing a comprehensive view of cardiovascular health.

Building upon its molecular specificity, label-free capabilities, speed of detection, deployability outside the laboratory settings without loss of performance (hence, the inherent portability) and the potential for biomarker discovery, RS-SkiNET has the potential to complement existing screening tools and contribute to early CVD diagnostics as well as opening a promising new avenue for an advanced first-line screening tool for cardiovascular diseases, ultimately benefiting healthcare providers and patients as well as saving lives.

6.5 Materials and Methods

6.5.1 Patient samples and plasma preparation

Consecutive patients referred to the Cardiology department at Queen Elizabeth Hospital Birmingham were initially screened for eligibility. Inclusion criteria included a diagnosis of CVD and being aged between 18-85. CVD diagnosis was established using clinical, serological, CT and ultrasound data. Exclusion criteria included pregnancy, chronic liver disease, sepsis infection and rheumatological disease. In total, 60 CVD patients were recruited with a control group of 40 healthy volunteers with no known underlying health conditions or personal or family history of CVD. The overall patient demographics is summarised in Table 1. Written consent was obtained from all participants included in this study. Follow-up patient samples were obtained three months after the initial hospital visit. 7 mL of blood was collected in a vacutainer containing EDTA as an anticoagulant. These were immediately centrifuged at 2,000 x g for 20 minutes, followed by 13,000 x g for 2 minutes to obtain blood plasma. Blood plasma was stored in sterile cryovials (Thermo Fisher, UK) at 80°C. Samples were thawed at room temperature immediately prior to their analysis.

Table 6.4. Baseline clinical and laboratory characteristics of study patients**.**

		Control (n = 40)	**CVD (n = 60)**
Age (years)		48.5 ± 11.5	65.5 ± 23.5
Sex (Male/Female)		28/12	49/11
Systolic blood pressure (mmHg)		122 ± 8	135 ± 25
Diastolic blood pressure (mmHg)		77 ± 9	84 ± 36
BMI (kg/m^2)		28.2 ± 4.2	34.3 ± 6.8
Hypertension, n (%)		0	32 (53%)
Diabetes Mellitus - T1, n (%)		0	0
	-T2, n (%)	0	9 (15%)
Hypercholesterolaemia, n (%)		0	23 (38%)
Previous MI, n (%)		0	13 (22%)
Previous HF, n (%)		0	5 (8%)
Previous Stroke, n (%)		0	13 (22%)
Presentation	STEMI, n (%)	-	3 (5%)
	NSTEMI, n (%)	-	16 (27%)
	Atrial Fibrillation, n (%)	-	8 (13%)
	Other	-	33 (55%)
Medication	Aspirin, (%)	0	19 (32%)
	Clopidogrel (%)	0	6 (10%)
	Ticagrelor, n (%)	0	4 (7%)

6.5.2 Biomarkers

Biomarkers were purchased from Generon (Slough, UK), Miltenyo Biotec (Bergisch Gladbach, Germany) and Merck Life Science (UK). All biomarkers were reconstituted to a final working concentration of 1 mg/mL and were analysed immediately.

6.5.3 Raman Measurements

Raman measurements were acquired using a Renishaw *inVia* Qontor Raman confocal spectrometer (Renishaw, Wootton-under-Edge, UK) equipped with a Leica DMLM and a laser light source operating at 785 nm. Calibration was performed with a silicon wafer. Prior to analysis, 5 µL of plasma was deposited onto an aluminium foil covered glass slide and air dried for 30 minutes at room temperature. Raman measurements were performed using 5% laser power (~5 mW at source), x50 objective and a 1200 l/mm

grating. Spectra were acquired in the fingerprint region of 750-1650 cm^{-1} with a total of 10 acquisitions of 1s each. For each sample 100 spectra were collected acquiring 10x10 map with a step-size of 5 μm.

6.5.4 Data Acquisition and Analysis

WiRE software (Renishaw, UK) was used for the spectral acquisitions, to remove cosmic rays and for baseline subtraction using a 5^{th} order polynomial fit. Spectra were normalised using the standard normal variate (SNV) using a custom python script (Python 3.7). Spectral data analysis was performed and plotted using OriginPro, version 2023 (OriginLab Corporation, Northampton, MA, USA). The student's *t*-test was used to determine statistically significant differences between peak intensities in the Raman spectra for different sample groups. *p*-values of <0.05 were considered significant. Multivariate analysis was performed using SKiNET with the accompanying Raman Toolkit web interface to build SOM models using training data (80%) and perform predictions against test data (20%). 10-fold cross validation of the training data was completed to optimise SKiNET models, along with fine tuning of the number of training steps, grid size and initial learning rate. The final model consisted of 10x10 grid of neurons, 46,080 (4 epochs of the data), and initial learning rate of 0.1. Classification of the test data was repeated 10 times from separate SOM initialisations. SOMDI, by assigning class vectors as labels for each spectrum with corresponding weight vectors for each neuron, allowed the identification of the data type responsible for activating a particular neuron. This was subsequently used to inspect weights across all neurons and allow the extraction of prominent spectral features belonging to each class. Barcoding was performed *via* Python with barcodes generated using the

Savitzky-Golay filter to calculate the second derivative of each spectrum. The smoothing window was set to 21 with polynomial order of 2. Values exceeding 40% of the maximum height were assigned as 1, while values below this threshold were categorised as 0. These values were then overlaid onto the representative mean spectra, aligning with the underpinning peak positions.

6.6 References

[1] M. Di Cesare *et al.*, "World Heart Report 2023," Geneva, 2023.

[2] P. (Babu) Balagopal *et al.*, "Nontraditional Risk Factors and Biomarkers for Cardiovascular Disease: Mechanistic, Research, and Clinical Considerations for Youth," *Circulation*, vol. 123, no. 23, pp. 2749–2769, Jun. 2011, doi: 10.1161/CIR.0b013e31821c7c64.

[3] T. Gaziano, K. Reddy, and P. F, "Cardiovascular Disease," in *Disease Control Priorities in Developing Countries*, 2nd ed., D. Jamison, J. Breman, and A. Measham, Eds., Oxford University Press, 2006, pp. 1–62.

[4] J. J. Ricotta, J. Pagan, M. Xenos, Y. Alemu, S. Einav, and D. Bluestein, "Cardiovascular disease management: the need for better diagnostics," *Med Biol Eng Comput*, vol. 46, no. 11, pp. 1059–1068, Nov. 2008, doi: 10.1007/s11517-008-0416-x.

[5] R. J. Bisoendial, S. M. Boekholdt, M. Vergeer, E. S. G. Stroes, and J. J. P. Kastelein, "C-reactive protein is a mediator of cardiovascular disease," *Eur Heart J*, vol. 31, no. 17, pp. 2087–2091, Sep. 2010, doi: 10.1093/eurheartj/ehq238.

[6] R. S. Vasan, "Biomarkers of Cardiovascular Disease," *Circulation*, vol. 113, no. 19, pp. 2335–2362, May 2006, doi: 10.1161/CIRCULATIONAHA.104.482570.

[7] C. Soliman, J. Faircloth, D. Tu, S. Mabbott, K. Maitland, and G. Coté, "Exploring the Clinical Utility of Raman Spectroscopy for Point-of-Care Cardiovascular Disease Biomarker Detection," *Appl Spectrosc*, Jul. 2023, doi: 10.1177/00037028231187963.

[8] P. Libby *et al.*, *Braunwald's Heart Disease*, 9th ed. Philadelphia: Elsevier Inc., 2022.

[9] R. Dhingra and R. S. Vasan, "Biomarkers in cardiovascular disease: Statistical assessment and section on key novel heart failure biomarkers," *Trends Cardiovasc Med*, vol. 27, no. 2, pp. 123–133, Feb. 2017, doi: 10.1016/j.tcm.2016.07.005.

[10] N. M. Ćulum *et al.*, "Characterization of ovarian cancer-derived extracellular vesicles by surface-enhanced Raman spectroscopy," *Analyst*, vol. 146, no. 23, pp. 7194–7206, 2021, doi: 10.1039/D1AN01586A.

[11] W. Zhang *et al.*, "Enabling Sensitive Phenotypic Profiling of Cancer-Derived Small Extracellular Vesicles Using Surface-Enhanced Raman Spectroscopy Nanotags," *ACS Sens*, vol. 5, no. 3, pp. 764–771, Mar. 2020, doi: 10.1021/acssensors.9b02377.

[12] K. Liu, Q. Zhao, B. Li, and X. Zhao, “Raman Spectroscopy: A Novel Technology for Gastric Cancer Diagnosis,” *Front Bioeng Biotechnol*, vol. 10, no. March, pp. 1–11, 2022, doi: 10.3389/fbioe.2022.856591.

[13] L. Wang *et al.*, “Raman spectroscopy, a potential tool in diagnosis and prognosis of castration-resistant prostate cancer,” *J Biomed Opt*, vol. 18, no. 8, p. 087001, Aug. 2013, doi: 10.1117/1.JBO.18.8.087001.

[14] E. Ryzhikova *et al.*, “Raman spectroscopy of blood serum for Alzheimer’s disease diagnostics: specificity relative to other types of dementia,” *J Biophotonics*, vol. 8, no. 7, pp. 584–596, Jul. 2015, doi: 10.1002/jbio.201400060.

[15] G. Cennamo *et al.*, “Surface-enhanced Raman spectroscopy of tears: toward a diagnostic tool for neurodegenerative disease identification,” *J Biomed Opt*, vol. 25, no. 08, p. 1, 2020, doi: 10.1117/1.jbo.25.8.087002.

[16] K. Dixon *et al.*, “Using Machine Learning and Silver Nanoparticle-Based Surface-Enhanced Raman Spectroscopy for Classification of Cardiovascular Disease Biomarkers,” *ACS Appl Nano Mater*, vol. 6, no. 17, pp. 15385–15396, Sep. 2023, doi: 10.1021/acsanm.3c01442.

[17] R. L. McCreery, *Raman spectrscopy for chemical analysis*. John Wiley & Sons, Ltd, 2005.

[18] S. Desai *et al.*, “Raman spectroscopy-based detection ofRNAviruses in saliva: A preliminary report,” *J Biophotonics*, vol. 13, no. 10, Oct. 2020, doi: 10.1002/jbio.202000189.

[19] K. Dana, C. Shende, H. Huang, and S. Farquharson, “Rapid Analysis of Cocaine in Saliva by Surface-Enhanced Raman Spectroscopy.,” *J Anal Bioanal Tech*, vol. 6, no. 6, pp. 1–5, 2015, doi: 10.4172/2155-9872.1000289.

[20] E. Smith and G. Dent, *Modern Raman spectroscopy: A Practical Approach*. John Wiley & Sons, Ltd, 2019.

[21] R. S. Das and Y. K. Agrawal, “Raman spectroscopy: Recent advancements, techniques and applications,” *Vib Spectrosc*, vol. 57, no. 2, pp. 163–176, Nov. 2011, doi: 10.1016/j.vibspec.2011.08.003.

[22] J. Popp and Mäyerhöfer, *Micro-Raman Spectroscopy: Theory and Application*, Illustrate. Walter de Gruyter GmbH & Co KG, 2020.

[23] K. Ember *et al.*, “Saliva-based detection of COVID-19 infection in a real-world setting using reagent-free Raman spectroscopy and machine learning,” *J Biomed Opt*, vol. 27, no. 02, 2022, doi: 10.1117/1.jbo.27.2.025002.

[24] M. Taniguchi *et al.*, “Multimolecular Salivary Mucin Complex Is Altered in Saliva of Cigarette Smokers: Detection of Disulfide Bridges by Raman Spectroscopy,” *Biomed Res Int*, vol. 2013, p. 168765, 2013, doi: 10.1155/2013/168765.

[25] M. Paraskevaidi *et al.*, “Raman spectroscopic techniques to detect ovarian cancer biomarkers in blood plasma,” *Talanta*, vol. 189, no. June, pp. 281–288, 2018, doi: 10.1016/j.talanta.2018.06.084.

[26] C. Morasso *et al.*, “Raman analysis reveals biochemical differences in plasma of crohn’s disease patients,” *J Crohns Colitis*, vol. 14, no. 11, pp. 1572–1580, 2020, doi: 10.1093/ecco-jcc/jjaa080.

[27] S. Bhargava, S. de la Puente-Secades, L. Schurgers, and J. Jankowski, “Lipids and lipoproteins in cardiovascular diseases: a classification,” *Trends in Endocrinology & Metabolism*, vol. 33, no. 6, pp. 409–423, Jun. 2022, doi: 10.1016/j.tem.2022.02.001.

[28] F. Mach *et al.*, “2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk,” *Eur Heart J*, vol. 41, no. 1, pp. 111–188, Jan. 2020, doi: 10.1093/eurheartj/ehz455.

[29] F. H. Verbrugge, W. H. W. Tang, and S. L. Hazen, “Protein carbamylation and cardiovascular disease,” *Kidney Int*, vol. 88, no. 3, pp. 474–478, Sep. 2015, doi: 10.1038/ki.2015.166.

[30] R. I. Litvinov, D. A. Faizullin, Y. F. Zuev, and J. W. Weisel, “The α-Helix to β-Sheet Transition in Stretched and Compressed Hydrated Fibrin Clots,” *Biophys J*, vol. 103, no. 5, pp. 1020–1027, Sep. 2012, doi: 10.1016/j.bpj.2012.07.046.

[31] C. Grajeda-Iglesias and M. Aviram, “Specific Amino Acids Affect Cardiovascular Diseases and Atherogenesis via Protection against Macrophage Foam Cell Formation: Review Article,” *Rambam Maimonides Med J*, vol. 9, no. 3, p. e0022, Jul. 2018, doi: 10.5041/RMMJ.10337.

[32] K. Nitz, M. Lacy, and D. Atzler, “Amino Acids and Their Metabolism in Atherosclerosis,” *Arterioscler Thromb Vasc Biol*, vol. 39, no. 3, pp. 319–330, Mar. 2019, doi: 10.1161/ATVBAHA.118.311572.

[33] S. H. Shah *et al.*, “Association of a Peripheral Blood Metabolic Profile With Coronary Artery Disease and Risk of Subsequent Cardiovascular Events,” *Circ Cardiovasc Genet*, vol. 3, no. 2, pp. 207–214, Apr. 2010, doi: 10.1161/CIRCGENETICS.109.852814.

[34] A. A. M. Vaarhorst *et al.*, “A metabolomic profile is associated with the risk of incident coronary heart disease,” *Am Heart J*, vol. 168, no. 1, pp. 45-52.e7, Jul. 2014, doi: 10.1016/j.ahj.2014.01.019.

[35] P. Wurtz *et al.*, “High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis,” *Eur Heart J*, vol. 33, no. 18, pp. 2307–2316, Sep. 2012, doi: 10.1093/eurheartj/ehs020.

[36] O. Rom and M. Aviram, “It is not just lipids: proatherogenic vs. antiatherogenic roles for amino acids in macrophage foam cell formation,” *Curr Opin Lipidol*, vol. 28, no. 1, pp. 85–87, Feb. 2017, doi: 10.1097/MOL.0000000000000377.

[37] I. Chernyavskiy, S. Veeranki, U. Sen, and S. C. Tyagi, “Atherogenesis: hyperhomocysteinemia interactions with LDL, macrophage function, paraoxonase 1, and exercise,” *Ann N Y Acad Sci*, vol. 1363, no. 1, pp. 138–154, Jan. 2016, doi: 10.1111/nyas.13009.

[38] R. Jauhiainen, J. Vangipurapu, A. Laakso, T. Kuulasmaa, J. Kuusisto, and M. Laakso, “The Association of 9 Amino Acids With Cardiovascular Events in Finnish Men in a 12-Year Follow-up Study,” *J Clin Endocrinol Metab*, vol. 106, no. 12, pp. 3448–3454, Nov. 2021, doi: 10.1210/clinem/dgab562.

[39] P. Wurtz *et al.*, “High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis,” *Eur Heart J*, vol. 33, no. 18, pp. 2307–2316, Sep. 2012, doi: 10.1093/eurheartj/ehs020.

[40] J. Vangipurapu, A. Stancáková, U. Smith, J. Kuusisto, and M. Laakso, “Nine Amino Acids Are Associated With Decreased Insulin Secretion and Elevated Glucose Levels in a 7.4-Year Follow-up Study of 5,181 Finnish Men,” *Diabetes*, vol. 68, no. 6, pp. 1353–1358, Jun. 2019, doi: 10.2337/db18-1076.

[41] Y. Deng, C. Huang, J. Su, C.-W. Pan, and C. Ke, “Identification of biomarkers for essential hypertension based on metabolomics,” *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 31, no. 2, pp. 382–395, Feb. 2021, doi: 10.1016/j.numecd.2020.11.023.

[42] N. R. Cook, “A Randomized Factorial Trial of Vitamins C and E and Beta Carotene in the Secondary Prevention of Cardiovascular Events in Women,” *Arch Intern Med*, vol. 167, no. 15, p. 1610, Aug. 2007, doi: 10.1001/archinte.167.15.1610.

[43] G. S. Omenn *et al.*, “Effects of a Combination of Beta Carotene and Vitamin A on Lung Cancer and Cardiovascular Disease,” *New England Journal of Medicine*, vol. 334, no. 18, pp. 1150–1155, May 1996, doi: 10.1056/NEJM199605023341802.

[44] I.-M. Lee, N. R. Cook, J. E. Manson, J. E. Buring, and C. H. Hennekens, “β-Carotene Supplementation and Incidence of Cancer and Cardiovascular Disease: the Women’s Health Study,” *JNCI: Journal of the National Cancer Institute*, vol. 91, no. 24, pp. 2102–2106, Dec. 1999, doi: 10.1093/jnci/91.24.2102.

[45] J. M. Rapola *et al.*, “Randomised trial of α-tocopherol and β-carotene supplements on incidence of major coronary events in men with previous myocardial infarction,” *The Lancet*, vol. 349, no. 9067, pp. 1715–1720, Jun. 1997, doi: 10.1016/S0140-6736(97)01234-8.

[46] W. Chen *et al.*, “Elevated plasma phenylalanine predicts mortality in critical patients with heart failure,” *ESC Heart Fail*, vol. 7, no. 5, pp. 2884–2893, Oct. 2020, doi: 10.1002/ehf2.12896.

[47] M. D. Shapiro and S. Fazio, “From Lipids to Inflammation,” *Circ Res*, vol. 118, no. 4, pp. 732–749, Feb. 2016, doi: 10.1161/CIRCRESAHA.115.306471.

[48] M. V. Holmes *et al.*, “Mendelian randomization of blood lipids for coronary heart disease,” *Eur Heart J*, vol. 36, no. 9, pp. 539–550, Mar. 2015, doi: 10.1093/eurheartj/eht571.

[49] K. Leander *et al.*, “Circulating Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) Predicts Future Risk of Cardiovascular Events Independently of Established Risk Factors,” *Circulation*, vol. 133, no. 13, pp. 1230–1239, Mar. 2016, doi: 10.1161/CIRCULATIONAHA.115.018531.

[50] Y. S. Byun *et al.*, “Relationship of Oxidized Phospholipids on Apolipoprotein B-100 to Cardiovascular Outcomes in Patients Treated With Intensive Versus Moderate Atorvastatin Therapy,” *J Am Coll Cardiol*, vol. 65, no. 13, pp. 1286–1295, Apr. 2015, doi: 10.1016/j.jacc.2015.01.050.

[51] R. F. Storey, H. M. Sanderson, A. E. White, J. A. May, K. E. Cameron, and S. Heptinstall, “The central role of the P 2T receptor in amplification of human platelet activation, aggregation, secretion and procoagulant activity,” *Br J Haematol*, vol. 110, no. 4, pp. 925–934, Sep. 2000, doi: 10.1046/j.1365-2141.2000.02208.x.

[52] M. R. Thomas and R. F. Storey, “Clinical significance of residual platelet reactivity in patients treated with platelet P2Y12 inhibitors,” *Vascul Pharmacol*, vol. 84, pp. 25–27, Sep. 2016, doi: 10.1016/j.vph.2016.05.010.

[53] T. Vajen, S. Mause, and R. Koenen, “Microvesicles from platelets: novel drivers of vascular inflammation,” *Thromb Haemost*, vol. 114, no. 08, pp. 228–236, Aug. 2015, doi: 10.1160/TH14-11-0962.

[54] M. Thomas and R. Storey, “The role of platelets in inflammation,” *Thromb Haemost*, vol. 114, no. 09, pp. 449–458, Nov. 2015, doi: 10.1160/TH14-12-1067.

[55] M. R. Thomas *et al.*, “Platelet P2Y 12 Inhibitors Reduce Systemic Inflammation and Its Prothrombotic Effects in an Experimental Human Model,” *Arterioscler Thromb Vasc Biol*, vol. 35, no. 12, pp. 2562–2570, Dec. 2015, doi: 10.1161/ATVBAHA.115.306528.

[56] S. D. Prabhu and N. G. Frangogiannis, “The Biological Basis for Cardiac Repair After Myocardial Infarction,” *Circ Res*, vol. 119, no. 1, pp. 91–112, Jun. 2016, doi: 10.1161/CIRCRESAHA.116.303577.

[57] F. K. Swirski and M. Nahrendorf, “Leukocyte Behavior in Atherosclerosis, Myocardial Infarction, and Heart Failure,” *Science (1979)*, vol. 339, no. 6116, pp. 161–166, Jan. 2013, doi: 10.1126/science.1230719.

[58] K. L. Wong *et al.*, “Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets,” *Blood*, vol. 118, no. 5, pp. e16–e31, Aug. 2011, doi: 10.1182/blood-2010-12-326355.

[59] A. D. Sniderman *et al.*, “Apolipoprotein B Particles and Cardiovascular Disease,” *JAMA Cardiol*, vol. 4, no. 12, p. 1287, Dec. 2019, doi: 10.1001/jamacardio.2019.3780.

[60] J. Behbodikhah *et al.*, “Apolipoprotein B and Cardiovascular Disease: Biomarker and Potential Therapeutic Target,” *Metabolites*, vol. 11, no. 10, p. 690, Oct. 2021, doi: 10.3390/metabo11100690.

[61] R. C. Maranhão, P. O. Carvalho, C. C. Strunz, and F. Pileggi, “Lipoprotein (a): Structure, Pathophysiology and Clinical Implications,” *Arq Bras Cardiol*, 2014, doi: 10.5935/abc.20140101.

[62] P. R. Kamstrup, “Lipoprotein(a) and Cardiovascular Disease,” *Clin Chem*, vol. 67, no. 1, pp. 154–166, Jan. 2021, doi: 10.1093/clinchem/hvaa247.

[63] J. H. Boyd, C. D. Fjell, J. A. Russell, D. Sirounis, M. S. Cirstea, and K. R. Walley, “Increased Plasma PCSK9 Levels Are Associated with Reduced Endotoxin Clearance and the Development of Acute Organ Failures during Sepsis,” *J Innate Immun*, vol. 8, no. 2, pp. 211–220, 2016, doi: 10.1159/000442976.

[64] S. Glerup, R. Schulz, U. Laufs, and K.-D. Schlüter, “Physiological and therapeutic regulation of PCSK9 activity in cardiovascular disease,” *Basic Res Cardiol*, vol. 112, no. 3, p. 32, May 2017, doi: 10.1007/s00395-017-0619-0.

[65] M. B. Pepys and G. M. Hirschfield, “C-reactive protein: a critical update,” *Journal of Clinical Investigation*, vol. 111, no. 12, pp. 1805–1812, Jun. 2003, doi: 10.1172/JCI18921.

[66] D. L. Cozlea *et al.*, “The impact of C reactive protein on global cardiovascular risk on patients with coronary artery disease.,” *Curr Health Sci J*, vol. 39, no. 4, pp. 225–31, Oct. 2013, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24778862

[67] E. J. Bounds and S. J. Kok, *D Dimer*. 2023. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/30701097

[68] E. S. Ford *et al.*, “Traditional risk factors and D-dimer predict incident cardiovascular disease events in chronic HIV infection,” *AIDS*, vol. 24, no. 10, pp. 1509–1517, Jun. 2010, doi: 10.1097/QAD.0b013e32833ad914.

[69] E. A. Said *et al.*, “Defining IL-6 levels in healthy individuals: A meta-analysis,” *J Med Virol*, vol. 93, no. 6, pp. 3915–3924, Jun. 2021, doi: 10.1002/jmv.26654.

[70] H. Susilo *et al.*, “The Role of Plasma Interleukin-6 Levels on Atherosclerotic Cardiovascular Disease and Cardiovascular Mortality Risk Scores in Javanese

Patients with Chronic Kidney Disease," *J Pers Med*, vol. 12, no. 7, p. 1122, Jul. 2022, doi: 10.3390/jpm12071122.

[71] S. J. Nicholls and S. L. Hazen, "Myeloperoxidase and Cardiovascular Disease," *Arterioscler Thromb Vasc Biol*, vol. 25, no. 6, pp. 1102–1111, Jun. 2005, doi: 10.1161/01.ATV.0000163262.83456.6d.

[72] C. Frangie and J. Daher, "Role of myeloperoxidase in inflammation and atherosclerosis (Review)," *Biomed Rep*, vol. 16, no. 6, p. 53, May 2022, doi: 10.3892/br.2022.1536.

[73] E. Trisnawati, W. Nontji, and St. Nurasni, "Tumour necrosis factor-α (TNF-α) serum levels in preeclampsia pregnant women and pregnant women at risk with preeclampsia," *Enferm Clin*, vol. 30, pp. 27–30, Mar. 2020, doi: 10.1016/j.enfcli.2019.07.021.

[74] F. L. Bellisarii, S. Gallina, and R. De Caterina, "Tumor necrosis factor-alpha and cardiovascular diseases.," *Ital Heart J*, vol. 2, no. 6, pp. 408–17, Jun. 2001, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11453575

[75] L. Dobsa and K. Cullen Edozien, "Copeptin and its potential role in diagnosis and prognosis of various diseases," *Biochem Med (Zagreb)*, pp. 172–192, 2013, doi: 10.11613/BM.2013.021.

[76] F. Schill, S. Timpka, P. M. Nilsson, O. Melander, and S. Enhörning, "Copeptin as a predictive marker of incident heart failure," *ESC Heart Fail*, vol. 8, no. 4, pp. 3180–3188, Aug. 2021, doi: 10.1002/ehf2.13439.

[77] V. Blanda, U. M. Bracale, M. D. Di Taranto, and G. Fortunato, "Galectin-3 in Cardiovascular Diseases," *Int J Mol Sci*, vol. 21, no. 23, p. 9232, Dec. 2020, doi: 10.3390/ijms21239232.

[78] E. Danese, G. L. Salvagno, B. Caruso, M. Montagnana, F. Sanchis-Gomar, and G. Lippi, "Physiologic determinants of serum galectin-3 in a general healthy population," *Ann Res Hosp*, vol. 1, pp. 1–1, 2017, doi: 10.21037/arh.2017.08.03.

[79] M. Li *et al.*, "Growth differentiation factor-15 is associated with cardiovascular outcomes in patients with coronary artery disease," *Cardiovasc Diabetol*, vol. 19, no. 1, p. 120, Dec. 2020, doi: 10.1186/s12933-020-01092-7.

[80] P. Welsh *et al.*, "Reference ranges for GDF-15, and risk factors associated with GDF-15, in a large general population cohort," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 60, no. 11, pp. 1820–1829, Oct. 2022, doi: 10.1515/cclm-2022-0135.

[81] K. Thrailkill, G. Cockrell, P. Simpson, C. Moreau, J. Fowlkes, and R. C. Bunn, "Physiological matrix metalloproteinase (MMP) concentrations: comparison of serum and plasma specimens," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 44, no. 4, Jan. 2006, doi: 10.1515/CCLM.2006.090.

[82] A. Yabluchanskiy, Y. Ma, R. P. Iyer, M. E. Hall, and M. L. Lindsey, "Matrix Metalloproteinase-9: Many Shades of Function in Cardiovascular Disease," *Physiology*, vol. 28, no. 6, pp. 391–403, Nov. 2013, doi: 10.1152/physiol.00029.2013.

[83] P. Welsh *et al.*, "Reference Ranges for NT-proBNP (N-Terminal Pro-B-Type Natriuretic Peptide) and Risk Factors for Higher NT-proBNP Concentrations in a Large General Population Cohort," *Circ Heart Fail*, vol. 15, no. 10, Oct. 2022, doi: 10.1161/CIRCHEARTFAILURE.121.009427.

[84] J. B. Echouffo-Tcheugui *et al.*, “NT-proBNP and All-Cause and Cardiovascular Mortality in US Adults: A Prospective Cohort Study,” *J Am Heart Assoc*, vol. 12, no. 11, Jun. 2023, doi: 10.1161/JAHA.122.029110.

[85] R. H. Parikh, S. L. Seliger, R. Christenson, J. S. Gottdiener, B. M. Psaty, and C. R. DeFilippi, “Soluble ST2 for Prediction of Heart Failure and Cardiovascular Death in an Elderly, Community-Dwelling Population,” *J Am Heart Assoc*, vol. 5, no. 8, Aug. 2016, doi: 10.1161/JAHA.115.003188.

[86] Y. Sun, H. Pavey, I. Wilkinson, and M. Fisk, “Role of the IL-33/ST2 axis in cardiovascular disease: A systematic review and meta-analysis,” *PLoS One*, vol. 16, no. 11, p. e0259026, Nov. 2021, doi: 10.1371/journal.pone.0259026.

[87] Y. Feng *et al.*, “The Role of Interleukin-6 Family Members in Cardiovascular Diseases,” *Front Cardiovasc Med*, vol. 9, Mar. 2022, doi: 10.3389/fcvm.2022.818890.

[88] P. Baran *et al.*, “The balance of interleukin (IL)-6, IL-6·soluble IL-6 receptor (sIL-6R), and IL-6·sIL-6R·sgp130 complexes allows simultaneous classic and trans-signaling,” *Journal of Biological Chemistry*, vol. 293, no. 18, pp. 6762–6775, May 2018, doi: 10.1074/jbc.RA117.001163.

[89] H. Kook *et al.*, “Physiological concentration of atrial natriuretic peptide induces endothelial regeneration in vitro,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 284, no. 4, pp. H1388–H1397, Apr. 2003, doi: 10.1152/ajpheart.00414.2002.

[90] W. Song, H. Wang, and Q. Wu, “Atrial natriuretic peptide in cardiovascular biology and disease (NPPA),” *Gene*, vol. 569, no. 1, pp. 1–6, Sep. 2015, doi: 10.1016/j.gene.2015.06.029.

[91] S.-O. OLOFSSON and J. BOREN, “Apolipoprotein B: a clinically important apolipoprotein which assembles atherogenic lipoproteins and promotes the development of atherosclerosis,” *J Intern Med*, vol. 258, no. 5, pp. 395–410, Nov. 2005, doi: 10.1111/j.1365-2796.2005.01556.x.

[92] A. von Zychlinski, M. Williams, S. McCormick, and T. Kleffmann, “Absolute quantification of apolipoproteins and associated proteins on human plasma lipoproteins,” *J Proteomics*, vol. 106, pp. 181–190, Jun. 2014, doi: 10.1016/j.jprot.2014.04.030.

[93] J. B. Echouffo-Tcheugui *et al.*, “NT-proBNP and All-Cause and Cardiovascular Mortality in US Adults: A Prospective Cohort Study,” *J Am Heart Assoc*, vol. 12, no. 11, Jun. 2023, doi: 10.1161/JAHA.122.029110.

[94] J. Kottwitz *et al.*, “Myoglobin for Detection of High-Risk Patients with Acute Myocarditis,” *J Cardiovasc Transl Res*, vol. 13, no. 5, pp. 853–863, Oct. 2020, doi: 10.1007/s12265-020-09957-8.

[95] L. M. Brewster, Y. C. Haan, A. H. Zwinderman, B. J. van den Born, and G. A. van Montfrans, “CK (Creatine Kinase) Is Associated With Cardiovascular Hemodynamics,” *Hypertension*, vol. 76, no. 2, pp. 373–380, Aug. 2020, doi: 10.1161/HYPERTENSIONAHA.120.14675.

[96] Y.-W. Wu *et al.*, “Potential impacts of high-sensitivity creatine kinase-MB on long-term clinical outcomes in patients with stable coronary heart disease,” *Sci Rep*, vol. 10, no. 1, p. 5638, Mar. 2020, doi: 10.1038/s41598-020-61894-3.

[97] P. Welsh *et al.*, “Cardiac Troponin T and Troponin I in the General Population,” *Circulation*, vol. 139, no. 24, pp. 2754–2764, Jun. 2019, doi: 10.1161/CIRCULATIONAHA.118.038529.

CHAPTER 7

ADVANCED INTEGRATED MULTIPURPOSE SPECTROSCOPIC LAB-ON-A-CHIP FOR THE TIMELY DETECTION OF EXTRACELLULAR VESICLES AS KEY-MARKERS OF DISEASES

The work presented in this results chapter is based on a paper currently in submission as we well as two patent applications;

1. **Buchan E.,** Rickard, J.J.S, Thomas M., Goldberg Oppenheimer P., [2023], 'Advanced Integrated Multipurpose Spectroscopic Lab-on-a-chip for Timely Detection of Extracellular Vesicles as Key-Markers of Diseases.' Submitted.
2. **Buchan E.,** Rickard J.J.S., and Goldberg Oppenheimer P. [2023], 'Lab-on-a-chip.' UK Patent Application No. GB2304236.9
3. **Buchan E.,** Rickard J.J.S., and Goldberg Oppenheimer P. [2023], 'Lab-on-a-chip method of diagnosis.' UK Patent Application No. GB2304237.7

Author contributions – **E Buchan**: Conceptualisation, Methodology, Formal analysis, Writing – original draft, review and editing. **J Rickard** – Methodolgy, Writing – original draft, review and editing. **M Thomas** – Resources, Writing – review and editing. **P Goldberg Oppenheimer** – Conceptualisation, Methodology, Resources, Writing – review and editing, funding acquisition, supervision.

7.1 Abstract

One of the biggest public health challenges faced worldwide is the growth of the global ageing population. Associated with this, is the increase in incidence of serious health conditions such cardiovascular diseases (CVD) and inflammatory bowel diseases (IBD). Often, there are no early underlying symptoms, with a heart attack or stroke being first indications of CVD and sudden abdominal pain for IBDs. At present, there is no easy way of detecting early disease of individuals or assessing if someone is in good health at the point-of-care or in the community with no portable technology existing for quantitative assessment of early-stage CVDs and IBDs with sufficient sensitivity and timeliness to aid the stratification with rapid, early-diagnosis and triage, whether this is at GP/consultant clinics, or *in-situ* monitoring following diagnosis. Therefore, the development of a quick, simple and low-cost bedside, point-of-care technique to detect biomarkers indicative of diseases, from readily available biofluids such as blood-plasma or saliva, are of a vital importance. Lipid membrane extracellular vesicles (EVs), varying in size and conformation, are known to play a role in numerous diverse processes including intercellular communication, recycling of membrane proteins and lipids, immune modulation and cellular proliferation. EVs can act as a biomarker, specifying the progression of the disease state of the cells in which they originate. Herein, we have developed a portable advanced integrated multipurpose spectroscopic lab-on-a-chip (AIMSpec-LoC) for timely and rapid detection of CVDs and IBDs *via* specific fingerprinting for low-cost, non-invasive, early-stage diagnosis. The hybrid on-chip isolation with our advanced AI discrimination provides simultaneous analysis of multiple EV subtypes indicative of disease, demonstrating classification accuracy of >97% in the diagnosis of CVD and IBD from blood plasma and saliva,

respectively. AIMSpec-LoC is subsequently validated, uniquely capturing and detecting EVs while delivering molecular fingerprint specificity with sensitivity of >96%, yielding an attractive tool for sensing target biomarker molecules for timely diagnosis and rapid monitoring. Being convenient and efficient, it is capable to not only reduce the time spent waiting for a definitive diagnosis, improve prognosis and save lives but also, allow for accurate measurement of the efficacy of specific clinical therapies and interventions, pinpointing clinical progress. Advanced computational spectral analysis for quantitative determination of biomarkers with clinical input on the validated panels will in turn, enable accurate prognostic models of pathogenesis. Overall, AIMSpec-LoC development, by leveraging the intrinsic sensing properties of advanced optofluidic lab-on-a-chip, lays the platform towards a disruptive, *real-time* diagnostic platform technology, addressing an unmet need for improved interventions, which in the long-term, will increase the period individuals spend in good health and ensure that they reach an old age in good health as well as reduce morbidity.

7.2 Introduction

Cardiovascular diseases (CVDs) are the leading cause of morbidity and mortality worldwide, with high-complication rates. 20.5 million people died from CVDs in 2019, accounting for 32% of global deaths [1]. Although, many CVDs could be prevented by addressing behavioural risk factors of heart diseases and stroke, *e.g.,* physical inactivity and tobacco use, often no underlying atherosclerotic arterial disease has yet been identified and only the fatal cardiac events such as stroke or heart attack constitute the first signs. Inefficient prevention of cardiovascular disease increases the incidence of cardiovascular-related health conditions, contributing to higher demands on healthcare system. Furthermore, the steadily growing elderly population has been

increasing the demand from the health and social services for prolonged periods as patients with multiple CVDs require long-term management with repeated visits to healthcare professionals. Therefore, it is crucial to detect CVDs as early as possible to reduce the major burden on the healthcare systems through the inclusion of management, interventions and medications. Similarly, inflammatory bowel disease (IBD), including Crohn's disease (CD) and ulcerative colitis (UC) affects over 10 million people worldwide with the incidences continuously increasing [2]. These idiopathic inflammatory bowel disorders are characterised by both acute and chronic diseased states and whilst CD causes transmural inflammation and can affect any part of the gastrointestinal tract, UC is typified by mucosal inflammation [3]. Although the exact aetiology of IBD remains largely unknown, research indicates immune responses, intestinal microbial flora, genetics and the external environment are all involved in the pathogenesis of disease. In addition to the decline in quality of life, individuals living with IBD face an increased risk of gastrointestinal cancer and despite therapeutic advancements, hospitalisation rates for IBD and CD have been significantly increasing, resulting in substantial socioeconomic costs and burden [4].

Accurate measurement of a patient's diseased state through condition-specific biomarkers is essential in contemporary healthcare to inform strategic therapeutic interventions [5], [6]. CVD- and IBD-indicative biomarkers can offer valuable insights into the biological pathways associated with specific pathologies. This facilitates effective categorisation and classification of diseases, aiding in the guidance of appropriate management strategies [6], [7], [8]. Nevertheless, standard time-to-results for numerous diagnostic and prognostic tests specific to CVDs including,

electrocardiogram, CT coronary angiography, MRI and radionucleotide tests are often greater than a month [9]. These tests are expensive and delayed access provides a barrier to initiation of effective preventive treatments, leading to avoidable mortality. Conventional diagnostic methods often involve invasive procedures such as, coronary angiography in CVD and endoscopy in IBD, which are associated with small risks of heart attack, stroke, infection and discomfort for patients [10]. Thus, there is a pressing need for rapid, reliable, non-invasive diagnostic approaches. The development of a disposable, simple and low-cost bedside device to measure potential biomarkers of early CVD and IBD, from readily available biofluids such as, blood-plasma or saliva, would be of vital importance and a major step towards the determination of an individual's disease state. It would also allow for the accurate measurement of the efficacy of specific clinical therapies and interventions, indicating how an individual is clinically progressing.

In recent years, extracellular vesicles (EVs) have been emerging as promising circulating biomarkers of disease [11], [12]. EVs are lipid bound membranes, released by cells into the extracellular space. They are present in most body fluids, including blood, urine and saliva, and therefore, are easily accessible for analysis [13], [14], [15], [16]. EVs are considered as reservoirs of important information as they hold key evidence relating to the cells from which they originate. Acting as intracellular messengers, their role in the potential spread of disease has further garnered increasing interest [17], [18], [19], [20]. Although still largely unknown, there are multiple routes for generation of EVs, with subpopulations divided into three broad groups based on their size, density and composition, which typically include exosomes,

microvesicles and apoptotic bodies [21], [22]. Exosomes form as a result of the inward budding of endosomal membranes inducing the formation of multivesicular bodies (MVBs). MVBs are subsequently released into the extracellular environment by fusion with the plasma membrane [23], [24]. Upon their release into either the parenchymal space or the urinary space, exosomes can participate in molecular signalling events [25]. Microvesicles, on the other hand, are formed by direct budding from the plasma membrane [26]. Their formation is induced in the presence of a stimulus such as, oxidative stress, which drives cellular events *e.g.,* phospholipid-binding proteins. This in turn, leads to the budding and release of microvesicles from plasma [27]. Apoptotic bodies are released during the latter stages of cell death and contain a multitude of cellular organelles, cytosolic content and nuclear materials [28]. Despite each EV subgroup exhibiting apparent differences in their mechanisms of formation, once secreted from a cell, it is difficult to discern one EV subtype from another. These vesicles do not contain a specific molecular marker or descriptive property distinguishing the sub-groups [29]. The overall potential as key biomarkers has attracted a growing interest in EVs, which carry and protect cellular cargoes, including nucleic acids, proteins and lipids from their place of origin and thus, reflect the physiological and pathological state of their parent cells [30], [31], [32]. The detection and analysis of EVs therefore, offers valuable insights into the disease mechanisms, diagnosis, prognosis as well as response to therapeutics.

EVs are heterogenous in origin, molecular constituents and size, and whilst they are present in an array of complex biofluids, these comprise differing amounts of non-vesicular structures, which may impede analytical results [30]. The characterisation of

exosomes and other types of EVs pose significant challenges beyond establishing consistent parameters for defining each subtype with this task recently becoming more complex due to a growing recognition that proteins (and nucleic acids) loosely associated with the surface of the EVs, once regarded as artifacts of purification play functional roles [33], [34], [35]. Conventionally, EVs are described by their spherical phospholipid bilayer, within which transmembrane proteins such as, tetraspanins are embedded and until recently it was believed this bilayer defined the outer extent of exosomes, with any proteins adhering to the surface post-purification considered contaminants [36], [37]. However, as functional significance is attributed to these proteins forming the EVs corona, the field is undergoing a profound paradigm shift. It is understood the composition of such EV corona is likely to be influenced by the material the EV was isolated from, for example, urine typically contains fewer proteins compared to serum. EVs can either be washed in a biological fluid to remove the corona or washed in an appropriate buffer, such as one containing proteinase K, to eliminate the corona [38], [39], [40]. Moreover, non-EV components often co-isolated (proteins, lipoproteins) can contribute to the formation of the EV corona and thus interfere with downstream analysis. For example, the presence of abundant plasma proteins in the EV corona may mask or obscure the signals of low abundance EV biomarkers, leading to reduced sensitivity and specificity of detection methods [34], [35]. The isolation and separation of EVs is therefore a required pre-analytical step with a variety of analysis techniques used to separate and characterise both EVs and corona proteins. Conventional methods including, for instance, ultracentrifugation, size exclusion chromatography (SEC) and immunoaffinity approaches used to investigate EVs suffer from low throughput times and require expensive equipment, highlighting

the increasing demand for improved analytical methods [41], [42], [43], [44], [45]. Microfluidics has recently been emerging as an additional technique in EV research [46], [47], [48] with the lab-on-a-chip (LoC) based technologies holding promise for overcoming many of the challenges associated with existing EV isolation methods. By leveraging microfluidic systems, LoC devices can provide enhanced control over fluid flow, enabling precise and efficient isolation of EVs [49]. Such control and tunability enables the optimisation of isolation conditions and minimises the formation of non-specific binding events that may contribute to, for example, the EV corona [34]. The miniaturised nature of these devices allows for reduced volumes, rendering them suitable for time-sensitive or limited sample sources. They also allow the integration of functionalised surfaces to selectively isolate EVs based on surface markers or cargo molecules, thereby, reducing non-specific binding events [50], [51]. Moreover, LoC platforms offer added opportunities for automation, standardisation and multiplexed processing, leading to improved reproducibility and scalability of EV isolation protocols[52]. The miniaturisation and integration of quantitative analytical techniques within the chip, such as molecular profiling methods *via* Raman spectroscopy (RS), can enable the further facilitation of the downstream rapid, non-invasive and accurate analysis of isolated EVs.

Raman spectroscopy is a highly sensitive analytical technique, which offers a distinctive spectral fingerprint of target analytes *via* the inelastic scattering of light. Through close examination of spectroscopic bands associated with specific molecular vibrations, RS discriminates between molecular species and provides insights into their relative abundance as well as information regarding the secondary structure of

proteins, confirmation of nucleic acids and the ordering of lipids within the EV membranes, offering overall valuable insights into vesicle composition and organisation [53], [54]. The integration of RS with LoC technology for EV separation, detection and analysis therefore has a significant potential to enable the efficient and high-throughput isolation along with a controlled and reproducible environment for analysis, whilst performing *real-time* and *in-situ* detection of isolated EVs, yields rapid characterisation of their molecular composition.

Herein, we have developed a portable advanced integrated multipurpose spectroscopic lab-on-a-chip (AIMSpec-LoC) for isolation, detection and analyses of EVs *via* specific RS fingerprinting for low-cost, non-invasive, early-stage diagnosis. Design and development of our unique optofluidic Raman lab-on-a-chip device technology for detection of CVD- and IBD-indicative biomarkers enables timely detection of EVs as key-disease indicators, as a platform towards providing transformative triaging and diagnostics. AIMSPec-LoC rapidly extracts EVs from biofluids and detects these as key markers with high-sensitivity, specificity and timeliness. The developed lab-on-a-chip technology combined with Raman spectroscopy subsequently, is validated for the detection of cardiovascular and inflammatory bowel disease indicators, establishing a powerful analytical tool for biomolecular analysis. The acquired spectral data *via* AIMSPec-LoC are subsequently classified using our advanced artificial neural networks (NAA) self-optimising Kohonen index network (SKiNET) algorithm as a decision support tool for easy readouts for clinicians. SKiNET is based on self-organising map (SOM) with a classification *via* the self-organising map discriminant index (SOMDI). AIMSPec-LoC is further shown to

overcome several of the challenges associated with existing methodologies for the isolation of EVs. For instance, field-flow fractionation (FFF) used in EV isolation, where a sample is injected into a chamber is often affected by a cross flow with smaller particles eluting earlier [55], takes one hour to perform, requires significant sample pre-preparation to remove larger cells and debris and can only be operated by highly trained specialists. In contrast, AIMSPec-LoC requires no sample pre-preparation and analyses biofluids in their native state. A further isolation technique commonly applied in microfluidic EV studies is immunoaffinity capture (IAC). IAC relies on the antigenic identification of EV surface components (primarily proteins) and although is rapid and selective when used for either preparation or analysis, one intrinsic drawback of this technique is the requirement for prior knowledge of the surface antigen to be targeted, and which, may be present across multiple subpopulations of EVs [56]. Targeting only a single antigen is further exacerbated by the loss of potentially important EV populations.

Overall, by leveraging the ability of RS to analyse complex molecular signatures, we have identified disease specific EV markers associated with CVD and IBD. Through a comprehensive analysis combined with advanced computational algorithm, our developed novel approach for the accurate and early detection of CVD and IBD, ultimately would lead to improved patient outcomes and more effective disease management. Measuring abnormal changes in specific biomarkers would be indicative of CVDs / IBDs, providing a quantitative evaluation in the initial stages whilst also aiding in quantifying the extent of the damage. It would be interpreted by clinicians as an indication to initiate treatment before advanced disease has developed. In the long-

term, this platform could be extended to further applications *e.g.,* early-recognition of neurodegenerative diseases, degeneration and timely interventions specifically targeting the identification of CVDs and IBDs in otherwise healthy individuals. The integration of simultaneous EVs isolation with Raman spectroscopy holds great promise for disease diagnostics. It can improve the discovery of future cardiovascular and IBD markers for early-diagnostics and intervention as well as advance the understanding, prevention and treatment of CVDs, IBDs *etc*. and the associated detrimental conditions.

7.3 Results and Discussion

A unique lab-on-a-chip device capable of isolating and separating extracellular vesicles as key targets of cardiovascular and inflammatory bowel diseases has been developed, offering a highly versatile application with faster analysis times compared to conventional methods such as ultracentrifugation. The miniaturised design, eliminates the need for multiple separation steps, reduces the required sample volume and enables the integration of isolation and downstream analysis onto a single platform (**Fig.7.1a**). The portability and compactness of the LoC is well suited for point-of-care applications, allowing for their deployment in various clinical settings to provide real-time diagnostics.

The AIMSpec LoC was designed and engineered to enable sufficient isolation of various populations of EVs, including exosomes, microvesicles and apoptotic bodies within size ranges of 40-100, 101-200, 201-500 and 501-5000nm, respectively. In addition, the platform was designed to facilitate isolation and separation of up to 1mL of sample at final concentrations ranging from $1x10^6$-$9x10^8$ EVs from an array of

biofluids, while simultaneously allowing the separation of EVs into respective size-based subgroups. Based on guidelines set out in the MISEV2018 EVs can be categorised as either small (<200 nm) or medium/large EVs (>200 nm) according to their size. Here, we divide our EV subgroups into small (<100 nm), medium (100 – 250 nm) and large (>250 nm) particles. The unprecedented, generated ability of the LoC to analyse various vesicles in parallel from multiple body-fluid sources provides important and valuable insights into the tissue-specific release of EVs and their potential roles in different diseases and biological processes. **Figure 7.1** provides a comprehensive overview of the experimental setup, illustrating the integration of the designed LoC, the biogenesis of EVs with the key components of the methodological workflow (**Fig.7.1a**) including, the sample collection, on-chip isolation, separation and subsequent Raman analysis with an AI algorithm data classification, all of which were designed and engineered to achieve optimal outputs of speed, specificity and accuracy. Simultaneously, it highlights the cellular processes involved in the biogenesis of different EV subgroups, *e.g.,* the outward budding of EVs from the plasma membrane and their release into the extracellular space (**Fig.7.1b**) as well as representative size-based differences of each group (**Fig. 7.1c**). Typically, micro-platforms focus on the isolation of specific EV subtypes, either exosomes or microvesicles and incorporate capture mechanisms based on immunoaffinity or surface-marker selection. Here, we have uniquely fabricated LoC device capable of a broad-spectrum EV isolation, designed to capture a wide range of subtypes, collectively enabling not only a comprehensive analysis of the EV population present in the sample but also, allowing for a more inclusive and unbiased characterisation of EVs and their cargo.

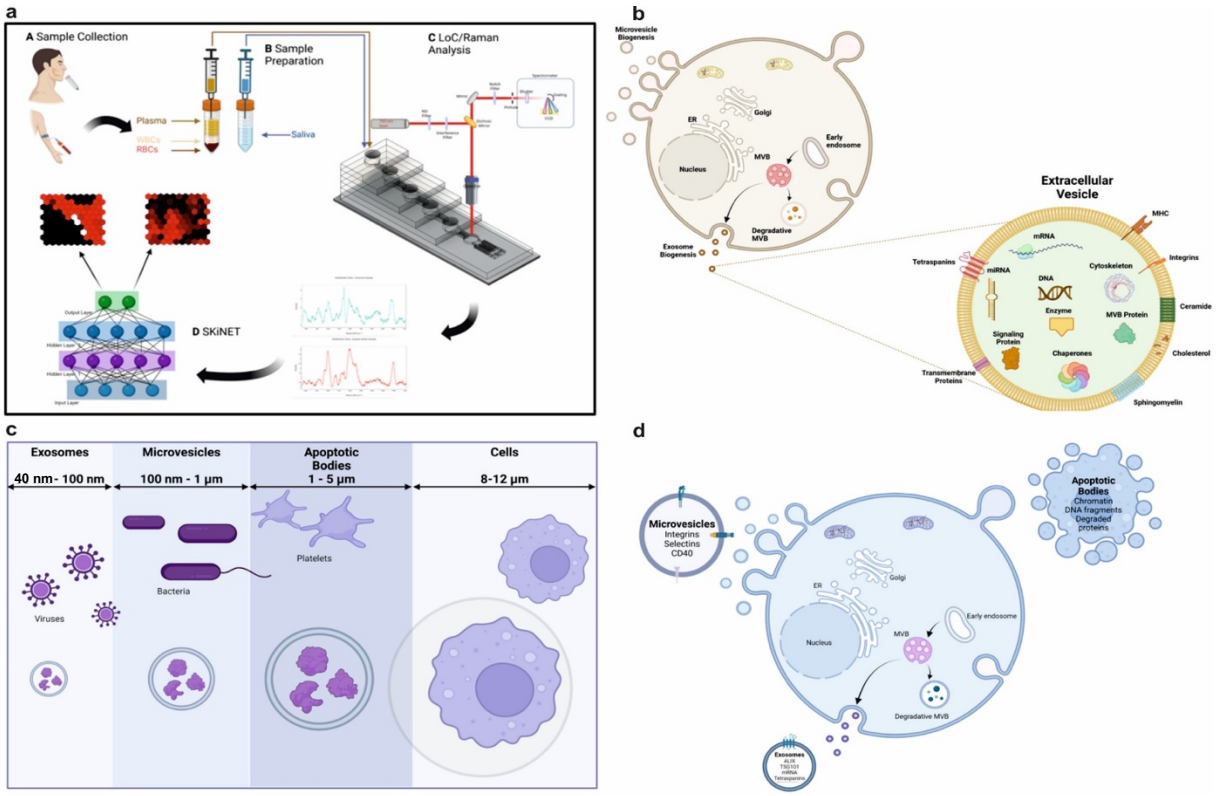

Figure 7.1. Overview of the EVs Diagnostic Principles. (a). Schematics of LoC methodology including the (A) sample collection from biofluids *e.g.,* saliva and blood, followed by (B) sample preparation to isolate plasma *via* centrifugation and dilution of saliva with (C) on-chip isolation, separation and the subsequent Raman detection of EVs obtained from saliva and blood plasma and (iv) rapid analysis *via* artificial neural network algorithm, SKiNET, acting as a decision support tool in the classification of acquired data, output in both the diseased state as well as subgroup of isolated EVs (D). **(b).** EV biogenesis illustrating the formation of exosomes *via* the inward budding of multivesicular bodies (MVBs) within the cell, resulting in the release of intraluminal vesicles as exosomes when MVBs fuse with the cell membrane and microvesicle formation through the direct outward budding or shedding of the plasma membrane. Composition and structure of EV illustrating phospholipid bilayer surrounding protein (signalling protein, MVB protein) and nucleic acids (DNA and RNA). Membrane proteins including tetraspanins, MHC, receptors, and adhesion molecules (integrins). **(c).** Typical size ranges of the three main EV subgroups compared with cells. Apoptotic bodies range from 1-5µm, smaller microvesicles range from 100 to 1000nm, overlapping with bacteria and exosomes are commonly

defined as vesicles with sizes in the range of 40-100nm. **(d).** EV formation highlighting the main contents and surface markers of each vesicle type. Exosomes contain ALIX, TSG101, mRNA and tetraspanins whilst integrins, selectins and CD40 are characteristic of microvesicles and apoptotic bodies are primarily composed of chromatin, DNA fragments and degraded proteins. EV images created in Biorendor.com from EV template available. (Figures created in Biorendor.com, adapted from EV formation template).

The LoC is designed using a polyethersulfone nanoporous membrane (PES) based on a size-based exclusion and filtration approach (**Figure 7.2**), fabricated to separate EVs from biofluids, within which a filtered plasma is acquired from 50µL of blood in <10 minutes with a similar performance enabled for the filtration of 50µL saliva (1:2v/v in PBS) in <15 minutes. Comprehensive characterisation of the AIMSPec-LoC, essential in the assessment of chip performance and its suitability for the efficient isolation and separation of EVs, incorporated multiple techniques including, nanoparticle tracking analysis (NTA), optical microscopy, fluorescence microscopy and dynamic light scattering (DLS), in accordance with the international society for extracellular vesicles (ISEV) recommendations of at least two characterisation methods to be used for EVs. The combined use of these analytical techniques provides a systematic evaluation of the LoC functionality, EV size distribution and capture efficiency as well as highlights the robustness and an overall performance of the chip in EV isolation and separation, providing further insights for its potential straightforward integration into existing healthcare pathways as a diagnostic modality. The engineered LoC consists of six adhered PDMS layers, with the top layer serving as the primary sample inlet (**Fig.7.2, a-d**). Each subsequent layer consists of a further sample inlet and a sample collection well. The PES membranes, of varying sizes (40, 100, 250, 500 and 5000nm) act as

filters placed between each PDMS layer at the base of the sample collection wells. Capillary channels (0.05 x 0.05mm^2) are placed in the base layer to aid fluid flow through the chip with the porous membranes in constant fluid contact. Whilst particles or EVs larger than the pore size are unable to pass through the membrane and trapped in their respective sample well, smaller particles / EVs pass through the membrane due to the faster diffusion. An additional membrane (pore size 40nm) is included in the design to trap proteins (including lipoproteins) and reduce contamination/co-isolation (**Fig. 7.2d**). The overall membrane placements and collection wells were designed to maximize the recovery of samples. The size distribution and concentration of EVs captured *via* the engineered LoC has been subsequently analysed *via* NTA whilst leveraging the Brownian motion of the particles in solution [57]. The representative size distribution profile of EVs isolated has been found to be in three main ranges of 41-100nm, 101-250nm and 251-500nm, (**Fig.7.2j**). Corresponding well-defined size distributions with mean diameters of 61±23nm, 189±84nm and 432±143nm, respectively, indicate a successful isolation of EVs within the desired ranges [25], [27]. The average EV concentrations were found to be 7.7x10^8 and 2.2 x10^8 particles/mL in each isolated sample well, confirming the efficient capture capabilities of the pore size-based separation of plasma based EVs, indicative of CVD. Further, size distribution of the larger isolated particles determined *via* DLS reveals distribution profile of 501-5000nm with a population consistent with the large particles [28] (**Fig.7.2h**).

To further assess and evaluate the specific size-based capturing of the EVs, fluorescent polystyrene beads ranging from 5µm to 30nm were used to represent the EVs dimensions for each studied subpopulation. Representative fluorescence

microscopy images of the size-based distribution show the fluorescent particles flowing through the microfluidic channels (**Fig.7.1, i-j**) exhibiting intense fluorescent signal corresponding to the specific sample labelling *i.e.,* PS beads in the range of 41-100nm (red), 101-250nm (blue) and 251-500nm (green). This indicates specific capturing of the target ranges with negligible sample signal crossover and an overall separation efficiency of 92% (calculated from the total number of beads collected across all filtration membranes divided by the number of those gathered from a single filter membrane), demonstrating the capability of the LoC to selectively capture size based EVs, delivering a robust platform for downstream analysis and rapid molecular profiling. Optical microscopy images of the nanoporous membrane structure (**Fig.7.2j**) showing the surface and internal features to determine pore size, shape and distribution, further enable identification of defects or irregularities as well as *real-time* monitoring to eliminate blockages. SEM images of the EVs in solution, captured on surface of the nanoporous membrane from each of the filtration steps (**Fig.7.2k**), further confirm the EV subpopulations isolated with size ranges of 40-100nm and 101-200nm and average EV sizes of 61±3nm and 189±84nm, respectively, in correspondence with the NTA analysis.

Important to note, the fabricated chip enables the label-free isolation of EVs and their associated subpopulations *via* size-based and capillary functions by applying *g*-force to the bioparticles, which is several orders of magnitude lower than the typical forces applied in conventional ultracentrifugation, reducing the overall particle aggregation and damage due to shear stress. The LoC, as a passive cataloguing technique, exhibits further advantages over traditional methods such as ultracentrifugation and

immunocapture, including the overall lower cost of the system, requiring no additional external force fields in contrast to syringe pumps routinely used in EV microfluidic systems [50], nor does it require antibodies to capture the EVs. The cost-effective engineered design is based on gravity forces combined with capillary channels, enabling high throughput fluid flow through the LoC, yielding a smaller, more compact chip than other microfluidic devices [58], [59]. Typical blockage issues, often associated with microfluidic systems as well as channel deformations are avoided within the AIMSPec-LoC due to the larger channel dimensions and overall lower flow rate. The ability to isolate EVs directly from saliva and blood plasma offers several key benefits for use in clinic or at bedside including, rapid time-to-results (<20 mins), portability (75 x 25 x 50mm^3) and high accuracy (AUC of 0.95-1). It also enables ease of deploy ability in remote or resource limited settings for timely diagnosis and treatment even in areas with limited access to laboratory equipment, automation and minimal user intervention reducing the potential human error, lower cost per test with reduced consumption of reagents as well as inherent scalability for future mass production. Finally, the unique ability of the designed LoC to isolate EVs from various biofluids, such as saliva and blood plasma, enables multiplexed assay detection. Whilst the microfluidic devices typically are applicable for a single biofluid for the EVs capture [59], [60], [61], AIMSPec-LoC platform is suitable for simultaneous multi-biofluid EV capture and isolation, to provide a superior overview of a patient's health status whilst reducing the requirement for multiple costly tests.

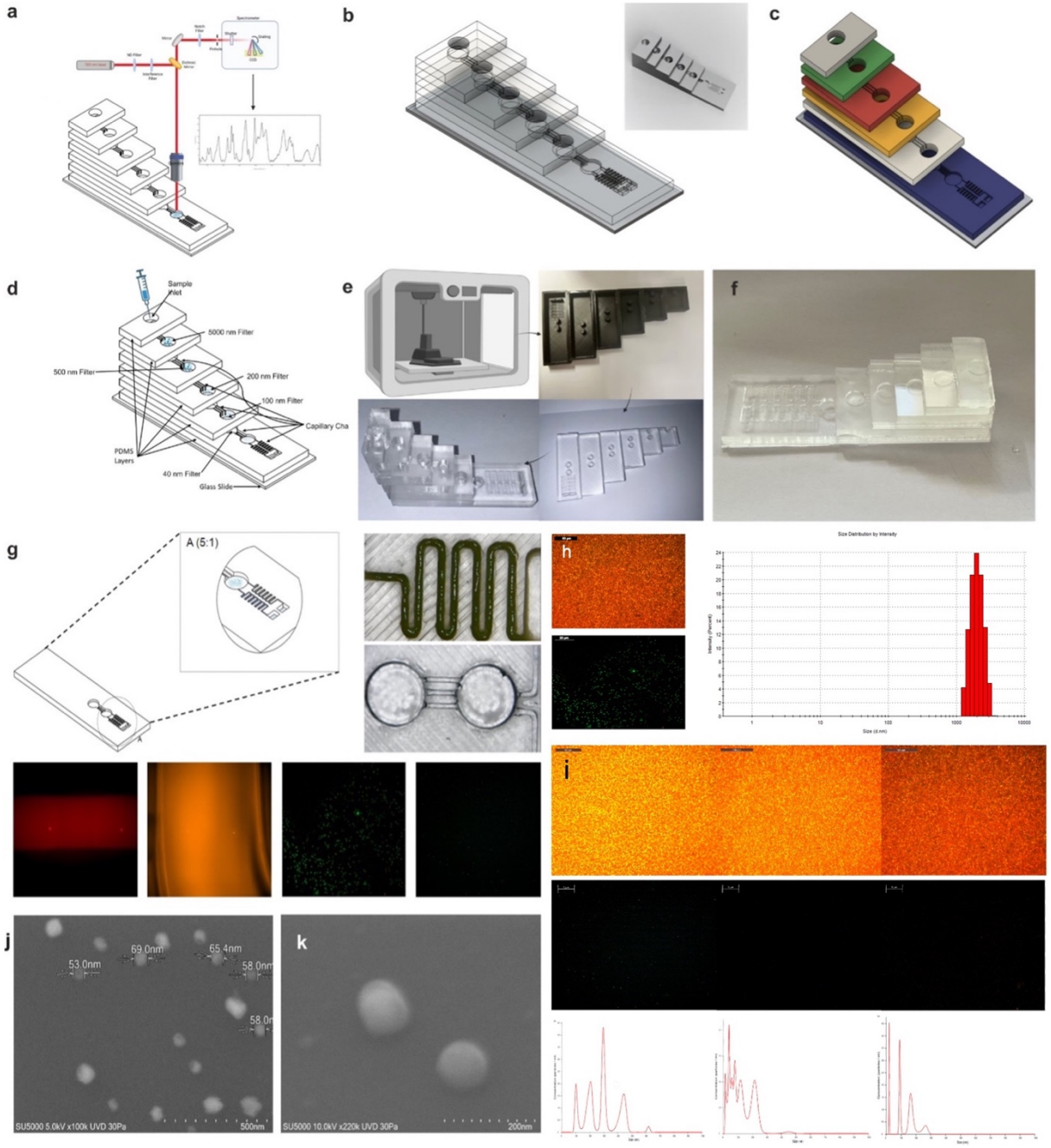

Figure 7.2. Fabrication and Characterisation of AIMSPec-LoC for the Isolation and Separation of EVs. (a). Schematic illustrating overall chip design complete with integrated downstream analytical EV Raman analysis. **(b).** Transparent 3D rendered design of the engineered LoC with embedded sample inlets and outlets within each layer with **(inset)**. **(c).** Colour-coded rendered LoC design with each layer, associated with a specific EV size range, highlighted in a different colour. **(d).** Detailed annotated diagram of the LoC showing the sample inlet in the first PDMS layer with the 5000, 500, 200, 100 and

40nm filters in subsequent layers and the capillary channels, aiding fluid flow within the system. **(e).** Overview of the LoC fabrication including, 3D printing (top left), PDMS casting (top tight and bottom right) and the 3D-assembly (bottom left), yielding the **(f)** prototype AIMSPec-LoC model. **(g).** Schematics of the capillary flow within the LoC base layer (left) with optical imaging of capillary channels (top right) and microfluidic channels (bottom right) connecting sample inlet and sampling well. **(h).** Optical microscopy image (top left) of the 500nm pore size nanoporous filter with the corresponding fluorescence microscopy image (left bottom) of collected 500-5000nm EVs and DLS measured *via* Panalytical Zetasizer HPPS (right handside) exhibiting abundance of nanovesicles in the range of 1000-5000nm. **(i).** Fluorescence microscopy images of the movement of fluorescent beads, acting as visual proxies for EVs within the microfluidic channels and their associated collection in the sampling well, demonstrating the separation efficiency of the LoC. Nanoparticle tracking analysis measured (NanoSight N300) demonstrated abundance and size distribution of isolated EVs for each filter size. **(j).** SEM images of the 100nm nanoporous membrane filter indicative EVs. **(k)** Representative SEM image of isolated EVs.

The overall separation efficiency of the LoC of 92%, confirmed by fluorescence microscopy combined with NTA (**Fig 7.2g** and **i**) is found to be significantly higher than size exclusion chromatography of 67.7±13.1% as well as than the polymer-based precipitation ~82% [62]. This separation efficiency can be further increased following dilution (1:10) of the tested biofluids however, this would be on the expense of the throughput and lower sample concentration for further downstream analysis. Our AIMSPec-LoC efficiently processes up to 500µL of saliva within 20 minutes (~1.5mL/hour), increasing to 650µL (~1.95 mL/hour) for blood plasma (**SI Figs.1-2)**. Although this is slower relative to microfluidic approaches such as ExoDFF, our system does not rely on any external forces for EV isolation [63] and has a unique ability to simultaneously isolate different size groups in a single step. This contrasts with 'gold-standard' of ultracentrifugation for EV isolation, which is unable to separate several EV

groups concurrently and requires further density gradient to reach higher purification [64]. Furthermore, to tackle the common issue with EV isolation of the co-elution contaminating lipoproteins and plasma proteins [65], we employed low protein binding PES filter membranes, successfully trapping the contaminating proteins *via* the 40nm filter as well as the lipoproteins including LDL with a size range of 18-35nm. Whilst traditional techniques such as size exclusion chromatography (SEC) are effective at removing proteins, they are unable to efficiently remove lipoproteins, requiring density gradient ultracentrifugation to aid in the exclusion of these contaminants and thus, the combined approach suffers from significant losses of EVs [66], [67]. In the AIMSPec-LoC, the hybrid isolation and spectroscopic detection enable the multiplexed ability whilst removing any signal associated with contaminating proteins or lipoproteins. Furthermore, although size-exclusion methods are commonly used in the isolation of EVs, they often suffer from clogging and trapping in the membrane, poor efficiency and deformation of EVs due to pre membrane pressure [68], [69]. These drawbacks are circumvented in the AIMSPec-LoC *via* the introduction of sequential filtration with the initial step using a large pore filter (5000nm), followed by sequential filtration steps through smaller pore filters (450, 200, 100 and 40nm). Additionally, removing the need for external syringe or centrifugal pumps and including capillary channels to propel the flow through the system overcomes pressure associated destruction of EVs.

Identifying and characterising different types of EVs can be challenging due to several different factors, including their small size, heterogeneity, and the lack of standardised isolation and characterisation methods. Due to the overlap in size between the two main types of EVs, microvesicles and exosomes, it is difficult to differentiate them

based solely on size. Therefore, the identification of biomarkers associated with EVs can significantly aid in their identification and characterisation. Biomarkers for example, provide specific characteristics that differentiate EVs from other particles or contaminants in biological samples as certain proteins, lipids, or nucleic acids may serve as unique markers for different EV types (e.g. CD9, CD63, CD82). As EV subgroups have distinct biomarker profiles, identifying these biomarkers enable the classification of EVs thus providing valuable information as to their cellular origin, biogenesis, and functional roles. In addition, changes in the composition of EV biomarkers are associated with various diseases. Identification of these specific markers can have diagnostic, prognostic, and therapeutic implications, with current research exploring the potential use of EV biomarkers for the early detection and monitoring of diseases such as cancer, neurodegenerative disorders, and cardiovascular diseases. To further validate the isolation of differing EV subgroups, a panel of known EV biomarkers were analysed (**Fig 7.3/SI Fig.S3)** to obtain their Raman spectral fingerprints in order to compare their signatures with those EVs isolated and separated with AIMSPec-LoC.

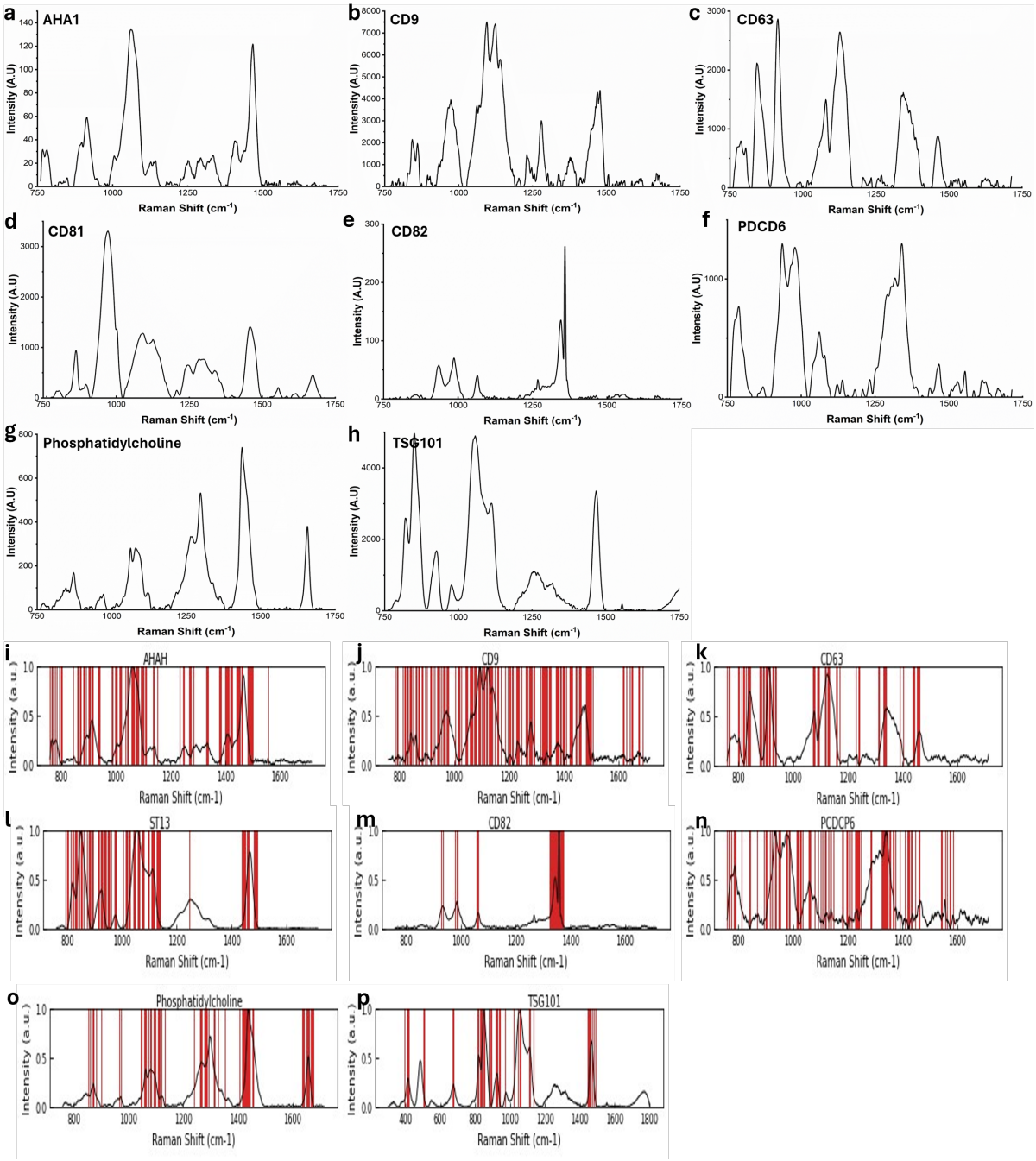

Figure 7.3. EV indicative biomarkers **(a)** AHA1, **(b)** CD9, **(c)** CD63 and **(d)** CD81 **(e)** CD82 **(f)** PCDC6IP, **(g)** Phosphatidylcholine and **(h)** TSG101. Molecular barcoding of corresponding biomarkers of EVs **(i)** AHA1, **(j)** CD9 and **(k)** CD61, **(l)** CD81, **(m)** CD82, **(n)** PCDC6IP, **(o)** Phosphatidylcholine and **(p)** TSG101.

Multiple EV surface markers have been reported and are routinely used in the identification of extracellular vesicles thus, biomarkers including tetraspanins CD9, CD63, CD81 and CD82, programmed cell death 6 interacting protein (PCDC6IP) as well as tumour susceptibility gene 101 protein (TSG101), phosphatidylcholine, heat shock protein 70 and heat shock protein 90 were herein evaluated. Although present in both small and large EVs CD63, CD9 and PCDC6IP are routinely found enriched in exosomes and compared to microvesicles [70], [71]. Phosphatidylcholine was also analysed as a key indicator of EVs with reports suggesting phosphatidylcholine composes 46-89% of the lipid components in all exosomes [73]. Heat shock proteins, considered intracellular proteins can shuttle between the cytoplasm and the nucleus and although typically not identified on the surface of healthy cells, heat shock proteins have been indicated as a constitutive EV protein [74]. From the identified biomarkers, significant peaks of interest ($p^{**} < 0.01$) were identified at Raman shifts of 785, 960 and 1448 cm^{-1} for the small particles, 1250-1350 and 1656 for medium particles and 1003 and 1120 cm^{-1} for large particles.

Micro engineered AIMSPec-LoC has been subsequently validated *via* analysing clinically relevant samples. Biochemical fingerprints of EVs derived from blood plasma and saliva correlating with CVD were investigated *via* Raman spectroscopy. Through the subsequent SKiNET examination of their spectral shift, generated from the molecular vibration frequencies of specific molecules including amino, fatty and nucleic acids, the disease state of the patients has been determined. The clinical blood plasma and saliva were collected as part of the CASCADE and TRAFIC studies (Ethics Ref. 19/SW/0010), providing samples from CVD and healthy control subjects (**Table 10.1**)

with the consequent EVs collected from the LoC directly detected and analysed *via* Raman spectroscopy (10x10μm^2 sample area, 785nm laser). Overall, a total of 4,000 measurements were collected from blood plasma samples of CVD group (*n*=20), 20 healthy volunteers acting as a control group and 4,500 saliva samples from 41 subjects including IBD (*n*=21) and healthy controls (*n*=24) (**Methods**).

Advanced ANN SKiNET analysis on the EVs collected from blood plasma from CVD patients revealed clearly separated clusters, recapitulating the fine spectral differences of the molecular fingerprinting spectra, effectively discriminating the three EV subgroups from raw blood plasma (**Fig 7.4c**). Combination of RS with SKiNET to investigate the disease state of a patient, identifies which spectral features, equivalent to the biochemical changes, are responsible for clustering and rapidly discriminates disease, *e.g.,* CVD, from a pool of healthy control groups.

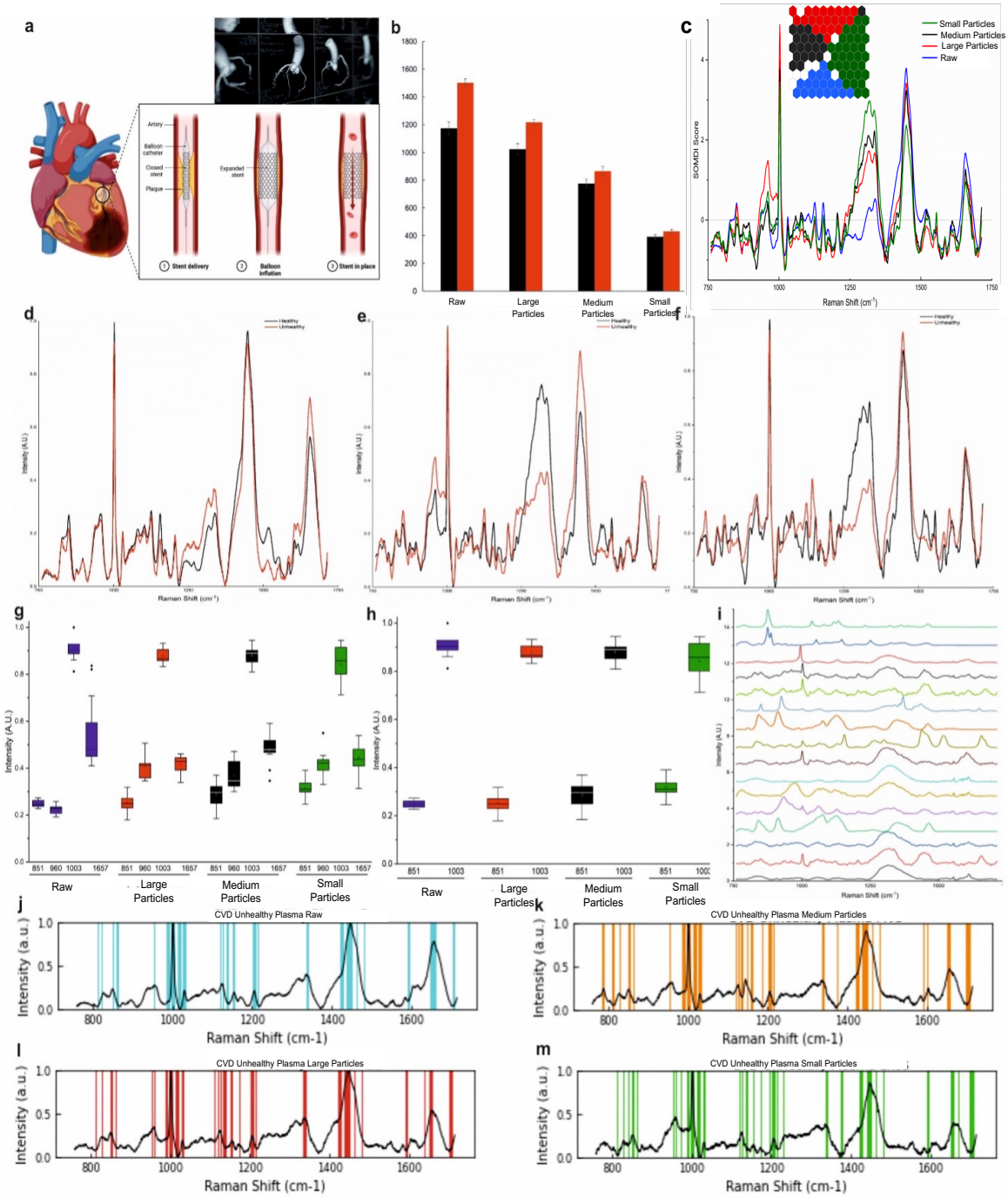

Figure 7.4. Spectral Detection and Differentiation of Cardiovascular Disease from Blood Plasma EVs. (a). Illustration of damage sustained to the heart after myocardial infarction with typical stenting procedure shown. Coronary angiogram **(inset)** typically used for clinical CVD diagnosis. **(b).** BCA protein assay indicating differences in protein concentration at each filtration step indicative of 3 main subgroups: large particles, medium particles, small particles and raw blood plasma. **(c).** SOM

(inset)/SOMDI of EVs classified with an accuracy of 96.5±0.4%. Sub classification according to EV subgroup are represented by blue hexagons for raw blood plasma, red for large particles, black for medium particles and red for small particles. Representative average Raman spectra of EV subgroups of healthy (black) and CVD (red) of **(d)** large particles, **(e)** medium particles and **(f)** small particles. **(g-h).** Box and whisker plots represent the minima, maxima, interquartile ranges, whiskers and the median in peak intensities identified *via* SOMDI as having the greatest effect on the classifier, *i.e.,* the spectral changes of greatest importance with the largest changes observed of 851, 960, 1003 and 1657cm^{-1}. **(i).** Representative Raman spectral fingerprints of a range of CVD-indicative biomarkers of interest, including apolipoprotein B (ApoB), lipoprotein A (LpA), proprotein convertase subtilisin/kexin type 9 (PSCK9), C-reactive protein (CRP), d-dimer, interleukin-9 (IL-9), myeloperoxidase (MPO), tumour necrosis factor-alpha (TNF-α), copeptin, galectin-3, growth and differentiation factor-15 (GDF-15), matrix metalloproteinase 9 (MMP9), N-terminal pro-B-type natriuretic peptide (NT-proBNP), suppression of tumorigenicity 2 (ST2), soluble interleukin-6 (SIL-6) and atrial natriuretic peptide (ANP). Characteristic molecular barcodes derived from the corresponding average fingerprint spectra with spectral differentiation between EV subgroups of **(j)** raw blood plasma, **(k)** medium particles, **(l)** large particles and **(m)** small particles.

Subsequently, we carried out EVs profiling to obtain molecular-level information including the composition of EVs as well as identifying biomolecules specific to EVs *e.g.,* lipids, proteins and nucleic acids. When CVDs occur clinically, the metabolites and protein contents in the blood plasma and saliva increase and the spectroscopic peaks indicative of these change in correlation within the detected spectra. The intensities of the primary peaks of interest at 960, 1003, 1136 and 1657cm^{-1} could subsequently, be used not only for the differentiation of the CVD but also, to monitor the progress post cardiac or inflammatory event since the ratio of characteristic peaks changes depending on the pathological status (**Fig 7.4a**), visible from the obtained stenting. We used the hybrid AIMSPec-LoC with AI to assess the patients' profiling

data and analyse the ability to differentiate CVD from circulating EVs in blood plasma, whilst identifying the underpinning biochemical changes associated with the disease (**Fig 7.4d-f**). From the SOM/SOMDI we determined the intrinsic classification accuracy of 97.1% for the large particles, 98.1% for the medium particles and 100% for the small particles, clearly discriminating between CVD and healthy group (**Table 7.1**).

Table 7.1. CVD Disease Differentiation *via* EVs. Classification performance and accuracy of healthy and CVD patient in blood plasma compared to healthy controls including sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV).

Differentiation	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
Healthy *vs.* CVD Raw blood plasma	95.2 ± 0.7	100 ± 0.4	100 ± 0.3	95 ± 0.7	97.5 ± 0.2
Healthy *vs.* CVD large particles	97.4 ± 1.1	96.8 ± 0.5	96.8 ± 0.9	97.5 ± 0.9	97.1 ± 0.5
Healthy *vs.* CVD medium particles	97.5 ± 0.9	98.7 ± 0.6	98.7 ± 01.2	97.5 ± 0.9	98.1 ± 0.3
Healthy *vs.* CVD small particles	100	100	100	100	100

To evaluate the discrimination of the differing EV populations within blood plasma samples of patients post cardiac event, the acquired spectra were barcoded *via* the dominant peaks of the highest intensity and spectral differences, comparing between the CDV and control cohorts (**Fig 7.4 j-m**). Further to the average spectra the most prominent SOMDI extracted (**Fig 7.4c**) spectral features, responsible for the clustering observed in SOM (**Fig 7.4c, inset**) are 759, 785, 882, 960, 1003, 1124, 1131, 1307, 1336, 1440, 1600 and 1657 cm^{-1}. With the detailed biochemical attributions of these

peaks summarised in **Table 7.2**, the most prominent characteristic changes are attributed to the nucleic acids (720-820cm^{-1}), phenylalanine (1003cm^{-1}), glycine backbone and proline side chain (1341-1348cm^{-1}) and lipid and protein markers of the CH and CH_2 groups (1440cm^{-1}), whereas proteins, lipids and amino acids dominating the overall spectra, in correspondence with the literature on the varying cargo within the EV subgroups [75], [76], [77], [78], [79].

Further key differences provide insights into the separation of detected subpopulations as well as the raw blood plasma of CVD patients **(Fig 7.4c/SI Figs.S4-5)** with the major intensity differences at 960cm^{-1} (polysaccharide structure), 1003cm^{-1} (phenylalanine), 1341 m^{-1} (glycine backbone), 1440cm^{-1} (lipid and protein) and 1656cm^{-1} (Amide I/lipid). For the filtered EVs, the final fraction containing the small particles yields a significantly lower peak intensity at 1440cm^{-1} yet an increased intensity at 1341 cm^{-1} (p***<0.0001), attributed to glycine backbone, while each of the three EV subpopulations found to have less intense lipid and protein peaks, *i.e.,* 1440 and 1656cm$^{-1,}$ than raw blood plasma. A significant decrease is also measured in protein concentration at each subsequent filtration step in the LoC, pertinent to raw blood plasma, large, medium and small particles (**Fig 7.4b**). Blood plasma is known to contain a wide range of proteins including albumins, clotting factors, enzymes and hormones as well as various other proteins involved in immune response, transport, and regulation [80]. On the other hand, EVs protein content varies depending on the cell type and conditions and these are known to carry a subset of proteins derived from their cells of origin with their cargo involved in intercellular communication and cell-to-cell signalling. While EVs can contain a diverse range of proteins, they are typically present at much lower

concentrations compared to blood plasma [81] (**Fig 7.4b**). Amongst the three EV subpopulations, small particles exhibit the least strong peak attributed to proteins whilst the large particles, the most intense. Large particles (including apoptotic bodies) are released during programmed cell death, *i.e.,* apoptosis, and are larger in size compared to the other EV subgroups and contain a diverse range of proteins derived from the dying cell [28], [82]. Small particles (including exosomes), being considerably smaller and formed by the inward budding of multivesicular bodies, also contain further proteins including, membrane and cytosolic proteins. However, they generally have lower protein levels compared to large EVs (such as apoptotic bodies) due to their size and the selective packaging of cargo during their biogenesis, corresponding to the lesser measured protein content [82], [83], [84].

Determining the total protein concentration of a specific subpopulation is further useful in assessing the separation efficiency of the LoC designed for EVs isolation. A higher protein concentration in the retained fraction indicates better capturing and enrichment of the subgroup and thus, a higher efficiency of the LoC device in isolating a specific EV population. Conversely, if the protein concentration is similar in both fractions, it indicates limitations in the separation efficiency for the targeted subpopulation. In the AIMSPec-LoC, as the vesicles passed through the LoC platform, the overall protein concentration decreased, indicating effective size-based separation. To address a challenge associated with EV isolation and separation due to contaminations, the LoC was designed to allow contaminating proteins (including lipoproteins) to flow through and since the majority are smaller than EVs, these contaminants are trapped by the 40nm filter. This results in successful separation of contaminants since the overall

protein concentration has decreased during the blood plasma passing through each filtration step (**Fig 7.4b**), in agreement with the 92% separation efficiency calculated *via* a combination of fluorescence microscopy and NTA (**Fig 7.2i**).

Table 7.2. Raman spectroscopy detected assignments of the dominant, characteristic spectral peaks [75], [76], [77].

Peak (cm^{-1})	Assignment
759	Tryptophan -ring breathing vibration
720-820	Nucleic Acids
830-853	Tyrosine doublet (protein)
875, 882	Phosphatidylcholine
920	Phosphate deformation and bending
960	Polysaccharide Structure
1003	Phenylalanine, carotenoids
1045	Proline
1124	Lipids
1131	C-N stretching of proteins
1125	Myoglobin (haem core), phospholipids, proteins
1307	Phospholipids, lipids, adenine, myoglobin (haem core)
1341-1348	Glycine backbone and proline side chain
1368	Tryptophan, guanine, thymine, myoglobin (haem core)
1440	Lipid and protein
1450	CH_2-deformation
1472	CH_2 bending of lipids and proteins
1550	Amide II/Protein
1553	Tryptophan
1656	Lipid/Amide I
1657	C=C stretching in lipids
1658	(v(C = O)), Amide I and lipids

A further statistically significant difference is found for the dominant bands at 1124, 1440 and 1656cm^{-1} (p***<0.0001) associated with lipids at decreased intensities in EVs compared to raw blood plasma, with exosomes exhibiting the lowest levels (**Fig 7.4g-h**). Blood plasma contains a variety of lipids, including triglycerides, cholesterol, phospholipids and fatty acids, which are essential for physiological functions including

energy storage, membrane structure and signalling [80]. In contrast, when EVs are isolated, the specific method employed to separate the vesicles from the bulk of the blood plasma typically removes lipids which are not associated with EVs, resulting in a relatively lower lipid content in isolated EVs compared to the raw plasma [85]. Bands at 882cm^{-1} (phosphatidylcholine) and 1553cm^{-1} (tryptophan) exhibited significantly increased intensities in EV subgroups relative to the raw blood plasma. Combined with the changes observed in protein and lipid peaks (*i.e.,* 1124, 1440, 1656cm^{-1}), these bands could be employed as spectral markers for the EV characterisation, which would also aid in the development of further studies to assess the clinical utility of the developed technology.

Disease indicative EV spectral differences are found to vary significantly with the representative Raman spectra of healthy and CVD derived EVs for each subpopulation exhibiting major intensity changes (**Fig 7.4d-f**). Both, the medium and the small particles, exhibited significant intensity increases (~17-68%) ($p^{**}<0.001$) in CVD group at bands of 752cm^{-1} (nucleic acids), 851cm^{-1} (tyrosine), 960cm^{-1} (polysaccharide structure), 1144cm^{-1} (lipids), 1440cm^{-1} (lipids and proteins) and 1656cm^{-1} (Amide I/lipids) with an additional increase at 1003cm^{-1} for small particles. The increased intensities of nucleic acids (**Fig 7.4e-f**) can be attributed to the EVs' representing an important intracellular communication mechanism through the containment and transportation of bioactive molecules such as, proteins and micro-ribonucleic acids (miRs) to target cells [17], [86], [87]. EVs therefore, pass on favourable, neutral or harmful effects on recipient cells, which include modulating gene expression and affecting molecular pathways. In the case of cardiovascular diseases, EVs can be

released by cardiovascular-related cells such as, platelets, monocytes, leucocytes and cardiomyocytes[11] and are capable of inducing pathological changes associated with CVDs as the composition of proteins and miRs transferred by EVs, which maintain cardiovascular balance, can be altered thus, giving rise to the development of CVDs [88]. Patients with CVDs often exhibit increased levels of nucleic acids circulating in their blood, resulting in packaging of these within the EVs and subsequent release into the blood stream [89].

The detected increase in intensity of the amino acid tyrosine in CVD-derived EVs would indicate an associated increased risk for CVD. Tyrosine is the precursor of catecholamines, which cause high blood pressure when found at increased levels in the body [90]. Previous studies have shown that tyrosine causes tachycardia and hypertension in small doses, whereas in larger doses the opposite was observed *i.e.,* bradycardia and hypotension [91]. Closely associated with the effect of elevated tyrosine levels in CVDs, is the increased concentration of phenylalanine. In two of the investigated EV subgroups, we have found both peaks at 1003cm^{-1} associated with phenylalanine as well as the 851cm^{-1} associated with tyrosine are increased in CVD patients. Inflammation and immunological activation are directly linked to the onset and progression of CVDs, with indications that they furthermore hamper the conversion of phenylalanine to tyrosine in individuals with clinical disorders. Intrinsically, increased phenylalanine concentrations and a higher phenylalanine-to-tyrosine ratio have been linked to immunological markers [92] thus, suggesting a potential spectral biomarker of CVD. The observed differences in intensity of each EV subgroup compared to raw blood plasma for both tyrosine and phenylalanine indicate the different biochemical

composition of EVs isolated *via* the LoC. Whilst amino acid levels in plasma have been used for early detection and diagnosis of diseases such as cervical cancer, pancreatic cancer and type 2 diabetes, differing amino acid levels in EVs have only been reported in small particles (such as exosomes), analysed with high performance liquid chromatography with fluorescence [93], where distinct differences were observed between small particles (such as exosomes), and native serum amino acid compositions. Of a particular interest, in the study by Onozata *et* al. histidine, arginine, glutamic acid, cysteine, lysine and tyrosine were found to be significantly increased ($p^*<0.05$) small particles (including exosomes), indicating that certain amino acids are enriched in small particles (including exosomes). We have similarly identified the most significant increase in amino acids (tyrosine and phenylalanine) in the fraction associated with small particles (including exosomes), in agreement with the study by Onozato *et* al. [93]. In addition, the 1553cm^{-1} peak associated with tryptophan has exhibited a significantly decreased intensity ($p^*<0.05$) in CVD blood plasma in each of the three EV subpopulations. Tryptophan breakdown into downstream metabolites is accelerated at the onset of CVD *via* interferon-*γ*-mediated inflammation [94]. Yu *et* al. determined that a lowered risk of CVD was strongly correlated with an increase in plasma tryptophan content, in line with the findings presented here, whereby tryptophan is inversely associated with the incidence of CVD [95].

To further categorise and determine disease-specific molecular markers within the EVs, RS was used to profile and classify a panel of CVD-indicative biomarkers. The identified panel of CVD biomarkers including apolipoprotein B (ApoB), lipoprotein A (LpA), proprotein convertase subtilisin/kexin type 9 (PSCK9), C-reactive protein

(CRP), d-dimer, interleukin-9 (IL-9), myeloperoxidase (MPO), tumour necrosis factor-alpha (TNF-α), copeptin, galectin-3, growth and differentiation factor-15 (GDF-15), matrix metalloproteinase 9 (MMP9), N-terminal pro-B-type natriuretic peptide (NT-proBNP), suppression of tumorigenicity 2 (ST2), soluble interleukin-6 (SIL-6) and atrial natriuretic peptide (ANP) were chosen based on their specificity for CVD and known physiological response in patients. The acquired average Raman spectral fingerprints from the panel of these biomarkers are shown in **Fig 7.4i**. The spectral fingerprints of healthy compared to the CVDs with ApoB, LpA, NT-ProBNP, PCSK9 and IL-9 exhibit a statistically significant difference ($p^*<0.05$) with an increased level of each marker in CVD-derived EVs, which is most likely due to atherosclerosis affecting the coronary arteries, primarily influenced by cholesterol and specifically, low-density lipoprotein cholesterol (LDL-C).

This is further underscored by the success of LDL-C lowering medications, such as statins as well as Mendelian randomisation studies, which clearly demonstrated the causative role of LDL-C's in CVD [96], [97]. For instance, Leander *et* al., identified the significant role of PCSK9 in the regulation of LDL-C and thus, demonstrated the potential of soluble PCSK9 to act as a novel indicator of CVD risk [98]. Further, the LpA, ApoB and PCSK9 demonstrated increased intensities at 1315cm^{-1} as well as between 1200-1340cm^{-1} are attributed to lipids and the identified upregulation of Amide I (1600-1655cm^{-1}) in both NT-ProBNP (~18%), IL-9 (~14%) and LpA (~32%), suggests the significant role of lipids and cytokines in the onset of CVD and demonstrates their potential as significant indicators of cardiovascular disease (**SI Tables. 1-4)**.

Further, *via* the application of SKiNET in the data classification, spectral barcodes of healthy and diseased EV subpopulations for CVD have been derived from the corresponding average fingerprint spectra selected from the bands of the highest intensity and significance (**Fig 7.4 j-m**) with SOMDI extracted features from SOMs, highlighting the most influential Raman peaks for each class. This establishes the multiplex barcoding from a complex biological matrix based on their distinct spectral signatures combined with the computational SKiNET algorithm for rapid classification, *via* the key-features from the spectral analysis visually represented in the coloured Raman maps, providing a selective and sensitive method for detection of CVD.

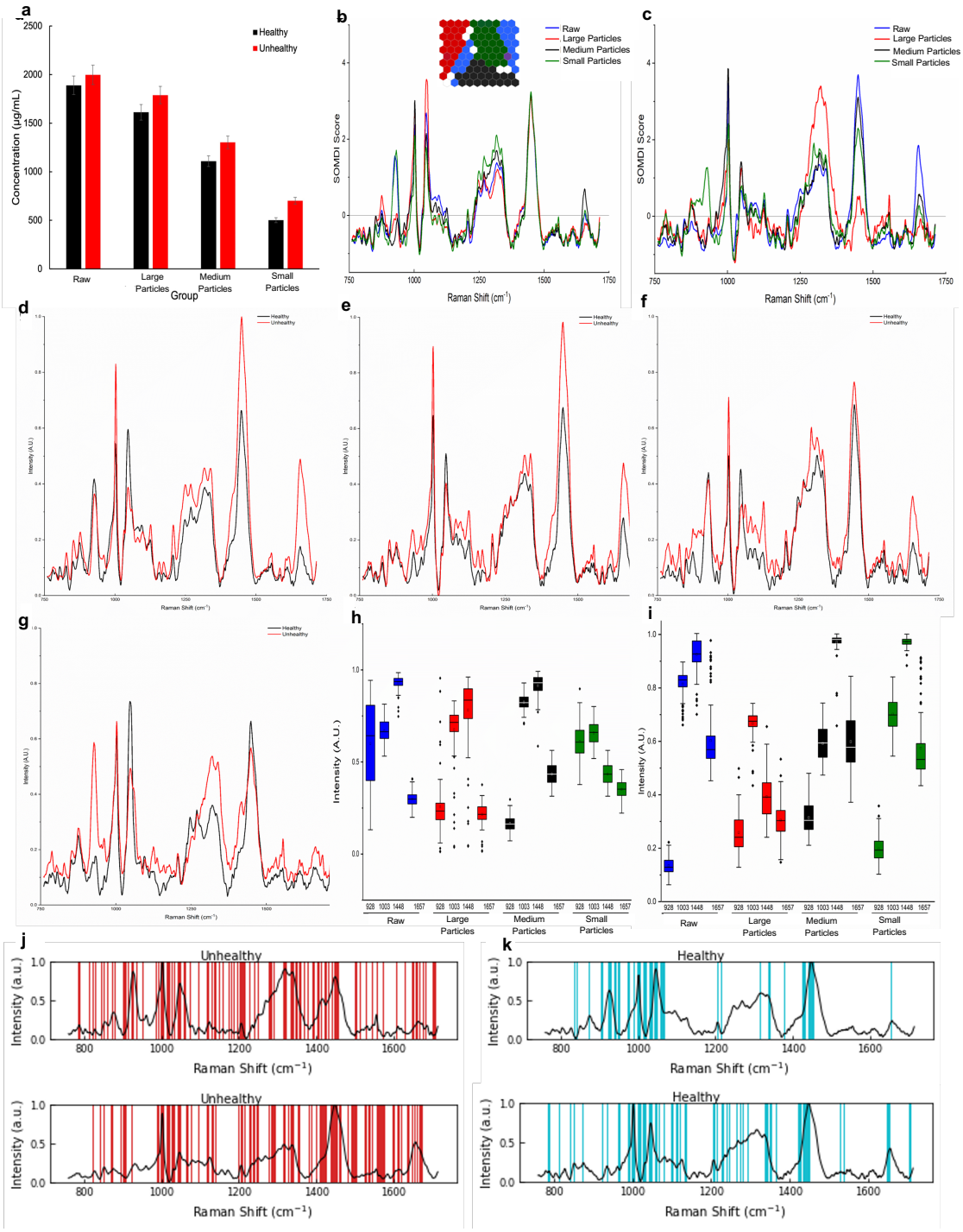

Figure 7.5. AIMSPec-LoC Versatility Validation-Spectral Detection and Differentiation of Inflammatory Bowel Disease from Saliva. (a). Protein concentration determined *via* BCA assay for each filtration step corresponding to raw saliva, large particles, medium particles and small particles. **(b)** Representative SOMDI and SOM **(inset)** of healthy EVs with a classification accuracy of 98.5±0.1%. Sub classification by EV subgroup is represented by blue hexagons for raw saliva, red hexagons for

large particles, black hexagons for medium particles and green hexagons for small particles. **(c).** SOM clustering (inset) and SOMDI classification of IBD saliva are classified with an accuracy of 96.7±0.2%. Representative average Raman spectra of healthy (black) and IBD (red) differentiating **(d)** large particles, **(e)** medium particles and **(f)** small particles as well as **(g)** raw saliva. **(h-i).** Box and whisker plots represent the minima, maxima, interquartile ranges, whiskers and the median in peak intensities identified *via* SOMDI as having the greatest effect on the classifier, *i.e.,* the spectral changes of greatest importance with the largest changes observed of 928, 1003, 1448 and 1657cm^{-1} between healthy **(h)** and IBD **(i)** saliva for each EV subgroup. Characteristic molecular barcodes derived from the corresponding average fingerprint spectra indicative of IBDs of **(j)**, medium particles (top) and small particles (bottom) and healthy **(k)** medium particles (top) and small particles (bottom).

In addition to blood plasma EVs offering instrumental roles in disease diagnosis (**Fig. 7.4/SI Figs.8-9**), EVs isolated from saliva hold further important diagnostic potential, particularly due to its non-invasive and readily accessible characteristics. The versatile nature of the developed AIMSPec-LoC easily adaptable to different biofluids, opens a path in the successful diagnosis, and monitoring of further diseases (**Fig 7.5**). We have subsequently, in addition to plasma, tested the developed lab-on-a-chip to efficiently capture saliva derived EVs, expanding the scope of EV-based diagnostics. The collected EVs from saliva from IBD patients were analysed *via* Raman spectroscopy AIMSPec chip profiling the fingerprint region of 750-1750cm^{-1} with an optimal signal-to-noise ratio and spectral reproducibility. Spectroscopic fingerprint datasets comprising raw saliva, large particles, medium particles and small particles were acquired from 50 clinical samples of n_{IBD}=21 and $n_{Healthy}$=24 (**Methods**), characterising the molecular profile of saliva derived EVs and elucidating their potential as diagnostic candidate biomarkers of IBD. The subsequently classified data *via* SKiNET demonstrates a clear separation of healthy and IBD (**Fig.7.5, e-f**) for each of the EV

subgroups with classification accuracy of 100% (**Table 7.3**) and of 96.7% for all EVs indicative of IBD.

The average spectra from the EV subgroups isolated from healthy and IBD saliva exhibit characteristic bands at 1656, 1550, 920 and 720-820cm^{-1} attributed to Amide I, Amide II, phosphate bending and deformation and nucleic acids, respectively. Across each spectrum, the lipid and protein contents, particularly *via* the CH, CH_2 and CH_3 Raman shifts of 1124, 1440, 1450, 1472 and 1657cm^{-1} vary significantly ($p^{**}<0.001$) between the healthy and IBD cohorts as well as between the three EV subpopulations, suggesting that the main detected difference within the sample sets largely involve both proteins and lipid components (**Fig 7.5, k-l)**. Peaks corresponding to lipids exhibit a higher intensity in exosomes than in either of the isolated fractions corresponding to medium and small particles. Previous studies indicate exosomes have a remarkably high lipid content compared to microvesicles and apoptotic bodies and EVs are well-established to contain a wide range of lipids including glycerophospholipids, sphingolipids, fatty acids and cholesterol [99] Alterations in lipid profiles, such as these identified *via* AIMSPec-LoC in patients with IBDs, have been previously recognised as responsible for the upregulation of lipid content within IBD-derived EVs. For instance, Jansson *et* al. analysed faecal samples and identified increased amounts of fatty acids such as, linoleic and stearic acids in patients with CD [100]. Also, Franzosa *et* al. reported increased levels of sphingolipids (ceramide and sphingomyelin) in serum of IBD patients compared to healthy subjects [101].

For the various filtration levels throughout the LoC, raw saliva is found to have the highest protein content of 1998µg/mL and the small particles, collected at the final filtration step, containing the lowest levels of 701µg/mL **(Fig 7.5d**). This is in agreement with the detected Raman spectra changes, whereby bands associated with proteins *i.e.,*1124, 1131 and 1440cm^{-1} are of a significantly higher intensity (p**<0.001) in raw saliva than the EV subgroups with the least intense protein peaks identified for small particles. This is in correspondence with previous studies investigating the potential of EV proteins derived from IBD saliva to act as biomarkers of disease indicate higher levels of ANXA1 in patients with mucosal inflammation [102]. Furthermore, Zheng *et* al. compared protein profiles of EVs isolated from saliva of IBD patients and detected eight proteins, which were only present in EVs of IBDs and not in healthy patients [103]. These are in agreement with our findings indicating the observed EV profiles significantly differ in the overall protein composition with more intense protein bands observed in patients with a known IBD diagnosis with characteristic bands of 1445, 1472 and 1550cm^{-1}.

Table 7.3. IBD Disease Differentiation *via* EVs. Classification performance and accuracy of healthy and CVD patient in saliva compared to healthy controls including sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV).

Comparison	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
Healthy *vs.* IBD Raw saliva	81.9 ±2.3	85.6 ± 2.8	86.3 ± 1.8	80.9 ±2.1	83.6 ±1.9
Healthy *vs.* IBD large particles	100	100	100	100	100

Healthy *vs.* IBD medium particles	100	100	100	100	100
Healthy *vs.* IBD small particles	100	100	100	100	100

Further identified peaks of significance (**Fig 7.5, d-f**) are attributed to saccharide content of the samples. Of a particular interest are the tryptophan ring detected at 759cm^{-1}, tyrosine doublet (protein) at 830-853cm^{-1}, 882cm^{-1} (phosphatidylcholine), 1003cm^{-1} (phenylalanine), 1049cm^{-1} (glycogen) and 1307cm^{-1} (phospholipids). These distinct bands combined with the protein and lipid peaks, also dominant from SOMDI analysis, constitute the major differences in the biochemical composition of EVs in healthy *versus* IBD saliva sample types and within each EV subpopulation. The identified increase in phenylalanine and tyrosine in IBD saliva is consistent with previous reports, which shown increased levels of these amino acids as indicators of IBD [100]. It has also been reported that increased levels of phenylalanine have been associated with immune activation and inflammation and in the diseased state, there is a greater infiltration of inflammatory cells leading to the increased quantity of phenylalanine (**Fig 7.5, g-j**). Inflammation disrupts the epithelial barrier, which prevents nutrients from being absorbed in the gut and leads to IBD patient samples such as faeces, to contain higher levels of amino acids *e.g.,* phenylalanine and tyrosine [104]. Similarly, tryptophan has been also shown to be altered in IBD patients with Jansson *et* al. and Schicho *et* al. identified increased levels in faecal and urine samples [100], [105]. Tryptophan plays significant roles in IBDs leading to its upregulation. It serves as a precursor for serotonin, which may influence gut function in patients whilst metabolised along the kynurenine pathway leading to production of metabolites *i.e.,*

kynurenine and quinolinic acid, both of which have been implicated in modulating immune responses and inflammation. Inflammation-induced tryptophan uptake has also been shown to lead to increased tryptophan levels since the inflammatory processes in IBD result in increased availability of tryptophan in immune cells and hence, increased circulating levels. Furthermore, patients with IBD typically show an overall decrease in their gut microbiome taxonomic diversity compared to healthy individuals with these changes reportedly altering many metabolites and pathways due to the inability of the gut microorganisms to transform dietary products into compounds essential for metabolism with amino acids being the most affected by this failure. Therefore, the significantly differing amino acid profiles identified *via* spectroscopic analyses may, in the longer-term, pave the way for these to act as significant EV-derived biomarkers of IBD.

Furthermore, comparing the presence of non-EV contaminants in saliva *versus* plasma is essential for understanding the composition and potential influence of these contaminants on EV studies. Non-EV contaminants in saliva may include various components such as mucins, proteins from oral bacteria, food residues, and other salivary proteins [106], [107]. These contaminants can originate from the oral cavity and surrounding tissues and may vary in composition and concentration depending on factors such as oral hygiene, diet, and oral health status[108]. However, many reports indicate that salivary EVs have a higher degree of purity compared to EVs from serum or plasma due to the low amounts of co-isolated proteins and the absence of lipoproteins [109], [110], [111]. In contrast, non-contaminants in plasma may include high protein concentrations of, for example, albumin as well as lipoproteins that have

similar dimensions and densities as EVs. Plasma samples typically contain EVs in the size range of 40 nm to 2 μm, as was evidenced in this study; while lipoprotein subgroups of high-density lipoproteins (HDL) are 5-12 nm in size, low density lipoproteins (LDL) are 18-25 nm in size, intermediate density lipoproteins (IDL) are 25-35 nm in size, very low-density lipoproteins (VLDL) are 25-35 nm, chylomicron remnants are 30-80 nm, and chylomicrons are 75-1200 nm in size [110], [112], [113], [114]. Thus, this size range makes it difficult to isolate each pair of particles respectively. In this study, low protein binding filter membranes of 40 nm have been applied to trap particles smaller than 40 nm in the size range smaller than EVs but larger than HDL, LDL, IDL, and VDL in order to reduce the effects of contamination from these components on the overall Raman spectra obtained for each fraction.

In addition, the presence of VLDL in plasma and triglycerides in serum post-meals is closely linked to the plasma concentration of EVs and thus implies that the quantification of particles *via* a technique such as NTA may be influenced by the presence of these molecules [114]. To reduce this effect, blood samples, similarly to saliva were collected 45 minutes prior to any ingestion of food. Moreover, studies indicate samples collected using Ethylenediaminetetraacetic Acid (EDTA) as a blood anticoagulant exhibits the highest plasma concentration of EVs, thus indicating that EDTA may stimulate the generation of micro-vesicles through platelet activation and thus contribute to the increased numbers of smaller EVs detected with NTA [115].

The presence of non-EV contaminants in saliva and plasma has the potential to compromise findings in EV studies such as the one presented here. They can for example, interfere with the detection and quantification of EV associated biomarkers,

masking or obscuring signals from EV-specific markers, reducing the overall sensitivity and specificity of biomarker assays. They can also lead to the generation of false positive or false negative results and thus introduce bias and misinterpretation [109], [110]. Therefore, it is paramount to apply an EV isolation technique that minimises this potential interference. Here, low-protein binding PES membranes were applied as well as the introduction of filters with pore sizes of 40 nm to trap smaller lipoprotein particles and reduce the error associated with non-EV contamination. Increased co-isolated proteins identified in plasma could also be attributed to EV protein corona. It is well understood that EVs circulating in the blood have protein corona, which comprises molecules tethered to the vesicles and cannot be separated from the outer membrane of EVs during isolation [33], [34]. Some of these proteins constitute the protein cargo absorbed onto EVs, while others are partially recruited from bodily fluids following vesicle shedding. Also, the calculated ratio of protein per fraction was greater for salivary EVs than those derived from serum, indicating salivary EVs may have a limited protein corona and a major protein load in their core [116], [117].

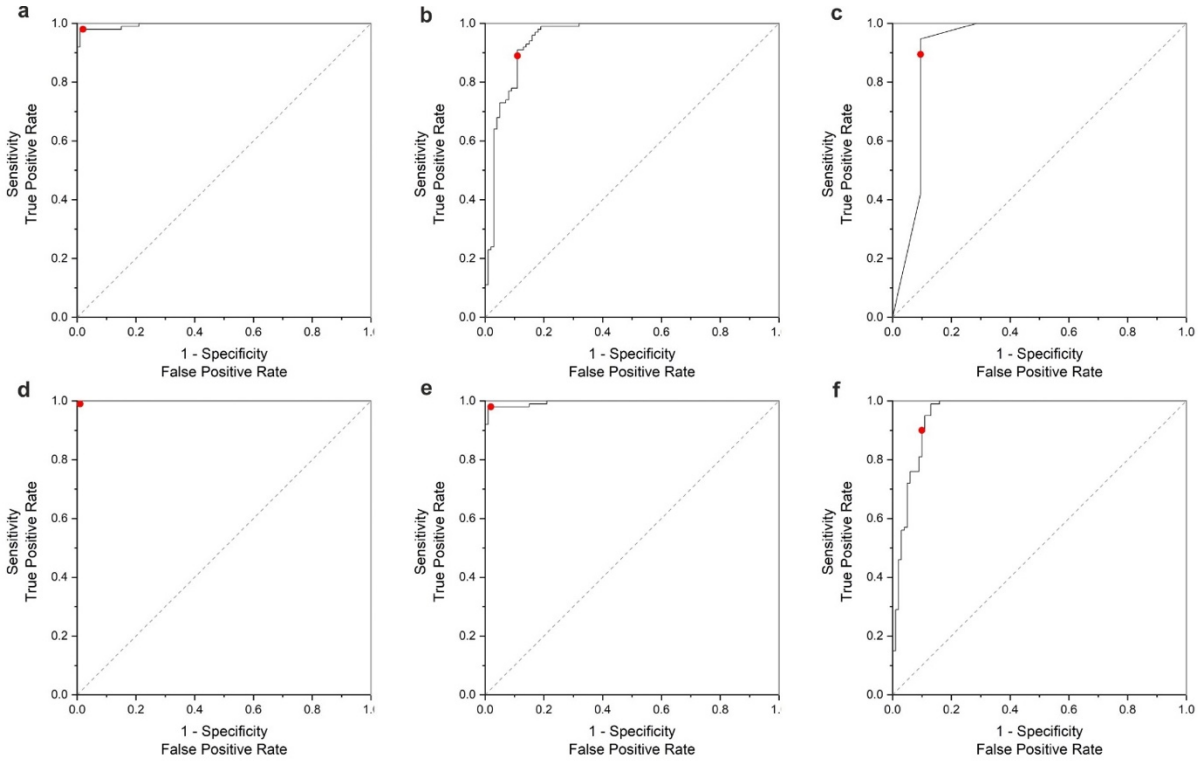

Figure 7.6. ROC plots representing the Sensitivity versus 1-Specificity derived from post probability assignment applied to determine the success for classifying the healthy and CVD-derived blood plasma for **(a)** small particles, **(b)** medium particles and **(c)** large particles and of healthy and IBD saliva for **(d)** small particles, **(e)** medium particles and **(f)** large particles, two-sided test with no multiple comparisons. The AUC indicates that the change in the protein and lipid levels of plasma and saliva-derived EVs are valuable markers to discriminate healthy from the IBD and CVD-indicative EVs with the CVD indicative EVs showing an excellent performance with AUC = and 0.99 (95% CI 0.99-1.0), 0.95 (95% CI 0.94-0.96) and 0.92 (95% CI 0.81-1.03) for small, medium and, large particles respectively, and AUC = 0.99 (95% CI 0.99-1.0) and, 0.99 (95% CI 0.99-1.0) and 0.96 (95% CI 0.95-0.97) for IBD-indicative small, medium and, large particles.

We used receiver operating characteristic (ROC) curves to assess the differentiation of healthy and CVD blood plasma as well as healthy and IBD saliva for each of the EV subpopulations by calculating the area under the curve (AUC) plotting the true positive versus the false negative rates (**Fig 7.5**). From the ROC curves we determined the

intrinsic classification accuracy in CVD-derived blood plasma of 99% for small particles (**Fig 7.6a**), 95% for medium particles (**Fig 7.6b**) and 92% for large particles (**Fig 7.6c**) and in IBD saliva of 99% for the small particles (**Fig 7.6d**), 99% for medium particles (**Fig 7.6g**) and 96% for large particles (**Fig 7.6f**) clearly discriminating between healthy and diseased groups, which is crucial for classifying CVD and IBD.

To note, the ROC's steep ascent with the consistent trajectory toward the upper-left corner, indicating a minimal rate of false positives while maximising true positive predictions of the classification. This is critical when assessing diagnostic or predictive models, due to the indication of the ability to accurately detect relevant diseases cases and states. The calculated AUC overall range from 0.92-0.99 with the higher values demonstrated for the small particle subpopulation in both plasma and saliva, in-line with SOM classification accuracy (**Fig 7.4c** and **7.5f**) with classification accuracies of >96%. These AUC values underscore the robustness of our combined AIMSPec-LoC with intrinsic AI classification, exhibiting accurate discriminatory capabilities, particularly valuable in applications of disease diagnostics, where precision and specificity are of a paramount importance.

Overall, the spectral barcodes derived from the corresponding average fingerprint spectra of healthy and IBD derived EVs used to identify biochemical components based on their unique Raman signature highlight the importance of proteins and lipids as the primary biomolecules responsible for the classification of patient samples, in agreement with SKiNET classification with SOMDI extracted features from SOM (**Fig 7.5, m-n**), identifying the most influential specific Raman bands. Through spectral

barcoding of the data sets, each component's unique barcode serves as a spectral fingerprint, allowing for the identification and characterisation of individual biochemicals without the need for prior knowledge or reference samples. It also allows for the differentiation of closely related species with similar Raman bands as well as any changes within the barcode, which may be due to the disease progression or drug interactions, cementing the AIMSPec-LoC suitability to non-invasively detect and monitor the diseased state of patients.

7.4 Conclusions

Numerous CVDs currently are only identified after a life-threatening event such as a heart attack or stroke, making it vital to recognise, detect and study, the often subtle, biochemical changes which precede these events. Similarly, with IBDs, oftentimes it is difficult to discriminate these from other gastrointestinal diseases such as, irritable bowel syndrome, with invasive and painful endoscopy being the current ‘gold standard’ diagnostic method. There is, therefore, an unmet need to diagnose and quantify the molecular profile of CVD and IBD at the early disease onset.

We have developed a novel lab-on-a-chip technology for simultaneous isolation and analysis of extracellular vesicles, which has been validated for the detection of EVs indicative of CVD and IBD from blood plasma and saliva. The AIMSPec-LoC enables label-free separation of extracellular vesicles and their associated subgroups of small, medium and large particles *via* its inherent size-based and capillary action functions. In contrast to traditional methods, our on-chip isolation and detection requires minimal sample pre-processing, eliminating degradation and subsequent loss of EVs. With the

diverse cargo of EVs offering a multiplexed approach for biomarker analysis, the integrated capabilities of the AIMSPec-LoC not only reduce the overall processing time but also yield the *real-time* analysis of multiple EV subtypes, enabling the discrimination of multiple EV markers and their cargos whilst also providing a comprehensive view of their underpinning biochemical characteristics, in a single run. It is worth noting the sample volume, which can be separated *via* AIMSPec-LoC for downstream analysis, is dependent on the size and length of the capillary channels serving to pull the biofluid through the system. When the channels and end collection wells are full, the flow ceases. At present, the design accommodates up to 1.5mL of sample at a time. Scaling up the overall size of the LoC or increasing the length of the capillary flow channels, would further enhance the overall ability of the LoC to efficiently separate larger volumes of biofluid sample. Furthermore, while the AUC is an important evaluation performance parameter to distinguish between classes, thorough assessment of other evaluation metrics including sensitivity, specificity, PPV and NPV are required to provide a comprehensive view of the system performance. With the methodology excelling across these metrics, achieving overall >96% sensitivity and specificity, reaffirms the effectiveness of the approach in rapidly detecting disease indicative biomarkers from a range of biofluids.

With progressing advances in hand-held Raman spectrometers, their integration with LoC devices and EV analysis has further potential to revolutionise diagnostics, disease monitoring, response to therapeutics and identify biomarkers of disease, which in turn, will ultimately lead to improving patient outcomes and healthcare management. The generated barcoded datasets in the context of CVD and IBD detection *via* Raman

spectroscopy provides the means for multiplex sample analysis, performing molecular profiling, organising and integrating data whilst leveraging machine learning classification. These collectively contribute to more efficient, comprehensive, and accurate detection, enhancing our understanding of the disease and potentially improving patient outcomes. The fast and efficient nature of the LoC also allows for *real-time* analysis, enabling continuous disease monitoring leading to potentially more effective and tailored treatment plans. The utilization of EVs as biomarkers, coupled with RS, offers a promising avenue for non-invasive and sensitive detection of diseases (*e.g.,* CVD/IBD), reducing patient discomfort and risks associated with traditional tissue biopsy. It can also provide *real-time* information about disease progression and treatment response, allowing for dynamic monitoring of patient's health. This study, therefore, holds significant implications for enhancing clinical decision making and paving the way for personalised medicine in the field of cardiovascular and bowel disease medicine and beyond.

This work overall marks an important proof-of-concept laying the groundwork towards the development of *in-vivo* clinical measurements *via* portable AIMSpec-LoC device for CVDs and IBDs. The acquired spectral data is rapidly and accurately classified using our artificial neural networks algorithm, SKiNET as a decision support tool for easy readouts for clinicians. Measuring abnormal changes in specific EV biomarkers would be indicative of CVDs or IBDs, providing a quantitative assessment at the earliest stages whilst simultaneously helping to quantify the damage. It would be interpreted by clinicians as an indication to treat the patients according to guidance without delay and help in initiating preventive treatments at an earlier stage, thereby

reducing the risk of avoidable mortality. In the long-term, this versatile platform enhancing the diagnostic utility of EVs will be extended to further applications *e.g.,* early-recognition of other major diseases and timely interventions specifically targeting the identification of CVDs/IBDs in otherwise healthy individuals. demonstrated with the excellent capability of this versatile technology to easily adapt to different biomarkers, which may further open an untraveled path in the successful diagnosis, prognosis and monitoring of major, devastating diseases.

7.5 Methods

7.5.1 Clinical Samples

Study participants were recruited through the Inflammatory Bowel disease and Cardiovascular Research Centres at Queen Elizabeth Hospital of Birmingham (UK) as part of the CASCADE and TRAFIC studies) (Ethics Ref. 19/SW/0010 and ERN_22-0290). Written informed consents were received from participants or valid proxy (family or a professional not directly involved in the study) prior to inclusion in the study. The samples were obtained from a standard, widely accepted classification of the CVD or IBD based on clinical criteria *i.e.,* full blood panels, as well as supported by in-hospital scans including electrocardiograms for CVD diagnosis and endoscopy with histological assessment for IBD. Patients were categorised into HV, CVD, IBD (to include UC and CD). All patients were gender and age matched to HVs. Patient demographics are shown in **Tables 7.4** and **7.5**. Blood samples from patients in each category were obtained at Queen Elizabeth Hospital, Birmingham. Once collected, blood samples were centrifuged initially at 2,000 x g for 20 minutes follow by a further centrifugation

cycle of 13,000 x g for 2 minutes. Extracted blood plasma was stored at -80°C until analysis. All samples were processed within two hours of venepuncture.

7.5.2 Saliva Collection and Preparation

Saliva was collected from 35 participants (10 ulcerative colitis, 11 Crohn's Disease and 14 healthy) (**Table 7.4)**. IBD diagnosis was confirmed *via* current clinical practice, primarily endoscopy with histological assessment. All healthy volunteers had no previous health issues or known conditions. All participants were evaluated by the research team at Queen Elizabeth Hospital Birmingham, UK, who were instructed not to consume any food or drink 45 minutes prior to sample collection. 5mL of unstimulated saliva was collected from each participant in a 50mL conical tube (Thermo Fisher, UK) using the passive drool method. Raw saliva was stored at -80°C.

Table 7.4. Demographics Including Epidemiological and Clinical Parameters of IBD Patients.

Characteristic	IBD (n=21)	Healthy (n=24)
Age (median, std, yrs)	43.7 ± 14.9	39.2 ± 7.4
Sex		
Female	14	13
Male	11	12
Disease duration (years)	72.6 ± 8.8	-
UC patients	10 (48.5%)	-
UC Montreal disease extension		
Proctosigmoiditis	7 (43.7%)	-
Left colitis	1 (6.25%)	-
Pancolitis	4 (25%)	-
CD patients	11 (51.5%)	-
CD Montreal disease localisation		
Ileal	3 (29.4%)	-
Colonic	2 (11.8%)	-
Ileocolonic	8 (47%)	-
Perianal disease	2 (11.8%)	-
CDAI	-	-
pMAYO	-	-
SES-CD > 2	-	-
Rutgeerts score ≥ 2	-	-
MES >1	5	-
MES ≤1	3	
UCEIS ≤1		
UCEIS > 1	2	
PICASSO	6	
Baseline anti-TNF-α therapy	6(18.1%)	-
Baseline AZA therapy	4(12.1%)	
Baseline Vedolizumab	3(9.09%)	
Baseline ustekinumab	5 (15.1%)	
Baseline tofacitinib	4(12.1%)	
Baseline mesalamine therapy	5 (15.1%)	
Baseline risankizumab	1 (3.03%)	
CRP (mg/L)	10.3 ± 9.8	-
FC (µg/g)	298 ± 8.9	-

7.5.3 Blood Plasma Collection and Preparation

Blood samples were collected from Queen Elizabeth Hospital Birmingham and University of Birmingham, UK from 40 participants (**Table 7.5)**. Among them, 20 were clinically diagnosed with CVD confirmed *via* current clinical practices and the remaining 20 participants were healthy with no known health issues or conditions. Blood plasma

was separated from whole blood using a 2-step centrifugation process; initial centrifuge at 2,000xg for 20 minutes, followed by removal of supernatant and further centrifuge at 13,000 x g for 2 minutes. Plasma was then extracted and stored at -80°C. During the recruitment process, written informed consent was obtained from each participant.

Table 7.5. Demographics comparing epidemiological and clinical parameters of CVD patients.

		Control (*n*=20)	CVD (*n*=20)
Age (years)		48.5 ± 11.5	65.5 ± 23.5
Sex (Male/Female)		28/12	49/11
Systolic blood pressure (mmHg)		122 ± 8	135 ± 25
Diastolic blood pressure (mmHg)		77 ± 9	84 ± 36
BMI (kg/m^2)		28.2 ± 4.2	34.3 ± 6.8
Presentation (*n*)	**STEMI**	-	10 (50%)
	NSTEMI	-	6 (30%)
	Atrial Fibrillation	-	4 (20%)
Medication (*n*)	**Aspirin**	0	18 (90%)
	Clopidogrel	0	10 (50%)
	Ticagrelor	0	14 (70%)

7.5.4 Fabrication of the Nanoporous Membrane Chip

All layers were fabricated using 3D resin printed micro moulds of high temperature resolution resistance resin (Siraya Tech, USA). Moulds were printed using a monochrome LCD/MSLA resin printer equipped with a 405nm light source (Elegoo Mars 3, China) and subsequently, thoroughly washed in isopropyl for 20 minutes, annealed at 120°C and finally, cured under an ultraviolet lamp (Elegoo, China) for 20 minutes to prevent mould/PDMS interaction. PDMS (Dow corning, USA) was mixed in a ratio of 5:1 with curing agent Sylgard ™ 184 (Dow Corning, USA). Access holes of 6-mm diameter were devised for the inlet and outlet chambers, respectively. Sterlitech polyethersulfone (PES) membrane filters of 25mm diameter with varying pore sizes

(5000, 500, 200, 100 and 40nm) were placed between the PDMS layers. PDMS layers were bonded through a combination of plasma for 10 minutes followed by hard baking at 120°C for 30 minutes.

7.5.5 Lab-on-a-chip Device Design and Fabrication

The designed microfluidic chip consisted of 5 layers of PDMS adhered to a glass slide. The top layer contained a well for the sample inlet, at the base of the well and between the top and middle layers a hydrophilic polyethersulfone nanoporous membrane with a pore size of 5000μm to filter out any cellular debris/food particles. Saliva dripped through the filter into the inlet of the next PDMS layer, moving down the microfluidic channel and into the sample collection area containing the PBS buffer. It subsequently, passed through a further filter membrane (500 nm) sandwiched between the two PDMS layers, where EVs of 5000-500 nm (large) were trapped and collected. Saliva continued to drip through a further three layers with pore sizes of 200, 100 and 40nm to collect and trap EVs in the size ranges of 500-200 nm (large), 200-100 (medium), 40-100 (small). 40nm filter was used to trap proteins. Capillary channels were used to aid flow throughout the filtering process. Prior to sample injection, 1mL of PBS was injected into the inlet to pre-wet the filter membranes. Sample outlet was filled with 100μL of PBS to aid capillary flow throughout the system. After 10minutes, inlet PBS was replaced with 100μL of sample. Separation was carried out at room temperature.

7.5.6 Nanoparticle Tracking Analysis

NTA NanoSight NS300 (Malvern, UK) was used to measure the size distribution and concentration of the collected EVs. Blood plasma samples were diluted with pre-filtered

PBS in a 1:5 ratio with saliva diluted 1:10 in line with recommended particle concentration per frame (20-100 particles/frame). For each measurement five 60s videos were captured. After capture, NanoSight software NTA 3.1 (Malvern, UK) was used to analyse video segments for each EV measurement with a detection threshold of 6.

7.5.7 Dynamic Light Scattering

DLS measurements were performed using the Malvern Panalytical Zetasizer HPPS equipped with a 633-nm He-Ne laser. Low volume quartz cuvettes were used for analysis of 15µL of samples. All measurements were carried out at a fixed position with an automatic attenuator and at a controlled temperature of 25°C. 5 measurements were averaged for each EV sample.

7.5.8 Fluorescence Microscopy

Particles were analysed using Zeiss LSM780 microscope equipped with excitation wavelengths of 405, 488 and 514 nm to visualise different coloured nanoparticles and x10/x50/x100 objective lens. All images were processed using open-source ImageJ software.

7.5.9 Scanning Electron Microscopy

EVs were analysed on the filter membrane, in solution (PBS) and in fixed form. EVs were fixed on silicon wafers using 2.5% glutaraldehyde prepared in PBS for 30 minutes. The sample was then sequentially dehydrated in an ascending sequence of ethanol. Samples were dried at room temperature prior to SEM observation. Several

randomly selected frames from each sample were captured for morphological assessment and scanning electron micrographs were acquired using a thermally assisted Field Emission Scanning Electron Microscope (FESEM, LEO VP 1530 and FEI Magellan and Helios) with a lateral resolution of 2-5nm. A LEO ULTRA 55 SEM instrument including a Schottky emitter (ZrO/W cathode) was also used for imaging the samples with a typical acceleration voltage of 2-5kV.

7.5.10 Raman Spectroscopy

Raman spectroscopy was performed directly on chip at each filter membrane. Raman spectra were acquired using a Renishaw InVia Qontor confocal Raman microscope equipped with a microscope Leica DMLM (Renishaw PLC, UK). 785nm excitation laser was used in the analysis and laser light was focussed using x50 objective. Output power at the sample was 5mW. All Raman spectra were acquired in the main fingerprint region of 700-1700cm^{-1}. Map scans were obtained over a 50x50μm^2 areas for each sample using a 5μm step size between map points, 10 accumulations of 1second acquisition per spectra. In summary, 1 map with 100 spectra were collected per sample for both saliva and plasma and used for data processing and analysis. All Raman data was acquired using WiRE 5.1 (Renishaw PLC, UK) also applied for the polynomial background subtraction and removal of the cosmic rays. Spectra were normalised using the standard normal variate (SNV) using Python (Python 3.7). Signal-to-noise ratio was improved using 10 averaged spectra at each map location.

7.5.11 SKiNET Classification

Multivariate Analysis was performed using the SKiNET, alongside the Raman Toolkit web interface to build the SOM models. Data sets were grouped and split into test and training data (20:80). 10-fold cross validation was performed on the training data with optimisation of grid-size, initial learning rate and number of training steps. The end model used to classify the test data consisted of a 10 x 10 grid of neurons, 46,080 training steps (4 epochs of the data) and an initial learning rate of 0.1. Python was used to complete multi-chemical barcoding of data. Savitzky-Golay filter was applied to calculate the second derivative of each spectrum. Smoothing window was set to 21 with a polynomial order of 2. Maximum peak heights with absolute values over 40% were assigned a value of 1 with values below 40% assigned a value of 0. Values were overlaid on the averaged spectra with reference to the main peaks of interest identified, thus creating the barcode.

7.5.12 Statistical Analysis

Spectral data collected was analysed using IBM SPSS Statistics (IBM Corp., USA). Descriptive statistics, including means, standard deviation and range were calculated to summarise demographic characteristics of the study participants. Statistical significance of the difference between two sets of data was assessed using the Wilcoxon rank sum/Mann-Whitney U test. The significance level for all statistical tests was set at 0.05. A p-value <0.05 was considered statistically significant. Comparisons across groups at each time and within groups over time were performed by the analysis of variance and Tukey's post hoc test on transformed data.

7.5.13 ROC Curves and Box Plots

ROC curves were generated based on patient profiling data for various cut-off points using non-parametric Mann-Whitney U and Kruskal Wallis tests, executed with SPSS statistics software. Each point on the ROC curve represented a sensitivity/specificity pair corresponding to a specific decision threshold, standard equations were used to calculate diagnostic values such as sensitivity, specificity, and accuracy. An ROC curve for a test with perfect discrimination (no overlap in the two distributions) passed the upper left corner (100% sensitivity, 100% specificity). The closer the ROC curve approached the upper left corner, the higher the overall accuracy of the test. Box plots were created using Origin Pro software, employing an x-y chart for each series to represent interquartile ranges and accommodate data with negative values. The median was denoted by a circle marker, and horizontal markers represented Q_1 and Q_3 without the need to shift the data. Two-sided normal-based 95% CIs *t-test* was used to compare mean levels between the patient groups and healthy volunteers control group. Classification sensitivity, accuracy, and specificity were determined based on disease detection results: Sensitivity=(TP)/(TP+FN), Specificity=(TN)/(TN+FP) and the Accuracy= (TP+TN)/(TP+TN+FN+FP) with TP being ‘true positive’, TN ‘true negative’, FP and FN ‘false positive’ and ‘false negative’, respectively.

7.6 References

[1] World Heart Federation, “WORLD HEART REPORT 2023 CONFRONTING THE WORLD’S NUMBER,” 2023, [Online]. Available: https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf

[2] S. Khalid, “IBD Statistics 2022: Crohn’s and Ulcerative Colitis,” Ampersand.

[3] Y.-Z. Zhang, “Inflammatory bowel disease: Pathogenesis,” *World J Gastroenterol*, vol. 20, no. 1, p. 91, 2014, doi: 10.3748/wjg.v20.i1.91.

[4] I. J. Pence *et al.*, “Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy,” *Biomed Opt Express*, vol. 8, no. 2, p. 524, Feb. 2017, doi: 10.1364/BOE.8.000524.

[5] R. Dhingra and R. S. Vasan, “Biomarkers in cardiovascular disease: Statistical assessment and section on key novel heart failure biomarkers,” *Trends Cardiovasc Med*, vol. 27, no. 2, pp. 123–133, Feb. 2017, doi: 10.1016/j.tcm.2016.07.005.

[6] P. (Babu) Balagopal *et al.*, “Nontraditional Risk Factors and Biomarkers for Cardiovascular Disease: Mechanistic, Research, and Clinical Considerations for Youth,” *Circulation*, vol. 123, no. 23, pp. 2749–2769, Jun. 2011, doi: 10.1161/CIR.0b013e31821c7c64.

[7] K. Dixon *et al.*, “Using Machine Learning and Silver Nanoparticle-Based Surface-Enhanced Raman Spectroscopy for Classification of Cardiovascular Disease Biomarkers,” *ACS Appl Nano Mater*, vol. 6, no. 17, pp. 15385–15396, Sep. 2023, doi: 10.1021/acsanm.3c01442.

[8] Y. Deng, C. Huang, J. Su, C.-W. Pan, and C. Ke, “Identification of biomarkers for essential hypertension based on metabolomics,” *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 31, no. 2, pp. 382–395, Feb. 2021, doi: 10.1016/j.numecd.2020.11.023.

[9] L. J. Laslett *et al.*, “The Worldwide Environment of Cardiovascular Disease: Prevalence, Diagnosis, Therapy, and Policy Issues,” *J Am Coll Cardiol*, vol. 60, no. 25, pp. S1–S49, Dec. 2012, doi: 10.1016/j.jacc.2012.11.002.

[10] J. J. Ricotta, J. Pagan, M. Xenos, Y. Alemu, S. Einav, and D. Bluestein, “Cardiovascular disease management: the need for better diagnostics,” *Med Biol Eng Comput*, vol. 46, no. 11, pp. 1059–1068, Nov. 2008, doi: 10.1007/s11517-008-0416-x.

[11] A. Dickhout and R. R. Koenen, “Extracellular Vesicles as Biomarkers in Cardiovascular Disease; Chances and Risks,” *Front Cardiovasc Med*, vol. 5, Aug. 2018, doi: 10.3389/fcvm.2018.00113.

[12] M. Nawaz *et al.*, “The emerging role of extracellular vesicles as biomarkers for urogenital cancers,” *Nat Rev Urol*, vol. 11, no. 12, pp. 688–701, Dec. 2014, doi: 10.1038/nrurol.2014.301.

[13] Y. Li *et al.*, “EV-origin: Enumerating the tissue-cellular origin of circulating extracellular vesicles using exLR profile,” *Comput Struct Biotechnol J*, vol. 18, pp. 2851–2859, 2020, doi: 10.1016/j.csbj.2020.10.002.

[14] I. Lozano-Ramos *et al.*, “Size-exclusion chromatography-based enrichment of extracellular vesicles from urine samples,” *J Extracell Vesicles*, vol. 4, no. 1, p. 27369, Jan. 2015, doi: 10.3402/jev.v4.27369.

[15] N. Comfort *et al.*, “Isolation and characterization of extracellular vesicles in saliva of children with asthma,” *Extracell Vesicles Circ Nucl Acids*, 2021, doi: 10.20517/evcna.2020.09.

[16] K. Iwai, T. Minamisawa, K. Suga, Y. Yajima, and K. Shiba, “Isolation of human salivary extracellular vesicles by iodixanol density gradient ultracentrifugation and their characterizations,” *J Extracell Vesicles*, vol. 5, no. 1, p. 30829, Jan. 2016, doi: 10.3402/jev.v5.30829.

[17] A. Dickhout and R. R. Koenen, “Extracellular Vesicles as Biomarkers in Cardiovascular Disease; Chances and Risks,” *Front Cardiovasc Med*, vol. 5, Aug. 2018, doi: 10.3389/fcvm.2018.00113.

[18] P. D. Robbins and A. E. Morelli, “Regulation of immune responses by extracellular vesicles,” *Nat Rev Immunol*, vol. 14, no. 3, pp. 195–208, Mar. 2014, doi: 10.1038/nri3622.

[19] L. Urbanelli, S. Buratta, K. Sagini, B. Tancini, and C. Emiliani, “Extracellular Vesicles as New Players in Cellular Senescence,” *Int J Mol Sci*, vol. 17, no. 9, p. 1408, Aug. 2016, doi: 10.3390/ijms17091408.

[20] H. Peinado *et al.*, “Pre-metastatic niches: organ-specific homes for metastases,” *Nat Rev Cancer*, vol. 17, no. 5, pp. 302–317, May 2017, doi: 10.1038/nrc.2017.6.

[21] J. C. Akers, D. Gonda, R. Kim, B. S. Carter, and C. C. Chen, “Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies,” *J Neurooncol*, vol. 113, no. 1, pp. 1–11, May 2013, doi: 10.1007/s11060-013-1084-8.

[22] G. Raposo and W. Stoorvogel, “Extracellular vesicles: Exosomes, microvesicles, and friends,” *Journal of Cell Biology*, vol. 200, no. 4, pp. 373–383, Feb. 2013, doi: 10.1083/jcb.201211138.

[23] G. van Niel, G. D’Angelo, and G. Raposo, “Shedding light on the cell biology of extracellular vesicles,” *Nat Rev Mol Cell Biol*, vol. 19, no. 4, pp. 213–228, Apr. 2018, doi: 10.1038/nrm.2017.125.

[24] P. D. Stahl and G. Raposo, “Extracellular Vesicles: Exosomes and Microvesicles, Integrators of Homeostasis,” *Physiology*, vol. 34, no. 3, pp. 169–177, May 2019, doi: 10.1152/physiol.00045.2018.

[25] L. Muller, C.-S. Hong, D. B. Stolz, S. C. Watkins, and T. L. Whiteside, “Isolation of biologically-active exosomes from human plasma,” *J Immunol Methods*, vol. 411, pp. 55–65, Sep. 2014, doi: 10.1016/j.jim.2014.06.007.

[26] R. Szatanek, M. Baj-Krzyworzeka, J. Zimoch, M. Lekka, M. Siedlar, and J. Baran, “The Methods of Choice for Extracellular Vesicles (EVs) Characterization,” *Int J Mol Sci*, vol. 18, no. 6, p. 1153, May 2017, doi: 10.3390/ijms18061153.

[27] V. Muralidharan-Chari, J. W. Clancy, A. Sedgwick, and C. D’Souza-Schorey, “Microvesicles: mediators of extracellular communication during cancer progression,” *J Cell Sci*, vol. 123, no. 10, pp. 1603–1611, May 2010, doi: 10.1242/jcs.064386.

[28] M. Battistelli and E. Falcieri, “Apoptotic Bodies: Particular Extracellular Vesicles Involved in Intercellular Communication,” *Biology (Basel)*, vol. 9, no. 1, p. 21, Jan. 2020, doi: 10.3390/biology9010021.

[29] W. Phillips, E. Willms, and A. F. Hill, “Understanding extracellular vesicle and nanoparticle heterogeneity: Novel methods and considerations,” *Proteomics*, vol. 21, no. 13–14, p. 2000118, Jul. 2021, doi: 10.1002/pmic.202000118.

[30] K. Burbidge *et al.*, “Cargo and cell-specific differences in extracellular vesicle populations identified by multiplexed immunofluorescent analysis,” *J Extracell Vesicles*, vol. 9, no. 1, p. 1789326, Sep. 2020, doi: 10.1080/20013078.2020.1789326.

[31] E. R. Dellar, C. Hill, G. E. Melling, D. R. F. Carter, and L. A. Baena-Lopez, “Unpacking extracellular vesicles: RNA cargo loading and function,” *Journal of Extracellular Biology*, vol. 1, no. 5, May 2022, doi: 10.1002/jex2.40.

[32] B. Mir and C. Goettsch, “Extracellular Vesicles as Delivery Vehicles of Specific Cellular Cargo,” *Cells*, vol. 9, no. 7, p. 1601, Jul. 2020, doi: 10.3390/cells9071601.

[33] M. Wolf *et al.*, “A functional corona around extracellular vesicles enhances angiogenesis, skin regeneration and immunomodulation,” *J Extracell Vesicles*, vol. 11, no. 4, Apr. 2022, doi: 10.1002/jev2.12207.

[34] E. Á. Tóth *et al.*, “Formation of a protein corona on the surface of extracellular vesicles in blood plasma,” *J Extracell Vesicles*, vol. 10, no. 11, Sep. 2021, doi: 10.1002/jev2.12140.

[35] B. W. Sódar *et al.*, “Low-density lipoprotein mimics blood plasma-derived exosomes and microvesicles during isolation and detection,” *Sci Rep*, vol. 6, no. 1, p. 24316, Apr. 2016, doi: 10.1038/srep24316.

[36] A. Németh *et al.*, “Antibiotic-induced release of small extracellular vesicles (exosomes) with surface-associated DNA,” *Sci Rep*, vol. 7, no. 1, p. 8202, Aug. 2017, doi: 10.1038/s41598-017-08392-1.

[37] S. Ritz *et al.*, “Protein Corona of Nanoparticles: Distinct Proteins Regulate the Cellular Uptake,” *Biomacromolecules*, vol. 16, no. 4, pp. 1311–1321, Apr. 2015, doi: 10.1021/acs.biomac.5b00108.

[38] S. Tenzer *et al.*, “Rapid formation of plasma protein corona critically affects nanoparticle pathophysiology,” *Nat Nanotechnol*, vol. 8, no. 10, pp. 772–781, Oct. 2013, doi: 10.1038/nnano.2013.181.

[39] Z. Varga *et al.*, “Size Measurement of Extracellular Vesicles and Synthetic Liposomes: The Impact of the Hydration Shell and the Protein Corona,” *Colloids Surf B Biointerfaces*, vol. 192, p. 111053, Aug. 2020, doi: 10.1016/j.colsurfb.2020.111053.

[40] K. P. Chan, S.-H. Chao, and J. C. Y. Kah, “Exploiting Protein Corona around Gold Nanoparticles Conjugated to p53 Activating Peptides To Increase the Level of Stable p53 Proteins in Cells,” *Bioconjug Chem*, vol. 30, no. 3, pp. 920–930, Mar. 2019, doi: 10.1021/acs.bioconjchem.9b00032.

[41] E. A. Mol, M.-J. Goumans, P. A. Doevendans, J. P. G. Sluijter, and P. Vader, “Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation,” *Nanomedicine*, vol. 13, no. 6, pp. 2061–2065, Aug. 2017, doi: 10.1016/j.nano.2017.03.011.

[42] K. Brennan *et al.*, “A comparison of methods for the isolation and separation of extracellular vesicles from protein and lipid particles in human serum,” *Sci Rep*, vol. 10, no. 1, p. 1039, Jan. 2020, doi: 10.1038/s41598-020-57497-7.

[43] M. Yu. Konoshenko, E. A. Lekchnov, A. V. Vlassov, and P. P. Laktionov, “Isolation of Extracellular Vesicles: General Methodologies and Latest Trends,” *Biomed Res Int*, vol. 2018, pp. 1–27, 2018, doi: 10.1155/2018/8545347.

[44] M. Monguió-Tortajada, C. Gálvez-Montón, A. Bayes-Genis, S. Roura, and F. E. Borràs, “Extracellular vesicle isolation methods: rising impact of size-exclusion chromatography,” *Cellular and Molecular Life Sciences*, vol. 76, no. 12, pp. 2369–2382, Jun. 2019, doi: 10.1007/s00018-019-03071-y.

[45] Z. Zhao, H. Wijerathne, A. K. Godwin, and S. A. Soper, “Isolation and analysis methods of extracellular vesicles (EVs),” *Extracell Vesicles Circ Nucl Acids*, 2021, doi: 10.20517/evcna.2021.07.

[46] S.-C. Guo, S.-C. Tao, and H. Dawn, “Microfluidics-based on-a-chip systems for isolating and analysing extracellular vesicles,” *J Extracell Vesicles*, vol. 7, no. 1, p. 1508271, Dec. 2018, doi: 10.1080/20013078.2018.1508271.

[47] S. Zhang *et al.*, “Advanced microfluidic technologies for isolating extracellular vesicles,” *TrAC Trends in Analytical Chemistry*, vol. 157, p. 116817, Dec. 2022, doi: 10.1016/j.trac.2022.116817.

[48] S. Razavi Bazaz, S. Zhand, R. Salomon, E. H. Beheshti, D. Jin, and M. E. Warkiani, “ImmunoInertial microfluidics: A novel strategy for isolation of small EV

subpopulations," *Appl Mater Today*, vol. 30, p. 101730, Feb. 2023, doi: 10.1016/j.apmt.2022.101730.

[49] X. Dong *et al.*, "Efficient isolation and sensitive quantification of extracellular vesicles based on an integrated ExoID-Chip using photonic crystals," *Lab Chip*, vol. 19, no. 17, pp. 2897–2904, 2019, doi: 10.1039/C9LC00445A.

[50] S.-C. Guo, S.-C. Tao, and H. Dawn, "Microfluidics-based on-a-chip systems for isolating and analysing extracellular vesicles," *J Extracell Vesicles*, vol. 7, no. 1, p. 1508271, Dec. 2018, doi: 10.1080/20013078.2018.1508271.

[51] B. Talebjedi, N. Tasnim, M. Hoorfar, G. F. Mastromonaco, and M. De Almeida Monteiro Melo Ferraz, "Exploiting Microfluidics for Extracellular Vesicle Isolation and Characterization: Potential Use for Standardized Embryo Quality Assessment," *Front Vet Sci*, vol. 7, Jan. 2021, doi: 10.3389/fvets.2020.620809.

[52] D. Raju, S. Bathini, S. Badilescu, A. Ghosh, and M. Packirisamy, "Microfluidic Platforms for the Isolation and Detection of Exosomes: A Brief Review," *Micromachines (Basel)*, vol. 13, no. 5, p. 730, Apr. 2022, doi: 10.3390/mi13050730.

[53] A. Gualerzi *et al.*, "Raman spectroscopy uncovers biochemical tissue-related features of extracellular vesicles from mesenchymal stromal cells," *Sci Rep*, vol. 7, no. 1, p. 9820, Aug. 2017, doi: 10.1038/s41598-017-10448-1.

[54] A. Samoylenko *et al.*, "Time-gated Raman spectroscopy and proteomics analyses of hypoxic and normoxic renal carcinoma extracellular vesicles," *Sci Rep*, vol. 11, no. 1, p. 19594, Oct. 2021, doi: 10.1038/s41598-021-99004-6.

[55] H. Zhang and D. Lyden, "Asymmetric-flow field-flow fractionation technology for exomere and small extracellular vesicle separation and characterization," *Nat Protoc*, vol. 14, no. 4, pp. 1027–1053, Apr. 2019, doi: 10.1038/s41596-019-0126-x.

[56] H. Kang, J. Kim, and J. Park, "Methods to isolate extracellular vesicles for diagnosis," *Micro and Nano Systems Letters*, vol. 5, no. 1, p. 15, Dec. 2017, doi: 10.1186/s40486-017-0049-7.

[57] D. Bachurski *et al.*, "Extracellular vesicle measurements with nanoparticle tracking analysis – An accuracy and repeatability comparison between NanoSight NS300 and ZetaView," *J Extracell Vesicles*, vol. 8, no. 1, p. 1596016, Dec. 2019, doi: 10.1080/20013078.2019.1596016.

[58] H. Gwak *et al.*, "Microfluidic chip for rapid and selective isolation of tumor-derived extracellular vesicles for early diagnosis and metastatic risk evaluation of breast cancer," *Biosens Bioelectron*, vol. 192, p. 113495, Nov. 2021, doi: 10.1016/j.bios.2021.113495.

[59] R. T. Davies, J. Kim, S. C. Jang, E.-J. Choi, Y. S. Gho, and J. Park, "Microfluidic filtration system to isolate extracellular vesicles from blood," *Lab Chip*, vol. 12, no. 24, p. 5202, 2012, doi: 10.1039/c2lc41006k.

[60] Y.-Q. Xu *et al.*, "A Novel Microfluidic Chip for Fast, Sensitive Quantification of Plasma Extracellular Vesicles as Biomarkers in Patients With Osteosarcoma," *Front Oncol*, vol. 11, Aug. 2021, doi: 10.3389/fonc.2021.709255.

[61] Y. Kang *et al.*, "Isolation and Profiling of Circulating Tumor-Associated Exosomes Using Extracellular Vesicular Lipid–Protein Binding Affinity Based Microfluidic Device," *Small*, vol. 15, no. 47, p. 1903600, Nov. 2019, doi: 10.1002/smll.201903600.

[62] Y. Tian *et al.*, “Quality and efficiency assessment of six extracellular vesicle isolation methods by nano-flow cytometry,” *J Extracell Vesicles*, vol. 9, no. 1, p. 1697028, Sep. 2020, doi: 10.1080/20013078.2019.1697028.

[63] H. M. Tay *et al.*, “Direct isolation of circulating extracellular vesicles from blood for vascular risk profiling in type 2 diabetes mellitus,” *Lab Chip*, vol. 21, no. 13, pp. 2511–2523, 2021, doi: 10.1039/D1LC00333J.

[64] E. A. Mol, M.-J. Goumans, P. A. Doevendans, J. P. G. Sluijter, and P. Vader, “Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation,” *Nanomedicine*, vol. 13, no. 6, pp. 2061–2065, Aug. 2017, doi: 10.1016/j.nano.2017.03.011.

[65] T. Liangsupree, E. Multia, and M.-L. Riekkola, “Modern isolation and separation techniques for extracellular vesicles,” *J Chromatogr A*, vol. 1636, p. 461773, Jan. 2021, doi: 10.1016/j.chroma.2020.461773.

[66] N. Karimi *et al.*, “Detailed analysis of the plasma extracellular vesicle proteome after separation from lipoproteins,” *Cellular and Molecular Life Sciences*, vol. 75, no. 15, pp. 2873–2886, Aug. 2018, doi: 10.1007/s00018-018-2773-4.

[67] R. Stranska *et al.*, “Comparison of membrane affinity-based method with size-exclusion chromatography for isolation of exosome-like vesicles from human plasma,” *J Transl Med*, vol. 16, no. 1, p. 1, Dec. 2018, doi: 10.1186/s12967-017-1374-6.

[68] E. Zeringer, T. Barta, M. Li, and A. V. Vlassov, “Strategies for Isolation of Exosomes,” *Cold Spring Harb Protoc*, vol. 2015, no. 4, p. pdb.top074476, Apr. 2015, doi: 10.1101/pdb.top074476.

[69] L. Doyle and M. Wang, “Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis,” *Cells*, vol. 8, no. 7, p. 727, Jul. 2019, doi: 10.3390/cells8070727.

[70] D. Duijvesz *et al.*, “Differential tissue expression of extracellular vesicle-derived proteins in prostate cancer,” *Prostate*, vol. 79, no. 9, pp. 1032–1042, Jun. 2019, doi: 10.1002/pros.23813.

[71] E. Willms *et al.*, “Cells release subpopulations of exosomes with distinct molecular and biological properties,” *Sci Rep*, vol. 6, no. 1, p. 22519, Mar. 2016, doi: 10.1038/srep22519.

[72] L. Saludas *et al.*, “Isolation methods of large and small extracellular vesicles derived from cardiovascular progenitors: A comparative study,” *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 170, pp. 187–196, Jan. 2022, doi: 10.1016/j.ejpb.2021.12.012.

[73] S. Ghadami and K. Dellinger, “The lipid composition of extracellular vesicles: applications in diagnostics and therapeutic delivery,” *Front Mol Biosci*, vol. 10, Jul. 2023, doi: 10.3389/fmolb.2023.1198044.

[74] D. R. Sojka *et al.*, “Heat shock protein A2 is a novel extracellular vesicle-associated protein,” *Sci Rep*, vol. 13, no. 1, p. 4734, Mar. 2023, doi: 10.1038/s41598-023-31962-5.

[75] C. Krafft, L. Neudert, T. Simat, and R. Salzer, “Near infrared Raman spectra of human brain lipids,” *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 61, no. 7, pp. 1529–1535, May 2005, doi: 10.1016/j.saa.2004.11.017.

[76] G. Raposo and W. Stoorvogel, “Extracellular vesicles: Exosomes, microvesicles, and friends,” *Journal of Cell Biology*, vol. 200, no. 4, pp. 373–383, Feb. 2013, doi: 10.1083/jcb.201211138.

[77] K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor, and M. Baranska, “Raman spectroscopy of lipids: a review,” *Journal of Raman Spectroscopy*, vol. 46, no. 1, pp. 4–20, Jan. 2015, doi: 10.1002/jrs.4607.

[78] Z. Movasaghi, S. Rehman, and I. U. Rehman, “Raman Spectroscopy of Biological Tissues,” *Appl Spectrosc Rev*, vol. 42, no. 5, pp. 493–541, Sep. 2007, doi: 10.1080/05704920701551530.

[79] A. Gualerzi *et al.*, “Raman spectroscopy as a quick tool to assess purity of extracellular vesicle preparations and predict their functionality,” *J Extracell Vesicles*, vol. 8, no. 1, p. 1568780, Dec. 2019, doi: 10.1080/20013078.2019.1568780.

[80] J. M. Cameron *et al.*, “Vibrational spectroscopic analysis and quantification of proteins in human blood plasma and serum,” in *Vibrational Spectroscopy in Protein Research*, Elsevier, 2020, pp. 269–314. doi: 10.1016/B978-0-12-818610-7.00010-4.

[81] M. Palviainen *et al.*, “Extracellular vesicles from human plasma and serum are carriers of extravesicular cargo—Implications for biomarker discovery,” *PLoS One*, vol. 15, no. 8, p. e0236439, Aug. 2020, doi: 10.1371/journal.pone.0236439.

[82] S. Caruso and I. K. H. Poon, “Apoptotic cell-derived extracellular vesicles: More than just debris,” *Frontiers in Immunology*, vol. 9, no. JUN. Frontiers Media S.A., 2018. doi: 10.3389/fimmu.2018.01486.

[83] D. D. Gonda *et al.*, “Neuro-oncologic applications of exosomes, microvesicles, and other nano-sized extracellular particles,” *Neurosurgery*, vol. 72, no. 4, pp. 501–510, 2013, doi: 10.1227/NEU.0b013e3182846e63.

[84] A. V Vlassov, S. Magdaleno, R. Setterquist, and R. Conrad, “Exosomes: Current knowledge of their composition, biological functions, and diagnostic and therapeutic potentials,” *Biochimica et Biophysica Acta - General Subjects*, vol. 1820, no. 7. pp. 940–948, 2012. doi: 10.1016/j.bbagen.2012.03.017.

[85] K. Brennan *et al.*, “A comparison of methods for the isolation and separation of extracellular vesicles from protein and lipid particles in human serum,” *Sci Rep*, vol. 10, no. 1, p. 1039, Jan. 2020, doi: 10.1038/s41598-020-57497-7.

[86] M. Yáñez-Mó *et al.*, “Biological properties of extracellular vesicles and their physiological functions,” *Journal of Extracellular Vesicles*, vol. 4, no. 2015. Co-Action Publishing, pp. 1–60, 2015. doi: 10.3402/jev.v4.27066.

[87] G. Van Niel, G. D’Angelo, and G. Raposo, “Shedding light on the cell biology of extracellular vesicles,” *Nature Reviews Molecular Cell Biology*, vol. 19, no. 4. Nature Publishing Group, pp. 213–228, 2018. doi: 10.1038/nrm.2017.125.

[88] S. Fu, Y. Zhang, Y. Li, L. Luo, Y. Zhao, and Y. Yao, “Extracellular vesicles in cardiovascular diseases.,” *Cell Death Discov*, vol. 6, p. 68, 2020, doi: 10.1038/s41420-020-00305-y.

[89] V. Dinakaran, A. Rathinavel, M. Pushpanathan, R. Sivakumar, P. Gunasekaran, and J. Rajendhran, “Elevated Levels of Circulating DNA in Cardiovascular Disease Patients: Metagenomic Profiling of Microbiome in the Circulation,” *PLoS One*, vol. 9, no. 8, p. e105221, Aug. 2014, doi: 10.1371/journal.pone.0105221.

[90] S. Ekholm and H. Karppanen, “Cardiovascular effects of L-tyrosine: influence of blockade of tyrosine metabolism,” *Eur J Pharmacol*, vol. 163, no. 2–3, pp. 209–217, Apr. 1989, doi: 10.1016/0014-2999(89)90189-1.

[91] S. Ekholm and H. Karppanen, “Cardiovascular effects of L-tyrosine in normotensive and hypertensive rats,” *Eur J Pharmacol*, vol. 143, no. 1, pp. 27–34, Nov. 1987, doi: 10.1016/0014-2999(87)90731-X.

[92] H. Mangge, W. J. Schnedl, S. Schröcksnadel, S. Geisler, C. Murr, and D. Fuchs, “Immune activation and inflammation in patients with cardiovascular disease are associated with elevated phenylalanine-to-tyrosine ratios,” *Pteridines*, vol. 24, no. 1, pp. 51–55, Jun. 2013, doi: 10.1515/pterid-2013-0002.

[93] M. Onozato *et al.*, “Amino acid analyses of the exosome-eluted fractions from human serum by HPLC with fluorescence detection.,” *Pract Lab Med*, vol. 12, p. e00099, Nov. 2018, doi: 10.1016/j.plabm.2018.e00099.

[94] H. Mangge, I. Stelzer, E. Z. Reininghaus, D. Weghuber, T. T. Postolache, and D. Fuchs, “Disturbed tryptophan metabolism in cardiovascular disease.,” *Curr Med Chem*, vol. 21, no. 17, pp. 1931–7, Jun. 2014, doi: 10.2174/09298673216661403041055 26.

[95] E. Yu *et al.*, “Increases in Plasma Tryptophan Are Inversely Associated with Incident Cardiovascular Disease in the Prevención con Dieta Mediterránea (PREDIMED) Study,” *J Nutr*, p. jn241711, Feb. 2017, doi: 10.3945/jn.116.241711.

[96] M. D. Shapiro and S. Fazio, “From Lipids to Inflammation,” *Circ Res*, vol. 118, no. 4, pp. 732–749, Feb. 2016, doi: 10.1161/CIRCRESAHA.115.306471.

[97] M. V. Holmes *et al.*, “Mendelian randomization of blood lipids for coronary heart disease,” *Eur Heart J*, vol. 36, no. 9, pp. 539–550, Mar. 2015, doi: 10.1093/eurheartj/eht571.

[98] K. Leander *et al.*, “Circulating Proprotein Convertase Subtilisin/Kexin Type 9 (PCSK9) Predicts Future Risk of Cardiovascular Events Independently of Established Risk Factors,” *Circulation*, vol. 133, no. 13, pp. 1230–1239, Mar. 2016, doi: 10.1161/CIRCULATIONAHA.115.018531.

[99] T. Skotland, K. Sagini, K. Sandvig, and A. Llorente, “An emerging focus on lipids in extracellular vesicles,” *Adv Drug Deliv Rev*, vol. 159, pp. 308–321, 2020, doi: 10.1016/j.addr.2020.03.002.

[100] J. Jansson *et al.*, “Metabolomics Reveals Metabolic Biomarkers of Crohn’s Disease,” *PLoS One*, vol. 4, no. 7, p. e6386, Jul. 2009, doi: 10.1371/journal.pone.0006386.

[101] E. A. Franzosa *et al.*, “Gut microbiome structure and metabolic activity in inflammatory bowel disease,” *Nat Microbiol*, vol. 4, no. 2, pp. 293–305, Dec. 2018, doi: 10.1038/s41564-018-0306-4.

[102] G. Leoni *et al.*, “Annexin A1–containing extracellular vesicles and polymeric nanoparticles promote epithelial wound repair,” *Journal of Clinical Investigation*, vol. 125, no. 3, pp. 1215–1227, Mar. 2015, doi: 10.1172/JCI76693.

[103] X. Zheng *et al.*, “Salivary exosomal PSMA7: a promising biomarker of inflammatory bowel disease,” *Protein Cell*, vol. 8, no. 9, pp. 686–695, Sep. 2017, doi: 10.1007/s13238-017-0413-7.

[104] J. T. Bjerrum *et al.*, “Metabolomics in Ulcerative Colitis: Diagnostics, Biomarker Identification, And Insight into the Pathophysiology,” *J Proteome Res*, vol. 9, no. 2, pp. 954–962, Feb. 2010, doi: 10.1021/pr9008223.

[105] R. Schicho *et al.*, “Quantitative Metabolomic Profiling of Serum, Plasma, and Urine by 1 H NMR Spectroscopy Discriminates between Patients with

Inflammatory Bowel Disease and Healthy Individuals," *J Proteome Res*, vol. 11, no. 6, pp. 3344–3357, Jun. 2012, doi: 10.1021/pr300139q.

[106] T. S. Martins, M. Vaz, and A. G. Henriques, "A review on comparative studies addressing exosome isolation methods from body fluids," *Anal Bioanal Chem*, vol. 415, no. 7, pp. 1239–1263, Mar. 2023, doi: 10.1007/s00216-022-04174-5.

[107] K. Li *et al.*, "A signature of saliva-derived exosomal small RNAs as predicting biomarker for esophageal carcinoma: a multicenter prospective study," *Mol Cancer*, vol. 21, no. 1, p. 21, Dec. 2022, doi: 10.1186/s12943-022-01499-8.

[108] S. Sjoqvist and K. Otake, "Saliva and Saliva Extracellular Vesicles for Biomarker Candidate Identification—Assay Development and Pilot Study in Amyotrophic Lateral Sclerosis," *Int J Mol Sci*, vol. 24, no. 6, p. 5237, Mar. 2023, doi: 10.3390/ijms24065237.

[109] N. Comfort *et al.*, "Isolation and characterization of extracellular vesicles in saliva of children with asthma," *Extracell Vesicles Circ Nucl Acids*, 2021, doi: 10.20517/evcna.2020.09.

[110] M. Wu *et al.*, "Separating extracellular vesicles and lipoproteins via acoustofluidics.," *Lab Chip*, vol. 19, no. 7, pp. 1174–1182, Mar. 2019, doi: 10.1039/c8lc01134f.

[111] L. Cui *et al.*, "New frontiers in salivary extracellular vesicles: transforming diagnostics, monitoring, and therapeutics in oral and systemic diseases," *J Nanobiotechnology*, vol. 22, no. 1, p. 171, Apr. 2024, doi: 10.1186/s12951-024-02443-2.

[112] A. Rai, H. Fang, B. Claridge, R. J. Simpson, and D. W. Greening, "Proteomic dissection of large extracellular vesicle surfaceome unravels interactive surface platform," *J Extracell Vesicles*, vol. 10, no. 13, Nov. 2021, doi: 10.1002/jev2.12164.

[113] A. Radeghieri *et al.*, "Active antithrombin glycoforms are selectively physiosorbed on plasma extracellular vesicles," *Journal of Extracellular Biology*, vol. 1, no. 9, Sep. 2022, doi: 10.1002/jex2.57.

[114] S. Jamaly *et al.*, "Impact of preanalytical conditions on plasma concentration and size distribution of extracellular vesicles using Nanoparticle Tracking Analysis," *Sci Rep*, vol. 8, no. 1, p. 17216, Nov. 2018, doi: 10.1038/s41598-018-35401-8.

[115] S. Jamaly *et al.*, "Impact of preanalytical conditions on plasma concentration and size distribution of extracellular vesicles using Nanoparticle Tracking Analysis," *Sci Rep*, vol. 8, no. 1, p. 17216, Nov. 2018, doi: 10.1038/s41598-018-35401-8.

[116] Y. Sun *et al.*, "Facile preparation of salivary extracellular vesicles for cancer proteomics," *Sci Rep*, vol. 6, no. 1, p. 24669, Apr. 2016, doi: 10.1038/srep24669.

[117] Y. Yuana *et al.*, "Handling and storage of human body fluids for analysis of extracellular vesicles," *J Extracell Vesicles*, vol. 4, no. 1, Jan. 2015, doi: 10.3402/jev.v4.29260.

CHAPTER 8

SUMMARY AND FUTURE WORK

8.1 Summary

This thesis presents a multifaceted exploration of the application of Raman spectroscopy in advancing healthcare diagnostics across two distinct domains – IBD and CVD. Furthermore, it bridges the gap between bench and bedside by incorporating the development of a lab-on-a-chip platform for the isolation of EVs as potential biomarkers for these diseases.

The foundational chapters provide an extensive review of RS principles, emphasising its molecular specificity, non-invasive nature, and potential diagnostic utility. Subsequent Chapters 3-6 detail the application of RS in elucidating molecular signatures associated with CVD and IBD. Through meticulous analysis, we uncover distinct Raman biomarkers that exhibit promise in early disease detection and characterisation in the form of lipids for CVD and cytokines for IBD. Additionally introduced, is a comparative analysis with mass spectrometry, showcasing how RS stands as a viable alternative with its advantages in speed, simplicity, and compatibility.

The highly innovative aspect of this thesis lies in the integration of RS with an LoC platform outlined in Chapter 7. The LoC is designed for the combined on-chip isolation and downstream analysis of EVs, nano vesicles released by cells that carry molecular cargoes reflective of the cellular environment. Key design elements including stepwise isolation of EVs lead the way for the isolation of multiple populations of EVs ranging in size from 40 – 5000 nm. Leveraging RS on EVs, we showcase its potential as a non-invasive tool for disease diagnostics and monitoring. The LoC, a promising technology

for streamlined EV isolation, is poised to overcome current limitations such as scalability and standardisation, and practical implementation in routine clinical settings. The research not only identifies these challenges but proposes strategic avenues for refinement, aiming to make the LoC technology a robust and reliable tool for point-of-care diagnostics. The thesis also addresses the clinical translation of these findings. The development of the LoC technology holds immense promise for the routine diagnostics, offering a rapid and cost-effective method for EV isolation. Challenges such as standardisation and scalability are acknowledged, prompting future research directions aimed at optimising and validating this innovative platform.

The broader significance of this work lies in its contribution to the evolution of diagnostic methodologies for CVD and IBD. By integrating RS with cutting-edge microfluidic technologies and artificial intelligence, we move closer to realising personalised and non-invasive diagnostic approaches. In contemplating the clinical translation of these technologies, the thesis lays the groundwork for continued interdisciplinary research, fostering collaboration between researchers, engineers, and clinicians to propel these innovations towards clinical implementation with their seamless introduction into the fabric of modern healthcare, thus transforming the landscape of disease diagnosis and monitoring.

8.2 Future Outlook

The promising outcomes of this thesis pave the way for a multitude of avenues for future research, bridging the fields of RS and healthcare, with several initiatives proposed to enhance the impact and applicability of the developed methodologies.

Firstly, it will be necessary to undertake meticulous large-scale clinical validation studies with an emphasis on diverse patient cohorts, considering factors such as age, gender, and comorbidities. This will include the development of an internal reference system, such as a specific molecular marker, to serve as a constant benchmark, allowing for real-time calibration adjustments and normalisation. Our preliminary work looked to investigate such specific biomarkers of interest, however while the connection between specific chemical bonds and their corresponding spectra is well understood, interpreting the biological implications of such changes in the abundance or reduction of certain molecules can be more complex and may only sometimes have a straightforward interpretation. Here, comparison with diagnostic modalities will be essential to cement biomarkers, such as our identified cytokines and lipids as internal references. Collaboration with clinical experts to collect parallel datasets using traditional diagnostic methods, such as histopathology and close comparison with Raman-based outcomes will be central to clinical validation of our technologies and positioning them within the existing diagnostic landscape.

In the pursuit of advancing diagnostic capabilities, the integration of RS with other imaging modalities represents a paradigm shift towards a multimodal approach. By initiating research ventures into the integration of RS with other high-resolution imaging modalities such as mass spectrometry or advanced magnetic resonance imaging, such an approach could offer a more holistic diagnostic platform, capturing a spectrum of molecular information for a nuanced understanding of disease. For example, mass spectrometry excels in the identification and characterisation of complex mixtures and the trace detection of compounds. Although the introduction of SERS methodologies

has made this possible with RS, SERS-active substrates are expensive and achieving high reproducibility, especially in the laboratory setting can be challenging.

A further area of focus should be on patient stratification and personalised medicine. By propelling the research into the realm of precision medicine *via* conducting in-depth investigations into the utility of identified biomarkers of disease for patient stratification, it allows the exploration of correlations with genotypic variations, patient responses to treatment, as well as long term outcomes thus, setting the stage for personalised therapeutic interventions.

In addition, efforts should be focussed on the refinement of the LoC platform, delving into the microfluidic engineering approach to enhance EV isolation efficiency and explore the frontiers of automation and miniaturisation, aiming not just for point-of-care applications but for a revolutionary shift in how we approach extracellular vesicle analysis in the clinical setting. Microfluidic 3D printing as adopted in our approach allows design flexibility, rapid prototyping, customisation, integration of multiple layers and low setup costs, however compared to traditional lithography methods 3D microfluidic printing is associated with a more limited range of materials. For example, further adaption of the presently designed LoC may in the future include an enclosed system to ensure sample containment and prevent exposure to the end user.

The integration of RS into clinical laboratories for healthcare applications is an exciting area with considerable potential, however future work in this field should also focus on

addressing the following key aspects to facilitate its translation into routine clinical practice:

- Standardisation - Establishing standardized protocols and validation procedures is crucial for ensuring the reliability and reproducibility of Raman spectroscopy results across different laboratories and instruments. Developing consensus guidelines and participating in collaborative validation studies will contribute to building a robust foundation for clinical use.
- Automation and high-throughput platforms - Developing automated and high-throughput Raman spectroscopy platforms is critical for handling the demands of clinical laboratories. Automation can improve efficiency, reduce operator variability, and make the technology more amenable to routine diagnostic workflows.
- Biological tissue imaging - Enhancing the capabilities of Raman spectroscopy for *in vivo* and *ex vivo* imaging of biological tissues can open new avenues for non-invasive diagnostics and surgical guidance. Improving the depth of penetration and spatial resolution is essential for accurate tissue characterisation.
- Instrument miniaturisation - Miniaturising Raman spectroscopy instruments can contribute to the development of handheld or portable devices suitable for point-of-care applications. Smaller, more user-friendly instruments may facilitate broader adoption in diverse clinical settings.

Collaborative efforts among researchers, clinicians, engineers, and regulatory bodies are pivotal for advancing RS in healthcare. By addressing these challenges and opportunities, RS has the potential to become a valuable tool for non-invasive diagnostics, disease monitoring, and personalised medicine in the clinical setting.

APPENDICES

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 2

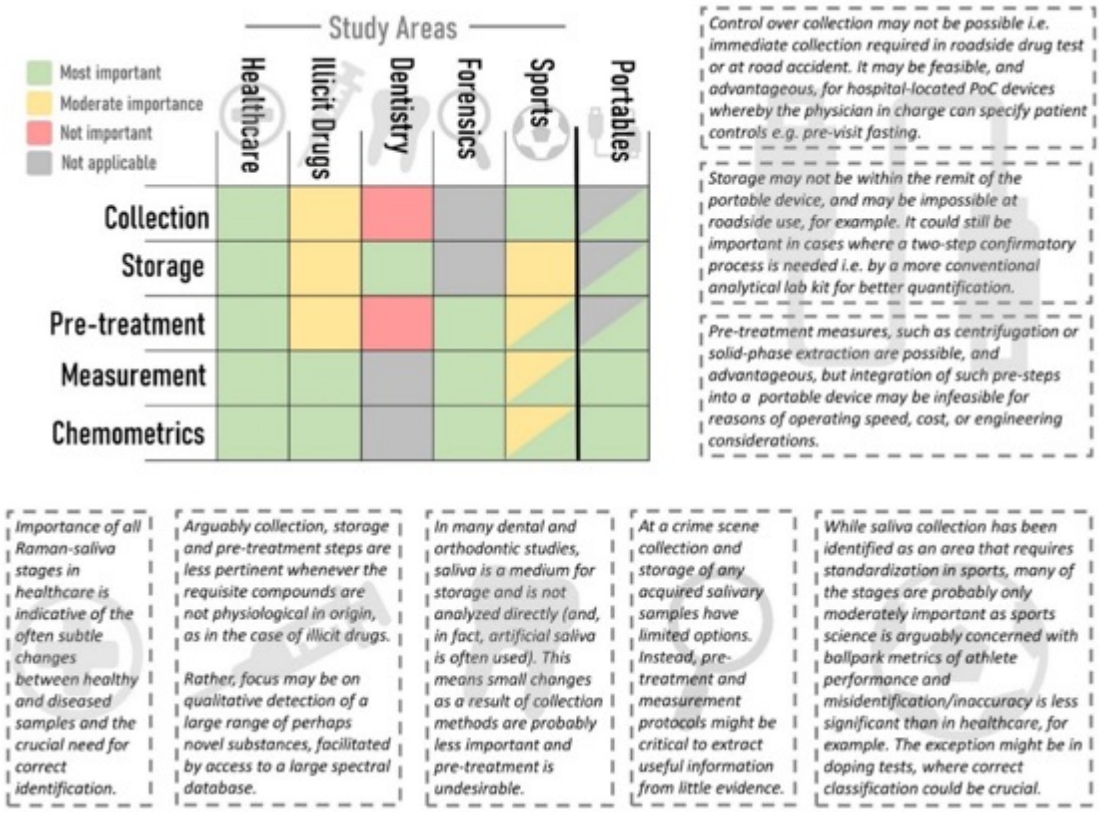

Figure SA.1. Importance of saliva methods in collection, storage, pre-treatment measures, Raman measurement protocols and chemometrics/data analysis to application areas. Further details on different Raman-saliva methodologies are summarised in Ref. (1).

Table SA.1. Raman saliva studies published per year, separated by category from 1990-2021. (DO: Dentistry, HC: Healthcare, ID: Illicit Drugs, F: Forensics and O: Other).

	DO	HC	ID	F	O
1990	0	0	0	0	0
1991	0	0	0	0	0
1992	0	0	0	0	0
1993	0	0	0	0	0
1994	0	0	0	0	0
1995	0	0	0	0	0
1996	0	0	0	0	0
1997	0	0	0	0	0
1998	0	0	0	0	0
1999	1	0	0	0	0
2000	0	0	0	0	0
2001	0	0	0	0	0
2002	0	0	0	0	0
2003	0	0	0	0	0
2004	0	0	0	0	0
2005	0	2	0	0	0
2006	0	0	0	0	0
2007	0	1	0	0	0
2008	1	1	0	1	1
2009	1	1	1	0	0
2010	1	1	1	2	1
2011	1	0	2	0	0
2012	1	2	0	0	0
2013	2	2	1	0	0
2014	6	5	0	0	0
2015	1	3	2	0	0
2016	3	6	0	2	0
2017	2	6	0	0	0
2018	7	9	4	0	0
2019	6	9	4	0	0
2020	1	6	2	1	0
2021	0	1	0	0	0
TOTAL	34	55	17	6	2
				TOTAL	**114**

Table SA.2. Raman saliva cumulative studies by year and category 1999-2021.

Author	Year	Category
Paluszkiewicz	2009	DO
Cândido	2019	DO

Daood	2015	DO
Silva Soares	2016	DO
Toledano	2014	DO
Daood	2018	DO
Musa Trolic	2019	DO
Wachesk	2016	DO
Wang	2014	DO
Wang	2014	DO
Wei	2010	DO
Zaharia	2014	DO
Zaharia	2017	DO
Hussein	2018	DO
Zhang	2018	DO
Zhang	2019	DO
Daood	2019	DO
da Silva	2008	DO
Fawzy	2019	DO
Hua	2020	DO
Iafisco	2018	DO
Li	1999	DO
Osorio	2014	DO
Osorio	2014	DO
Osorio	2018	DO
Silveira	2018	DO
Soares	2013	DO
Gonchukov	2013	DO
Gonchukov	2012	DO
Veys-Renaux	2016	DO
Condò	2017	DO
Passos	2018	DO
Hernández-Cedillo	2019	DO
Braga	2011	DO
Lawanstiend	2018	HC
Altuntas	2018	HC
Zhou	2019	HC
Kah	2007	HC
Malkovskiy	2019	HC
Maitra	2020	HC
Colceriu-Şimon	2019	HC
Feng	2015	HC
Li	2012	HC
Qian	2018	HC
Ralbovsky	2019	HC

Zermeño-Nava	2018	HC
Farquharson	2008	HC
Farquharson	2005	HC
Lin	2017	HC
Qiu	2016	HC
Eom	2019	HC
Li	2012	HC
Hernandez-Arteaga	2017	HC
Hernandez-Arteaga	2019	HC
Stefancu	2019	HC
Yan	2009	HC
Rekha	2016	HC
Žukovskaja	2017	HC
Brindha	2016	HC
Hou	2017	HC
Taniguchi	2013	HC
Feng	2014	HC
Zamora-Mendoza	2019	HC
Wu	2016	HC
Cao	2015	HC
Chen	2014	HC
Chen	2018	HC
Connolly	2016	HC
Falamas	2020	HC
Lin	2018	HC
Al-Ogaidi	2014	HC
Wang	2010	HC
Lotfi	2018	HC
Desai	2020	HC
Carlomagno	2020	HC
Wu	2014	HC
Yang	2018	HC
Othmana	2017	HC
Othman	2018	HC
Othman	2016	HC
Othmanb	2017	HC
Hole	2020	HC
Moisou	2020	HC
Zhang	2019	HC
Velička	2021	HC
Farquharson	2005	HC
Zheng	2015	HC
Liu	2014	HC

Radzol	2013	HC
Deriu	2019	ID
Salemmilani	2018	ID
Inscore	2011	ID
D'Elia	2018	ID
Qu	2010	ID
Dana	2015	ID
Farquharson	2011	ID
Anyu	2009	ID
Yang	2015	ID
Andreou	2013	ID
Sivashanmugan_1 (Tetrahydrocannabinol sensing)	2019	ID
Dies	2018	ID
Li	2020	ID
Sivashanmugan_2 (Trace detection)	2019	ID
Su	2019	ID
Mohammadi	2018	ID
Hong	2020	ID
Virkler	2010	F
Sikirzhytski	2010	F
Muro_1 (Forensic)	2016	F
Muro_2 (Sex Disc)	2016	F
Virkler	2008	F
Al-Hetlani	2020	F
Yuen	2010	O
Yuen	2008	O
Total	**114**	

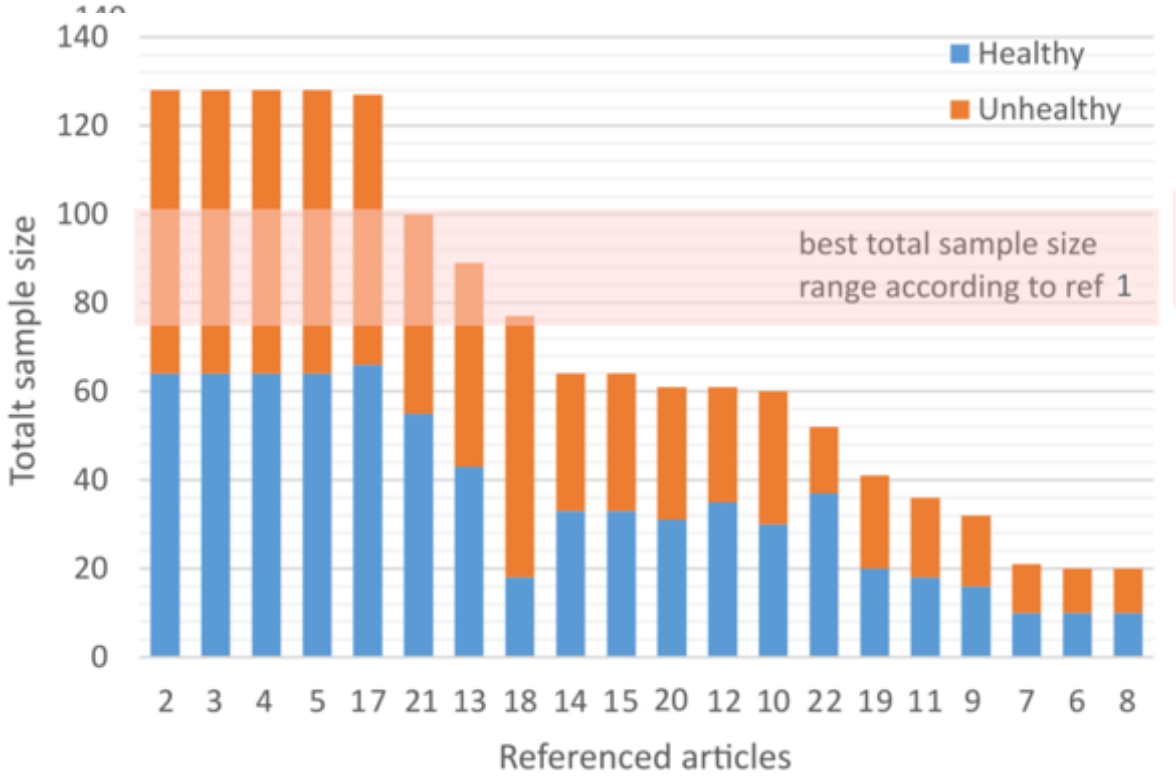

Figure SA.2. Stacked bar plot of the variation in sample size between different research papers. In orange, the number of unhealthy samples and the number of healthy samples in blue.

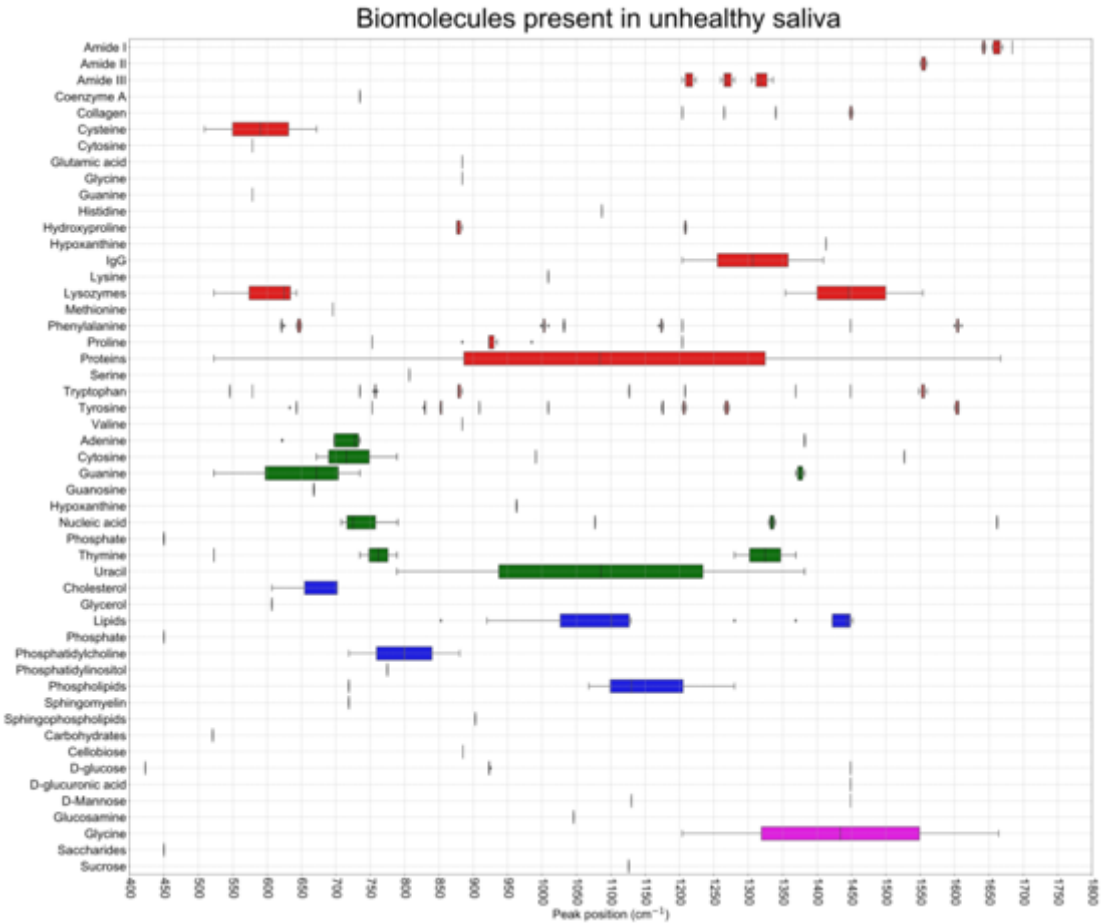

Figure SA.3. List of biomolecules present in saliva from cancerous patients with respective whisker plots for different Raman peak regions. Red is related to protein molecules, green nucleic acid, magenta sugars and blue lipids. When the peak assignment is singular a stripe represents the symbol-complemented with **Table SA.3**.

Table SA.3. Raman biomolecule information from cancerous patient saliva.

Protein Tentative assignment	Peak position (cm^{-1})
Amide I	1640[6],1645[28],1655[10,7],1661[8],1664[12],1667[7],1670[13],1684[15]
Amide II	1550[6],1555[15],1560[18]
Amide III	1204[11],1224[11],1260[6],1265[15],1268[17],1275[11],1280[19],1305[29],1318[30],1336[12]
Coenzyme A	735[19]
Collagen	1204[11],1265[15],1340[11],1447[15],1452[8]
Cysteine	509[13],672[17]
Cytosine	579[18]
Glutamic acid	884[19]
Glycine	884[19]
Guanine	579[18]
Histidine	1087[9]
Hydroxyproline	876[11,15],880[6],884[19],1207[7],1210[13]
Hypoxanthine	1413[7]
IgG	1204[11],1409[11]
Lysine	1009[19]
Lysozyme	523[19],625[7],643[7],1354[9],1446[9],1554[7]
Methionine	696[19]
Phenylalanine	620[6],621[7,10,13],622[19],625[7],643[11],650[6],999[17,16],1002[6,13],1003[5,10],1004[15,29],1009[19],1031[13],1032[10],1033[13],1170[6],1174[9,10],1176[7],1204[11],1449[17],1600[17],1603[18],1605[7],1606[16],1610[6]
Proline	753[7],884[19],923[7,17,18],925[19],935[15],985[7],1204[11]
Proteins	523[19],622[19],723[18],732[17],853[29],920[29],1047[6],1049[11],1077[19],1084[6,13],1125[29],1280[19],1369[19],1440[29],1449[13,17],1667[7]
Serine	807[6]
Tryptophan	546[18],579[18],735[19],755[13],757[7],758[6,10],760[11],878[9,10],880[7],884[19],1126[7],1127[9,10],1208[9,13],1369[19],1449[9,10],1548[9],1552[13],1554[7],1556[9,10],1560[18]
Tyrosine	633[6],643[7,9,10,13],753[7],828[13],830[7,9,10],852[17,18],853[13],854[9,10,15],909[19],1009[19],1174[18],1176[7,9],1204[15],1206[10],1207[7],1210[13],1265[9],1272[7],1600[17],1603[18],1606[9,10]
Valine	884[19]

Nuclei Acid Tentative assignment	Peak position (cm^{-1})
Adenine	622[19],723[18],732[17],735[19],1382[18]
Cytosine	672[17],696[19],735[19],789[19],991[9],1527[18]
Guanine	523[19],672[17],735[19],1369[19],1382[18]
Guanosine	668[29]
Hypoxanthine	963[7]
Unspecified	450[29],708[9],725[4],790[6],1077[19],1330[13],1339[6],1662[8]
Thymine	523[19],735[19],789[19],1280[19],1369[19]
Uracil	789[19],1382[18]

Lipids Tentative assignment	Peak position (cm^{-1})
Cholestrol	607[7],702[7]
Glycerol	607[7]
Unspecified	853[29],920[29],1062[7],1077[19],1125[29],1126[11],1129[6],1280[19],1369[19],1440[29],1447[15],1452[8]
Phosphatidycholine	719[9],880[7]
Phosphatidylinositol	775[7]
Phospholipids	450[29],719[9],1068[7],1130[18],1280[19]
Sphingomyelin	719[9]
Sphingophospholipids	903[7]

Sugars Tentative assignment	Peak position (cm^{-1})
Carbohydrates	521[9]
Cellobiose	885[15]
D-glucose	423[18,17],923[7,17,18],925[19],1449[18]
D-glucuronic acid	1449[18]
D-mannose	1130[6],1449[18]
Glucosamine	1046[17]
Glycine	1204[11],1664[12]
Saccharides	450[6]
Sucrose	1126[11]

Table SB.1. Representative Raman peaks and corresponding assignments summarised from the RS IBD studies.

Peak / cm^{-1}	Assignment	Reference
387	Uric acid	(69)
425	(δ(CCC) skeletal backbone)	(65)
492	Uric acid	(69)
500-1500	Region of interest for organic molecules	(69)
504	Disulphide bridges in proteins	(9)
523	V(S-S)	(67)
543	Cholesterol	(46)
559	Nucleic acids	(46)
591	Serine	(46)
602	DNA	(37)
606	Cholesterol	(46)
610	(ρ(CH) wagging in proteins)	(65)
627	Carbohydrates	(9)
646	Protein	(37)
685	DNA	(37)
695	Tyrosine	(69)
720	Phosphatidylcholine	(67,68),
743	Heme	(13)
750	Tryptophan, DNA	(68)
757/759/760	Tryptophan -ring breathing vibration	(66), (67), (9,46)
766	Proteins	(37)
781	DNA	(37)
813	ι-Serine and glutathione	(46)
815	Proteins, DNA	(37)
828	DNA (phosphate groups)	(9)
830-850	Tyrosine doublet (protein)	(67), (68)
873	(ρ(CH_2) in proteins)	(65)
875	Protein and lipids	(68)
875, 877	Phosphatidylcholine	(68)
922	Tricaprylin	(69)
940	Protein Backbone	(9)
960	Cholesterol	(37)
1002/1003	Phenylalanine, carotenoids	(66), (67), (13), (68), (62)(37)
1045	Proline	(69)
1068, 1128	Backbone of lipid, protein, and carbohydrate, Trans C-C Stretching	(68)
1080	(υ(C-C) of lipids), DNA & phospholipids	(65), (68)
1086	Backbone of lipid, protein, and carbohydrate	(68)
1087	Hypoxanthine	(69)
1110	C-C vibration mode of the gauche-bonded chain	(46)
1124	Lipids	(69)
1125	Myoglobin (Heme core), phospholipids, proteins	(62)
1160	(β-carotene)	(65)
1153	V(C-N), β-carotene, terpene	(67), (69)
1155	Carotenoids	(62), (9)
1160	(β-carotene)	(65)
1170	Protein	(37)
1199	Tryptophan	(69), (9)
1201-1209	Proteins at Amide III and CH2 wagging vibrations	(37)
1125	Myoglobin (heme core), phospholipids, proteins	(62)
1230-1285	Protein/Amide III	(68), (9)
1244	Amide III, β-sheet	(62)
1245	Heme/Amide III	(13)
1250-1750	Composite Band	(67)
1250	Protein backbone	(9)

1252	Amide III (Constituent of heme)	(66)
1257	Amide III	(69)
1265	(v(C-N) of Amide III)	(67)
1270	Lipids	(68)
1280	Protein Amide III	(46)
1284	Hypoxanthine	(69)
1290	DNA	(37)
1298	CH2 bending of lipids and proteins	(46)
1300	($\delta(CH_2)$ deformation of proteins and lipids	(65)
1303	Total Lipids	(67), (68)
1304	Phospholipids	(66)
1305	CH2-twisting	(13)
1307	Phospholipids, lipids, adenine, myoglobin (heme core)	(62)
1323	Lipids	(46)
1340	DNA	(37)
1342	Protein	(68)
1352	Nucleic acids	(46)
1359	Proteins	(37)
1361/62	DNA	(37)
1368	tryptophan, guanine, thymine, myoglobin (heme core)	(62)
1372	Lipid	(65)
1395	Myoglobin (heme core), Uracil	(62)
1397	Creatine	(69)
1412	Plasmatic lipids	(9)
1416	$V(C\text{-}O)_s$ of COO^- groups	(67)
1440	Shoulder($\delta(CH_2)$ deformation of proteins and lipids), phospholipids, lipids, collagen	(37,62,65)
1447	Lipid and protein	(68)
1450	CH2-deformation	(13)
1468	$\delta(CH_2)$	(67)
1472	CH2 bending of lipids and proteins	(46)
1485-87	DNA	(37)
1518	Carotenoids	(62)
1520	β-carotene, terpene	(69)
1523/25	(β-carotene)	(65), (9)
1548	Amide II	(67)
1549	Deoxy-Myoglobin (Heme core)	(62)
1580	Heme	(13)
1585	Phenylalanine, hydroxyproline, oxy-Myoglobin (heme core)	(62)
1600-1700	Protein	(68)
1617	Intermolecular β-sheet, strong hydrogen bonds (indicative of protein aggregates)	(67)
1620-1680	Amide I band	(37)
1622	C=C stretching mode of tyrosine and tryptophan	(46)
1634	Antiparallel β-sheet	(67)
1640-1654	α-helix structures	(67)
1640	Protein Backbone	(9)
1641	Oxy-Myoglobin (heme core)	(62)
1643	C=C stretching mode of lipids and the protein amide I	(46)
1650	Amide I	(67)
1655	Lipid/Amide I	(68), (9)
1657	C=C stretching in lipids	(66)
1658	(v(C = O)), Amide I and lipids	(65)
1660	Amide I, C=C stretching	(13)
1663	Disordered secondary structures (in proteins)	(67)
1690	Amide I - β-sheet	(62)
1694	Antiparallel β-sheet	(67)
1699	Proteins	(37)
1702	Amino acids aspartic and glutamic acid	(69)
1709	Phospholipids, triglycerides	(62)
1735	Lipid	(68)
1741	(v(C = O) in lipids)	(65)
2762	Phospholipids	(62)
2854	Lipids	(62)

2892	Lipids, proteins	(62)
2936	Lipids, proteins	(62)
2969	Lipids, proteins	(62)
2800-3200	CH stretching intensities	(37,63)

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 3

Spectroscopic Molecular-Fingerprint Profiling of Saliva

-Supporting Information-

*Emma Buchan[1], Liam Kelleher[1], Michael Clancy[1], Jonathan James Stanley Rickard[2] and Pola Goldberg Oppenheimer[1, 3, *]*

SC.1 Input Spectra

The 100 spectra measured across each sample were grouped according to class (summarized in **Table SC.1**). 20 % of the data was randomly selected from each group and reserved as test data, leaving the remaining 80 % for training (T**able SC.2**). Analysis of the training data was performed using SKiNET [39, 41], by randomly passing samples from the training data into the SOM over a number of iterations. SKiNET models were optimized by performing 10-fold cross validation on the training data, and tuning the number of neurons, initial learning rate and number of training steps. The final model used a 5x5grid of neurons (for age), 30625 training steps (5 epochs of the data), with an initial learning rate of 0.2. The initial neighbourhood size was maintained at 2/3 the edge length of the grid and cosine similarity used as the distance metric to determine the best matching unit. Finally, the optimized model was used to classify the previously unused test data, to give an indicator of the classification performance. Classification using the test data were repeated 10 times from separate SOM initializations and an average of the results output as a confusion matrix.

An illustration of the workflow is shown in **Fig. SC.1**. Studies assessing the suitability of Raman spectroscopy as a potential physiological monitoring tool are shown in **Fig. SC.2.** Spectroscopic Raman molecular fingerprint of fasting status (**Fig. 2C.a)** and time of day (**Fig. 4.2b)** are illustrated.

Gender	Age Group	Spectra per Sample	Participants	Total
Male	20-25	100	22	2200
	31-55	100	6	600
	56+	100	4	400
Female	20-25	100	16	1600
	31-55	100	11	1100
	56+	100	11	1100
Total				**7,000**

Table SC.1. Summary of spectra used as inputs for multivariate analysis across the three age groups.

Age Group	Total	Training Data	Test Data
20-25	2200	1760	440
31-55	600	480	120
56+	400	320	80
20-25	1600	1280	320
31-55	1100	880	220
56+	1100	880	220
Total	**7,000**	**5,600**	**1,400**

Table SC.2. Breakdown of data across each class and split into training and test data sets.

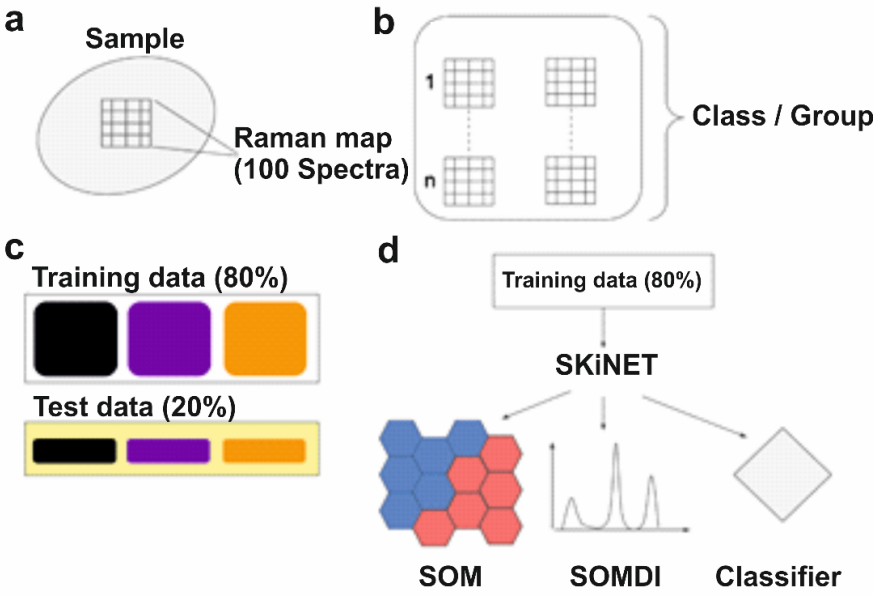

Figure SC.1. Illustration of data analysis workflow for saliva samples using SKiNET. Spectra measured from Raman maps **(a)** are grouped according to class or group studied. **(b)**. A 20% partition of the data is randomly selected and reserved as test data **(c)**. The remaining 80% is input into SKiNET, which directly provides dimensionality reduction (SOM), self-organising map discriminant index (SOMDI)

feature extraction and classification **(d)**. SKiNET is optimized on the training data using cross validation and adjusting the available parameters (number of neurons, initial learning rate and number of training steps) to maximize the classification accuracy on the training data. Finally, the optimized model is shown the previously unused test data and asked to classify each spectrum as either male, female or the various age sub-groups.

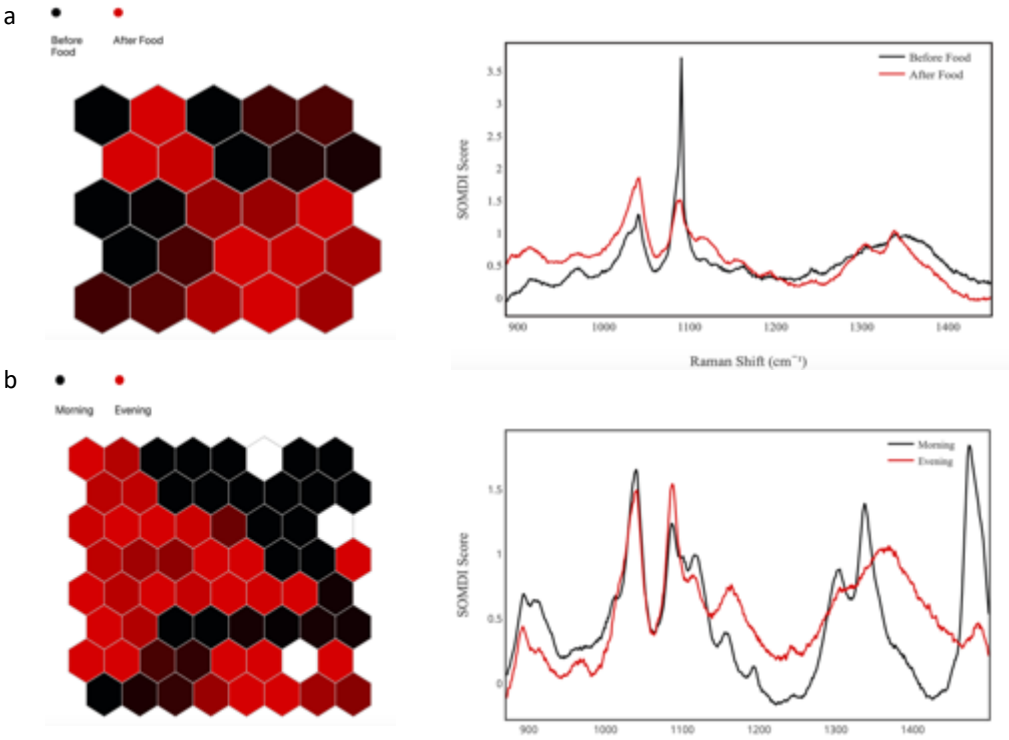

Figure SC.2. SOM (left) and SOMDI (right) of **(a)** saliva before and after food and **(b)** morning versus evening saliva, successfully classifying the saliva samples according to class with an accuracy of 90±0.2% and 77.5±1.1%, respectively.

The established process was further applied in the spectroscopic molecular-fingerprint profiling of saliva in the morning versus the evening and before and after eating food. Fresh saliva samples from 3 female participants aged 26±2 years were analysed using our recently developed neural network SKiNET algorithm. The derived weight vectors characteristic to each class are shown in **Fig.SC.2a-b.**

The most prominent spectral features of the saliva analysed before and after food are found at 912, 964, 1003, 1040, 1090, 1160, 1193, 1240, 1305 and 1340 cm^{-1}. Spectral differences of saliva pre- (black) and post- (red) eating are found to vary significantly. The accompanying SOMDI score indicates that in all but one region (amide III band) it

is the intensity changes within the saliva rather than the peak shift that contribute to the spectral differences. After eating there was an increased response at 912, 968, 1001 and 1040 cm^{-1} with an additional peak observed at 1305 cm^{-1}. Prior to eating increased responses were observed at 1090, 1240 and 1340 cm^{-1}. These spectral changes were largely attributed to proteins, lipids and amino acids.

The peak at 964 cm^{-1}, attributed to calcium-phosphate stretching band (cholesterol), is increased post intake of food. Although it is recognised that eating has only slight effects on three parts of the lipid profile, food does raise levels of triglyceride for several hours. Eating a high fat meal would increase these levels further [1]. Edwards *et al.* observed in their study that the overall rates of cholesterol synthesis in the liver during the regular circadian rhythm are due to the effects of eating [2]. The 1003 cm^{-1} peak is assigned to phenylalanine, with an increased response observed after food ingestion. A study by Biolo *et al.* looking at phenylalanine kinetics during meal ingestion, identified that phenylalanine levels of individuals increased during the administration of a mixed meal [3].

A further peak at 1160 cm^{-1} was assigned to lipids, 1040 cm^{-1} was glucose, with 1190 cm^{-1} assigned to carotene. As expected, glucose was identified with an increased intensity post food ingestion. As with any healthy individual blood glucose levels begin to rise whilst eating. Insulin then acts with it taking approximately 2 hours post food ingestion to return blood glucose levels to pre-meal status [4]. Carotene levels have also been shown to increase after ingestion of specific foods. A study by Micozzi *et al.* investigating the plasma carotenoid response to chronic intake of foods and β-carotene

supplements indicated that both α- and β- carotene increased in response to ingestion of foods such as carrots, broccoli and tomato juice [5]. Spectral changes observed at 1240, 1305 and 1340 cm^{-1} were attributed to changes in the amide III region. Prior to food there was a peak shift of 30 cm^{-1} when compared to post food intake.

Likewise, when analysing the Raman spectra of saliva in the morning (black) versus the evening (red), significant differences were observed. The most prominent spectral features are found at 892, 912, 964, 1040, 1090, 1116, 1160, 1194, 1240, 1300, 1337, 1372 and 1473 cm^{-1}. The SOMDI score also indicate it is again the intensity changes as opposed to Raman peak shift that is responsible for the observed spectral differences. Morning salivary spectra was distinct from evening spectra with increased responses at 892, 912, 964, 1040, 1116, 1300, 1337 and 1473 cm^{-1}. An additional peak at 1194 cm^{-1} was also identified. The evening spectra, however, has an increased response at 1090, 1160, 1240 and 1372 cm^{-1}. The majority of peaks were assigned to proteins, lipids and amino acids.

The morning Raman spectra indicated an additional peak at 1194 cm^{-1}. This peak is typical in the Raman spectra of cortisol. Cortisol, the stress hormone, has a powerful influence on sleeping and waking in the body. Sleeping and waking typically follow a circadian rhythm, with cortisol production also following a similar circadian rhythm. Levels of cortisol drop to their lowest at midnight with the cortisol peak an hour after waking [6]. In the majority of Raman saliva work the authors opted to avoid the cortisol peak in the morning and therefore, obtain saliva samples in the afternoon. Cortisol has a strong influence on the salivary spectra as evidenced in Fig. 2b, and with 15 to 18

smaller pulses of cortisol known to be released throughout the day and night, with some of the pulses corresponding to shifts in the sleep cycle, Raman analysis of cortisol has the potential to act as a tool to monitor the sleep cycle.

Peaks at 1116 and 1300 cm^{-1} were attributed to lipids, 964 cm^{-1} was assigned to calcium stretching band (cholesterol), 1040 cm^{-1} was glucose, and 1473 cm^{-1} was triglycerides. Versteeg *et al.* carried out a study monitoring the effects of morning light on glucose and triglyceride levels, they observed that ambient light affects plasma glucose and triglyceride levels. Both glucose and triglyceride peaks were observed at an increased intensity in the Raman spectral analysis of morning saliva, in agreement with the findings published by Versteeg *et al*. [7].

Peak at 1194 cm^{-1} was attributed to tyrosine. This peak was identified in the morning but not in the evening salivary spectra. A study by Fernstrom *et al.* indicated that as the protein content of the diet was increased throughout the day, the concentration of tryptophan, tyrosine and phenylalanine decreased [8]. After an evening meal protein levels are typically at their highest levels therefore tyrosine, tryptophan and phenylalanine levels would be at their lowest, thus no tyrosine peak would likely be observed as was evidenced in the Raman salivary spectra of evening saliva.

Further peak assignments were 1337 cm^{-1} as breathing modes of DNA, 1160 cm^{-1} as collagen and 1240 and 1372 cm^{-1} as amide III.

Here, Raman spectroscopy has been established as a useful tool in the molecular profiling of saliva samples. Fasting status and sampling time (morning or evening) have

been enabled from the differences observed in their spectral fingerprints. Classification of each molecular spectroscopic profile was enhanced using SKiNET algorithm, providing classification accuracies of 90% for time-of-day studies and 77.5% for fasting status classification. These preliminary results indicate the possibility of Raman spectroscopy to act as a useful physiological monitoring tool.

SC.2 References

[1] H. Ma, 'Cholesterol and Human Health.' *Nature and Science,* vol. 2, pp. 17-21, 2004.

[2] P.A. Edwards, H. Muroya, R.G. Gould, 'In vivo demonstration of the circadian rhythm of cholesterol biosynthesis inthe liver and intestine of the rat.' *J. Lipid Research,* vol. 13, pp. 396-401, 1972.

[3] G. Biolo, P. Tessari, S. Inchiostro, D. Bruttomesso, C. Fongher, L. Sabadin, M.G. Fratton, A. Valerio, A. Tiengo, 'Leucine and Phenylalanine kinetics during mixed meal ingestion: a multiple tracer approach.' *J. Physiology, Endocrinology and Metabolism,* vol. 262, pp. 455-463, 1992.

[4] Y. Yamada, S. Hosoya, S. Nishimura, T. Tanaka, Y. Kajimoto, A. Nishimura and O. Kajimoto, 'Effect of bread containing resistant starch on postprandial blood glucose levels in humans.' *Bioscience, Biotechnology and Biochemistry,* vol. 69, pp. 559-566, 2005.

[5] M.S. Micozzi, E.D. Brown, B.K. Edwards, J.G. Bieri, P.R. Taylor, F. Khachik, G.R. Beecher, J.C. Smith Jr, 'Plasma carotenoid response to chronic intake of selected foods and *β*-carotene supplements in men.' *J. Clinical Nutrition,* vol. 55, pp. 1120-1125, 1992.

[6] B. Bush, T. Hudson, 'The role of cortisol in sleep.' *J Natural Medicine,* vol. 2, 2010.

[7] R.I. Versteeg, D.J. Stenvers, D. Visintainer, A. Linnenbank, M.W. Tanck, G. Zwanenburg, A.K. Smilde, E. Fliers, A. Kalsbeek, M.J. Serlie, S.E. la Fleur, P.H. Bisschop, 'Acute effects of morning light on plasma glucose and triglycerides in healthy men and men with type 2 diabetes.' *J. Biological Rhythms,* vol. 32, pp. 130-142, 2017.

[8] J.D. Fernstrom, R.J. Wurtman, B. Hammarstrom-Wiklund, W.M. Rand, H.N. Munro, C.S. Davidson, 'Diurnal variations in plasma concentrations of tryptophan, tyrosine, and other neutral amino acids: effect of dietary protein intake.' *J. Clinical Nutrition,* vol. 32, pp. 1912-1922, 1979.

APPENDIX C

SUPPORTING INFORMATION FOR CHAPTER 5

Raman Scattered Spectroscopic Molecular Fingerprinting of Biomarkers for Inflammatory Bowel Disease

Emma Buchan[1], Jonathan James Stanley Rickard[2] and *Pola Goldberg Oppenheimer[1,3,*]*

Raman provides unique biomolecular spectral fingerprints of target analytes with rapid analytical response, enabling non-destructive, label-free, quantitative analysis of composition and structure with an inherently straightforward detection and no complex sample preparation, thus rendering itself as a powerful technique, which can be used to *in-situ* measure various biomarkers, yielding quantitative information about their concentration, highly-useful for monitoring disease progression and response to treatments. The availability of portable instruments makes it particularly attractive for point-of-care detection.

SD.1. Potential IBD Indicative Biomarkers Under Investigation

Collection of biofluids, *e.g.,* blood / urine, is a standard procedure in routine clinical practice and hence is ideal for a point-of-care or bedside setting. Biofluid derived biomarkers have many useful applications in healthcare including disease prevention and detection, determination of individual's risk and disease monitoring [1-2]. Currently, however, there are no validated blood biomarkers to accurately and reliably predict IBD. The complexity of interactions between host factors and the dynamic fluctuation of the gut microbiota during IBD hinders the identification of consistent changes in microbial composition, and thus the universal biomarkers for disease prediction [3]. Hence a large proportion of patients with non-specific abdominal pain often undergo unnecessary painful endoscopy or colonoscopy to rule out IBD [4]. Whilst effective, these are invasive and painful, and thus less suitable for routine use for timely and rapid IBD diagnosis. Not only are these procedures invasive but also a confirmatory IBD diagnosis requires further clinical examinations, imaging, and pathological investigations, which can be

often inconclusive. A pool of diagnostic and/or prognostic biomarkers of IBD would thus, considerably reduce or eliminate the need for invasive or nonspecific procedures and greatly improve early-stage diagnosis, management, and therapeutic interventions for IBDs. The ability to determine the type, severity and patient response to therapeutics has long been a priority of clinical researchers.

Although demonstrating correlation with IBD severity and holding potential diagnostic, prognostic and stratification value, biomarkers currently under investigation, are undetectable unless analysis is carried out in specialised laboratories.

Potential IBD indicative biomarkers currently under investigation include C-reactive protein (CRP), faecal calprotectin, anti-*Saccharomyces cerevisiae* antibodies (ASCA), anti-neutrophil cytoplasmic antibodies (ANCA) and serum amyloid A [5-9]. Cytokines, proteins produced by immune cells, are also known to play an active role in inflammation, and thus are additional candidate biomarkers in IBD [10]. Idiopathic IBDs typically arise in clinically immunocompetent individuals with cytokine-driven inflammation of the gastrointestinal tract giving rise to the distinctive signs and symptoms in individuals [11]. CD is most often associated with increased levels of IL-12/IL-23 and IFN-γ/IL-17, whereas UC is associated with excess IL-13 production [12]. Research studying the pathogenesis of IBD suggests that IBD arises due to a dysfunctional interaction between the mucosal immune system and the bacterial microflora of the gastrointestinal (GI) tract, and thus cytokines are well known to play a key-role in controlling intestinal inflammation and the associated diseases [13]. Cytokine testing therefore has the potential to support the IBD diagnosis due to its low cost and lack of invasiveness, compared to routine IBD diagnostics.

While demonstrating correlation with IBD severity and holding potential diagnostic, prognostic and stratification value, these potential biomarkers are undetectable unless analysis is carried out in specialised laboratories, using enzyme-linked immunosorbent assay (ELISA), chemiluminescent assay, high performance liquid chromatography (HPLC) or mass spectrometry. Although reliable, these techniques have time-to-results in the range of days, require complex equipment and trained personnel and do not meet the need for detection of ultra-low levels of target analytes in complex biological samples using rapid and direct readouts. ELISAs are relatively cost-effective but, despite advancements, suffer from poor stability, not rapid enough, require expensive labels for each biomarker and result in unacceptable rates of false negatives and false positives [14].

SD.2 References

[1]. Soubières AA. Emerging role of novel biomarkers in the diagnosis of inflammatory bowel disease. World J Gastrointest Pharmacol Ther. 2016, 7(1):41.

[2]. Wang A, Wang C, Tu M, Wong D. Oral Biofluid Biomarker Research: Current Status and Emerging Frontiers. Diagnostics. 2016 17, 6, 45.

[3]. Khan I, Ullah N, Zha L, Bai Y, Khan A, Zhao T, et al. Alteration of Gut Microbiota in Inflammatory Bowel Disease (IBD): Cause or Consequence? IBD Treatment Targeting the Gut Microbiome. Pathogens. 2019 13, 8(3):126.

[4]. Tefas C, Mărginean R, Toma V, Petrushev B, Fischer P, Tanțău M, et al. Surface-enhanced Raman scattering for the diagnosis of ulcerative colitis: will it change the rules of the game? Anal Bioanal Chem. 2021 7, 413(3):827–38.

[5]. Wakai M, Hayashi R, Tanaka S, Naito T, Kumada J, Nomura M, et al. Serum amyloid A is a better predictive biomarker of mucosal healing than C-reactive protein in ulcerative colitis in clinical remission. BMC Gastroenterol. 2020 3.20(1):85.

[6]. ROOZENDAAL C, KALLENBERG CGM. Are anti-neutrophil cytoplasmic antibodies (ANCA) clinically useful in inflammatory bowel disease (IBD)? Clin Exp Immunol. 2001, 24, 116(2):206-13.

[7]. Israeli E. Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease. Gut. 2005 1;54(9):1232.

[8]. Sherwood R, Walsham N. Fecal calprotectin in inflammatory bowel disease. Clin Exp Gastroenterol. 2016 Jan;21.

[9]. Vermeire S, van Assche G, Rutgeerts P. C-Reactive Protein as a Marker for Inflammatory Bowel Disease. Inflamm Bowel Dis. 2004 Sep;10(5):661–5.

[10]. Rogler G, Andus T. Cytokines in Inflammatory Bowel Disease. World J Surg. 1998 1, 22(4):382-9.

[11]. Strober W, Fuss I, Mannon P. The fundamental basis of inflammatory bowel disease. Journal of Clinical Investigation. 2007 1, 117, 514.

[12]. Guan Q, Zhang J. Recent Advances: The Imbalance of Cytokines in the Pathogenesis of Inflammatory Bowel Disease. Mediators Inflamm. 2017, 2017,1-8.

[13]. Fava F. Intestinal microbiota in inflammatory bowel disease: Friend of foe? World J Gastroenterol. 2011, 17(5):557.

SD.2. Self-Optimising Kohonen Index Network (SKiNET) Algorithm

The acquired data is classified using our new artificial neural network algorithm, self-optimising Kohonen index network (SKiNET) as a decision support tool, based on self-organising map (SOM) with a classification *via* the self-organising map discriminant index (SOMDI).

Through inspection of key differences between neuron weights and class weight vectors, the algorithm enables identification of the key spectral changes. These allow the identification of the types of data a given neuron activates, which are then used to inspect the weights across all neurons and extract prominent features belonging to each class by finding the weights that contribute most to a particular class. The peaks in SOMDI subsequently correspond to cm^{-1} and modes that contribute most to the clustering observed in the SOM. Training parameters used for the SOM include grid size, learning rate and optimal number of epochs and the separation of classes reveals the characteristic differences due to the classification of certain neurons. This enables a clear basis for differentiation *via* the characteristic weight vectors to be derived in SOMDI.

Inspired by the visual cortex in the brain, SOMs are trained for the neighbouring neurons to activate according to similar inputs, in this case Raman spectra. Each neuron has a weight vector with length equal to the number of variables in a spectrum. Through exposing the network to training samples over a number of iterations, the weights are gradually adjusted to be similar to the input data, so that each neuron only activates on a given spectral signature. The result is a projection of hyperspectral data into 2D space, that can be shown as visible clustering according to sample / disease type and state. SKiNET inherently employs SOMDI, which appends a set of label vectors to each neuron and allows us to study the most prominent

features that cause the activation of a particular neuron to a class label. Subsequently, a supervised learning step is introduced to optimise the network, and the class label associated with each neuron used to quickly identify new data presented to the SOM, allowing for diagnostics. Raw component spectra from chosen candidate molecules are fitted to SOMDI for a particular state, constituting a physically realistic fit, as our Raman spectra represent a mixed state of positive contributions from constituent components.

Our AI is based on SOMs with SKiNET as a framework for multivariate analysis that simultaneously provides (i) dimensionality reduction, (ii) feature extraction and (iii) multiclass classification (**Fig.SD.1**), where SKiNET performs visual separation to identify the underlying chemical differences between classes, providing accurate classification for simultaneously rich-information and high-classification specificity. SOMs provide visually intuitive 2D-clustering (according to disease / healthy *etc*. state) of high-dimensional data Raman spectra, that are otherwise difficult to interpret for large sample and measurement numbers. SKiNET incorporates supervised learning to additionally provide accurate classification, which could then be used to make diagnostic predictions. Whilst previous SOMs are usually an unsupervised method, our optimised SKiNET incorporates supervised learning to additionally provide accurate classification, which could then be used to make diagnostic predictions. Finally, a form of feature extraction (SOMDI) allows to understand which spectral features (=biochemical changes) are responsible for the clustering seen in SOMs.

When Raman spectroscopy combined with SKiNET is applied to investigate whether the identified biomarkers reflect the IBD, it distinguishes disease from a healthy control group, showing this to be because of similar chemical changes detected.

This algorithm enables a clear separation of the data from the different tissue or biofluid classes arranged as SOM, trained on spectra from these classes. Neurons (hexagons) are coloured according to the modal class they activate, from the training set of Raman spectra. Neurons that have no majority class or activate none of the training data are shown in white. Coloured circles within each neuron represent spectra from the training data that have been activated for that neuron. To aid visualisation, circles are forced to not overlap in space using the D3force library, providing an alternative mechanism to display sample frequency and class overlap for each neuron. For each class, there is a clearly defined block of neurons, with many of these activating only a single tissue type. An approximately even distribution in the number of neurons required to identify each class is observed. The SOMDI provides a representation of weights associated with neurons that identify a particular class. A higher SOMDI intensity indicates a greater importance of particular inverse centimetres along the axis of a spectrum. This, despite the level of overlap or noise in the original data, enables well defined peaks to be resolved, which are either more prominent or unique to each class. Automated classification of Raman spectra and assignment to a particular biomarker, tissue type or disease state is perhaps the most important step for the translation of Raman based diagnostic techniques to real world, clinical applications.

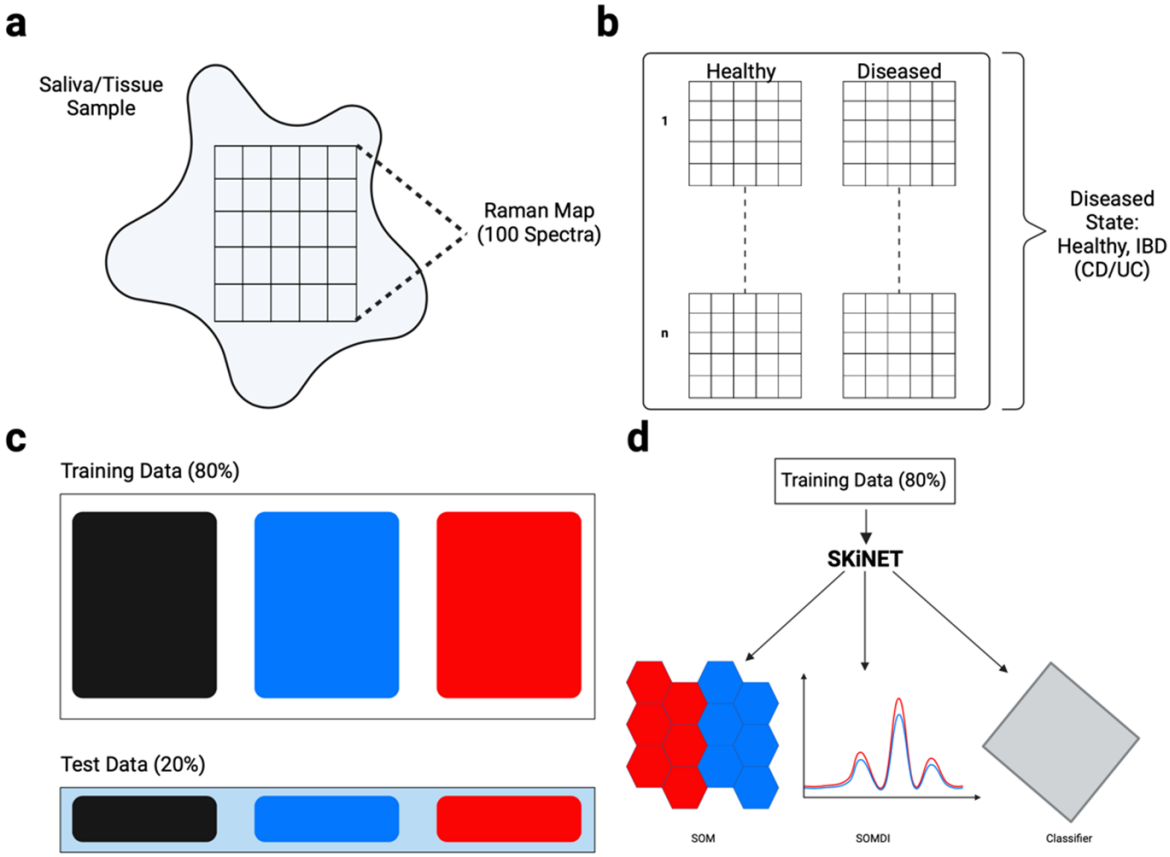

Figure SD.1. Illustration of data analysis pipeline using SKiNET. Spectra measured from Raman **(a)** of saliva or tissue are grouped according to class **(b)**. A 20% partition of the data is randomly selected and reserved as test data **(c)**. The remaining 80% is input into SKiNET, which directly provides dimensionality reduction (SOM), self-organising map discriminant index (SOMDI) feature extraction and classification **(d)**. SKiNET is optimised on the training data using cross validation and adjusting the available parameters (number of neurons, initial learning rate and number of training steps) to maximise the classification accuracy on the training data. Finally, the optimised model is shown the previously unused test data and asked to classify each spectrum as either IBD, healthy or IBD biomarker.

Table SD.1. Characteristic assignments for the identified dominant Raman peaks. [1-5]

Raman Shift / cm^{-1}	Assignment
759	Tryptophan ring breathing mode; Proteins
828	DNA phosphate groups
875	C-C symmetric stretching; Proteins and lipids
877	Choline stretching group; Phosphatidylcholine
936	C-C stretch amino acids; Protein Backbone
1003	Protein marker, Phenylalanine ring breathing mode, ρ(C-C); Phenylalanine
1110	C-C vibration of the gauche-bonded chain
1051	C-O and C-N stretch
1080	(ν(C-C) of lipids); PO_2 group in DNA and lipids
1128	Trans C-C stretching; backbone of lipid, protein, and carbohydrate
1201-1207	CH_2 wagging vibrations (glycine, proline, tyrosine, and phenylalanine
1230-1285	Amide III; Protein
1270	C-H deformation; Lipids
1305	CH_2 twisting; Phospholipids, adenine, myoglobin
1337	CH_2/CH_3 wagging and twisting (proteins, nucleic acids), nucleic acid bases (υ(C-H))
1340	C-H deformation; Protein
1368	ω(CH_2), δ(CH), υ4-υ(Pyr ¼ ring); tryptophan, guanine, thymine, myoglobin (Haem core)
1445	Shoulder (δ(CH_2) deformation of proteins and lipids
1447	CH_2 scissoring (lipid and protein)
1518	υ(C=C)
1620-1680	Amide I band
1622	C=C stretching mode of tyrosine and tryptophan
1643	C=C stretching modes of lipids and the protein Amide I
1656	C=C stretching in lipids, C=O stetch in proteins
1658	(υ(C=O))

[1]. Tefas C, Mărginean R, Toma V, Petrushev B, Fischer P, Tanțău M, et al. Surface-enhanced Raman scattering for the diagnosis of ulcerative colitis: will it change the rules of the game? Anal Bioanal Chem. 2021, 7;413(3):827-38.

[2]. Morasso C, Truffi M, Vanna R, Albasini S, Mazzucchelli S, Colombo F, et al. Raman Analysis Reveals Biochemical Differences in Plasma of Crohn's Disease Patients. J Crohns Colitis. 2020, 7;14(11):1572-80.

[3]. Pence IJ, Beaulieu DB, Horst SN, Bi X, Herline AJ, Schwartz DA, et al. Clinical characterization of in vivo inflammatory bowel disease with Raman spectroscopy. Biomed Opt Express. 2017, 1, 8(2):524.

[4]. Smith SCL, Banbury C et al. Raman spectroscopy accurately differentiates mucosal healing from non-healing and biochemical changes following biological therapy in inflammatory bowel disease. PLoS One. 2021;16, 1-16.

[5]. Pence I, Mahadevan-Jansen A. Clinical instrumentation and applications of Raman spectroscopy. Chem Soc Rev. 2016, 45, 1958.

Table SD.2. Normalised Raman intensity of main statistically significant and dominant spectral peaks identified.

Marker	**Raman Shift / cm^{-1}**				
	936	**1003**	**1340**	**1445**	**1656**
Healthy Saliva	0.29	0.77	0.56	0.81	0.46
IBD Saliva	0.28	0.84	0.51	0.78	0.45
Healthy Biopsy	0.25	0.63	0.60	0.75	0.51
IBD Biopsy	0.19	0.73	0.45	0.99	0.69
IL-4	0.37	0.99	0.49	0.63	0.45
IL-8	0.28	1.00	0.60	0.59	0.42
IL-12	0.17	0.91	0.39	0.97	0.43
IL-17	0.19	0.99	0.50	0.84	0.58

Table SD.3. Classification performance of tissue biopsy compared to healthy controls including sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV).

Comparison	Sensitivity	Specificity	PPV	NPV	Accuracy
Healthy v. IBD	83.9	84.6	87.6	88.2	82.5
Healthy v. UC	94.7	90.4	90.0	95.0	90.0
Healthy v. CD	94.4	86.3	85	95	90.2

Table SD.4. Classification performance of saliva compared to healthy controls.

Comparison	Sensitivity (%)	Specificity (%)	PPV	NPV	Accuracy
Healthy v. IBD	87.5	89	89	87.3	88.3
Healthy v. UC	91.6	88.1	88.5	88.5	80.1
Healthy v. CD	87.3	87.7	86.1	88.9	80.3

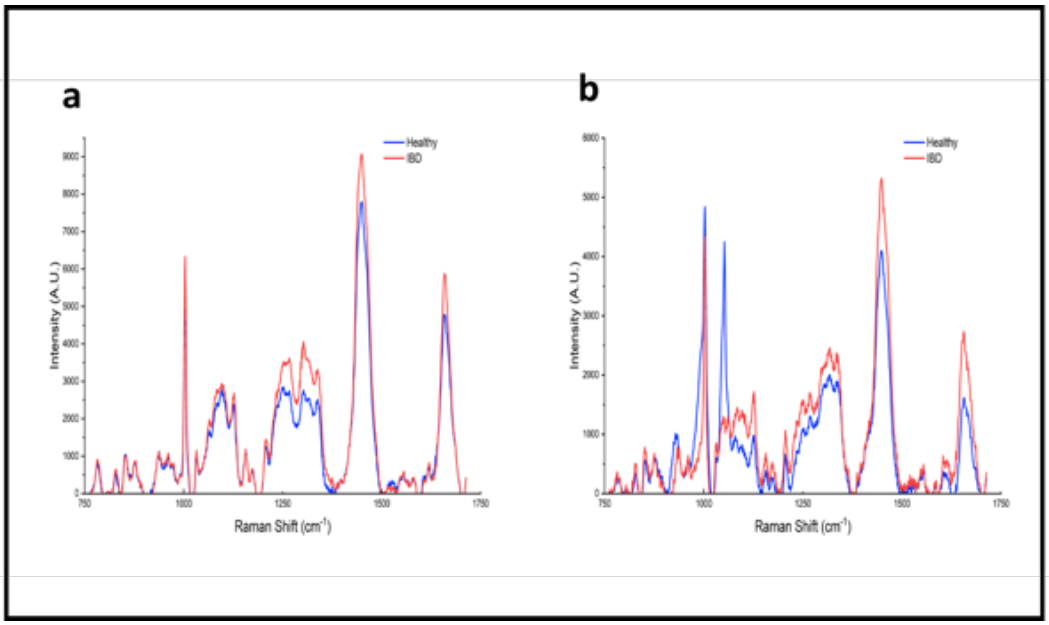

Figure SD.2. Raman spectral data of non-normalised molecular fingerprints of **(a)** saliva and **(b)** tissue biopsy (blue=healthy, red=IBD). The identified Raman spectral fingerprints obtained from non-normalised data align with the observed SNV normalised data, with the only differences observed in the scaling of the data.

SD.3. Spectroscopic Data Analyses and Interpretation Details – Extended Discussion

A trio of characteristic bands at 821, 851, 924cm^{-1} were identified for PKM2, emerging as a significant mediator of inflammatory processes with serum levels 6-fold higher in IBD *versus* healthy patients, [1] posing PKM2 as a putative IBD-biomarker (**Fig.3.2b/SD.3**). Prior research by Czub *et* al. identified increased fecal PKM2 levels 100% of active CD patients and 94.3% of active UC patients with associated enzyme reactivity higher in all IBD patients than those healthy controls. Here we identified PKM2 at increased levels in both saliva and tissue at each of the three characteristic Raman bands, thus indicating its potential as an identifier of IBD in both tissue and saliva as opposed to its current use in faecal samples. In addition, whilst CRP

exhibited increased intensities at 841, 912, 1076 and 1126cm^{-1} (p***<0.0001), this marker alone was found to be non-specific due to the elevation in systemic inflammatory disease other than those identified within the intestinal tract, for example in cancers of the body or infections, CRP levels can also be significantly elevated. However, when combined with other potential markers of IBD, such as PKM2, IL-4, IL-8, IL-12 and IL-23 CRP provides invaluable insight into the level of inflammation in the gut. Increased levels of CRP for example, can help differentiate mucosal active disease from quiescent IBD as well as acting as a predictor as to the need for colectomy through the reflection of severe ongoing and uncontrolled inflammation of the gut [2].

Notably, the peak at 936cm^{-1} exhibits a decreased intensity in the IL-4 compared with IL-8, IL-12, and IL-17 due to the C-C stretch of the amino acids in the protein backbone, where the vibrational frequency depends on the local chemical environment including hydrogen bonds and steric hindrance and represents the local conformation of the peptide bond predominantly influenced by the neighbouring amino acid residues. The 1340cm^{-1} is attributed to C-H deformation of proteins backbone/side chains, providing information on the presence of hydrogen bonds, electrostatic interactions and intermolecular forces. Since the position and intensity of this band can be influenced by presence of ligands, cofactors, or changes in the native environment, analysing these provides important underpinning insights on the structure and dynamics of the proteins. Herein, the rise in the C-H deformation band indicates an increase in either number or strength of C-H bonds in each of the interleukin molecules, and a change in the protein secondary structure due to the diseased state of the patient.

The peak at 1656cm^{-1} is associated with the amide-I band of the C=O stretching of backbone conformation [3]. Given that amide-I is the most intense absorption band in proteins and its exact position is determined by the backbone confirmation and the pattern of hydrogen bonding, its increase is indicative of these changes.

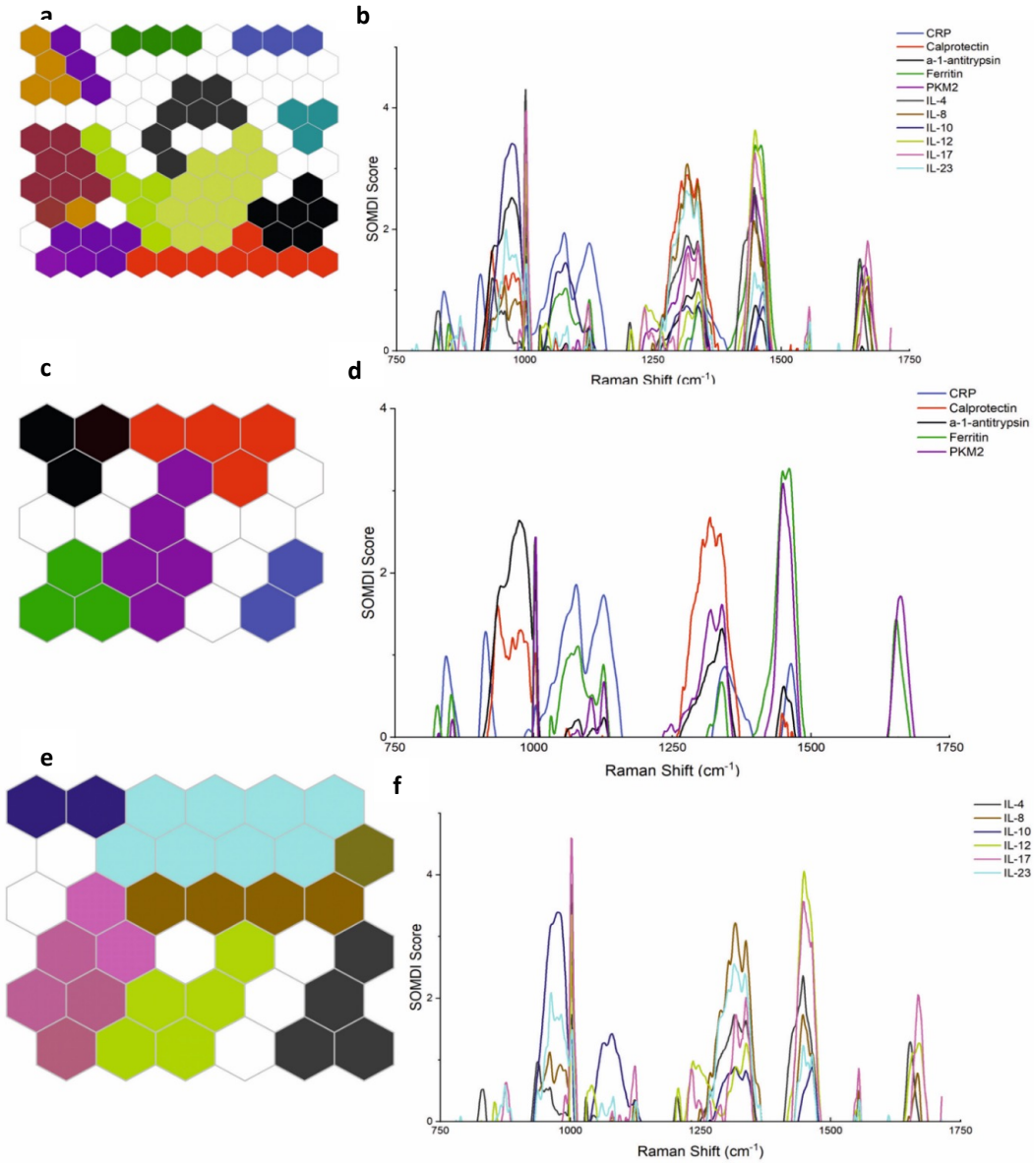

Figure SD.3. Representative SOM clustering and classification according to biomarker type of **(a)** all the studied biomarkers, **(c)** CRP, calprotectin, α-1-antitrypsin, ferritin and PKM2 and

(e) cytokine; IL-4, IL-8, IL-10, IL-12, IL-17 and IL-23 with the corresponding SOMDI extracted features from SOM **(b, d and f)** classifying biomarker subtype with an accuracy of 98.2%.

We have successfully detected and identified several Raman key-bands, which were characteristic in patients with IBD including the 936cm^{-1} (C-C stretching of amino acids, protein backbone), 1003cm^{-1} (Phenylalanine), 1340cm^{-1} (C-H deformation, Protein), 1445cm^{-1} (Shoulder (δ(CH2) deformation of proteins and lipids and 1656cm^{-1} (C=O stretch in proteins, Amide I). In IBD disease state, we also found that the intensity of the spectral peaks at 936 and 1340cm^{-1} to be considerably downregulated, whilst the bands at 1003, 1445 and 1656cm^{-1} were upregulated. These results are consistent with study Morasso *et* al. [4] where the authors identified similar assignments to these spectral changes via comparison of dried plasma *ex-vivo* samples using Raman spectroscopy and analysed using principal component analysis followed by linear discriminant analysis.

The fact that the same set of peaks are visible in both the healthy and diseased sample types would thus indicate that their biochemical composition is similar, which is consistent with our general understanding of inflammation. Inflammation does not introduce new metabolites into the system but instead, leads to an overproduction or overuse of the existing metabolites. Therefore, it is expected to see the same molecular changes as identified *via* intensity changes of the peaks in the same region in both tissue types as was evidenced in our study, supporting that these are due to the inflammatory response, where the variation in peak intensities between the two types are thus due to the detected differences in the concentration of the biomolecules.

Furthermore, the significant increase in intensity at 1003cm^{-1} is an additional indication of inflammation due to IBD, since in previous studies, this peak has been associated with higher levels of phenylalanine, due to the immune activation and inflammation. Smith *et* al. observed a significant reduction in intensity due to mucosal healing and Addis et al. recognised a similar effect, following biological therapy. [5] An additional finding relating to the peak changes at 1656cm^{-1} (Amide I) could indicate conformational changes induced by factors evolving from the inflammatory state. Although, it is possible that factors stemming from age, dehydration or radiological conditions, could also be responsible for the changes detected, these do not affect the data reported in our study as the recruited subjects were of a highly similar age, with no known health conditions thus predominantly, the biochemical changes detected are indicative of the inflammatory response associated with IBD.

Our study subsequently, has proceeded to identify which biomarkers associated with IBD could be responsible for the intensity changes found between healthy and IBD patient samples within these regions. Changes in response to IBD, have shown downregulation at 936cm^{-1}, (representative of C-C stretching associated with amino acids such as glutamine and arginine which influence the progression of IBD in individuals) with the main statically significant differences arising from amino acid composition ($p^{***}<0.001$). Our results of decreased intensity at 936cm^{-1} in IBD patients is further consistent with previous studies, where changes in pro-inflammatory cytokines, which mediate interaction between immune cells and non-immune cells, have been shown [6-9] to contribute to the inflammatory status of the intestine. In ulcerative colitis, the T-helper 2 response consists of IL-4 and IL-13. IL-4 is an

important contra inflammatory cytokine, which limits monocyte and macrophage activation. IL-4-mediated downregulation of activation has been shown by Ruckert *et* al. to be impaired in IBD [10]. Hoving *et* al. have also indicated that mice deficient in IL-4, IL-13 and IL-4 receptor-alpha on all cells develop an exacerbated IBD phenotype. [11] A further study by Xiong *et* al. identified the effects of IL-4 and IL-10 gene therapy on TNBS-induced murine colitis, shown to significantly inhibiting the TNBS-induced colon tissue damage and disease activity index as well as a marked block in expression of IFN-γ and TNF-α, highlighting the significance of monitoring IL-4 levels in IBD patients as well as the potential of IL-4 as a first line of defence in the treatment of IBD [9].

Furthermore, we have detected an upregulation of the band at 1003cm^{-1} attributed to the phenylalanine changes indicative of an increased inflammatory response. In IBD, neutrophils are known to be important cellular mediators with the IL-8 being a powerful neutrophil chemoattractant found in increased quantities in the mucosa. Grimm *et* al., by isolating macrophages and monocytes from intestinal resections and detecting IL-8 by *in-situ* hybridisation, found that in the inflamed bowel, IL-8 was detected in both macrophages and neutrophils however, not detected in uninflamed mucosa and the IL-8 expression was significantly more common by macrophages from IBD affected mucosa than in their healthy counterparts [12]. These results support that the detected increased levels of IL-8 with marked intensity increase at 1003cm^{-1} combined with an increase in the Amide I band at 1656cm^{-1} are indicative of protein confirmational changes, suggesting that IL-8 can act as a significant indicator of disease. Additionally, previous study by Billiet *et* al. identifying the correlation of primary IFX therapy

response with lower serum IL-8 levels, further suggests the ability of cytokines to predict the response to different therapies used in patients with IBD [6]. Additionally, the detected increase of statistically significant ($p^{***}<0.0001$) bands at 1003 and 1656cm^{-1} combined with the decrease at 1445cm^{-1} are indicative of the IL-12, playing a key-role in the activation and regulation of multiple cytotoxic immune cells including macrophages, natural killer cells and T cells, [6,13] produced predominantly by macrophages in response to bacteria and their products. IL-12 is rarely detected in normal intestinal mucosa and therefore, acts as an appealing potential marker of IBD. Recently, IL-12, in combination with other interleukins such as IL-23, has been emerging as a target molecule for treatment in IBD patients, for example, the use of Ustekinumab, an anti-IL-12/23p40 antibody has been approved for CD. Almradi *et* al. have also indicated that in early stages of IBD, IL-12 is the dominant p40-containing cytokine, driving inflammation in response to intestinal barrier disruption. [14] Similarly, IL-17 is known to exert a strong proinflammatory response. Its secretion has been reported to be limited to T lymphocytes with major indications of it being a strong mediator of inflammatory response in various tissues [15] as well as enhancing the proinflammatory response induced by IL-1β and TNF-α. Previous studies have shown that IL-17 expression has been detected in inflamed mucosa of active UC and CD patients *via* immunohistochemical techniques and serum ELISA. [16] Levels of IL-17 have been also previously found to increase significantly from 371.5 pg/mL to 1365.1 pg/mL in patients with IBD [17].

Building upon the above studies correlating with our results, *via* the ability to identify differing intensities within IBD and healthy individuals, we firstly highlight the significance of IL-4 in the pathogenesis of disease given its widely accepted role in IBD

as well as the diagnostic value of IL-8 as a biomarker with its potential to monitor mucosal healing and potential monitoring of patients in responses to therapeutics. The detected changes further highlight the importance of IL-12 in the pathogenesis and progression of IBD and of the IL-17 ability to act as a multiplexed IBD biomarker. Overall, real-time, rapid *in-vivo* spectroscopic measurements in IBD patients will enable establishing insights into biological pathways underlying the associated pathophysiology and could conceivably allow tracking the passage and dosage of current and emerging pharmacological therapeutics.

[1]. Czub E, Herzig KH, Szaflarska-Popawska A, Kiehne K, Socha P, Woś H, et al. Fecal pyruvate kinase: a potential new marker for intestinal inflammation in children with inflammatory bowel disease**.** Scand J Gastroenterology**.** 2007 10, 42(10): 1147-50.

[2] Vermeire S, Assche GA, Rutgeerts P, C-reactive protein as a marker for inflammatory bowel diease. Inflammatory Bowel Disease. 2004 9, 10(5):661-5.

[3]. Tefas C, Tanțău M. Clinical applications of raman spectroscopy in inflammatory bowel diseases. A review. Journal of Gastrointestinal and Liver Diseases. 2018, 27(4):433-8.

[4] Marasso C, truffe M, Vanna R, et al. Raman analysis reveals biochemical differences in plasma of Crohn's disease patients. Journal of Crohn's and Colitis, 2020, 1572-1580.

[5] Addis J, Mohammed N, Rotimi O, et al. Raman spectroscopy of endoscopic colonic biopsies from patients with ulcerative colitis to identify mucosal inflammation and healing. Biomedical Optics Express, 2016, 7(5).

[6]. Billiet T, Cleynen I, Ballet V, Claes K, Princen F, Singh S, et al. Evolution of cytokines and inflammatory biomarkers during infliximab induction therapy and the impact of inflammatory burden on primary response in patients with Crohn's disease. Scand J Gastroenterol. 2017 3, 52(10):1086-92.

[7] West G, Matsuura T, Levine A, et al. Interleukin 4 in inflammatory bowel disease and mucosal immune reactivity, 1996, 110(6):1683-1695.

[8] Jansson J, Willing B, Lucio M, et al. Metabolomics reveals metabolomic biomarkers of Crohn's Disease. PLoS One, 2009, 4(7):e6386.

[9] Xiong J, Lin Y, Bi L, et al. Effects of Interleukin-4 pr Interleukin-10 gene therapy on trinitrobenzenesulfonic acid-induced murine colitis, BMC Gastroenterology, 2013, 13 (165).

[10] Rückert Y, Schnidler U, Heinig T, et al. IL-4 Signalling mechanisms in inflammatory bowel disease mononuclear phagocytes, Inflamm Bowel Dis, 1996, 2(4):244-52.

[11] Hoving J, Keeton R, Höft M, et al. IL-4 receptor-alpha signalling intestinal epithelail cells, smooth muscle cells, and macrophages plays a redundant role in Oxazolone colitis. Mediators Inflamm, 2020, 32410852.

[12] Grimm M, Elsbury S, Pavli P, Doe W. Interleukin 8: cells of origin in inflammatory bowel disease, Gut, 1996, 38(1), pp.90-98.

[13] Greving C, Towne J. A role for IL-12 in IBD after all? Immunity, 2019, 51(2), pp. 209-211.

[14] Almradi A, Hanzel J, Sedano R, et al. Clinicla trials of IL-12/IL-23 inhibitors in inflammatory bowel disease. Clinical Trials BioDrugs, 2020, 34(6):713-721.

[15] Fujino S, Andoh A, Bamba S, et al. Increased expression of interleukin 17 in inflammatory bowel disease. Gut, 2003, 52(1):65-70.

[16] Lucaciu L, Ilies M, Vesa S, et al. Serum interleukin (IL)-23 and IL-17 profile in inflammatory bowel disease (IBD) patients could differentiate between severe and non-severe disease. J Pers Med, 2021, 11(11): 1130.

[17] Fujino S. Increased expression of interleukin 17 in inflammatory bowel disease. Gut. 2003, 1, 52(1):65-70.

SD.5. Materials and Methods

Sample Collection and Preparation. Saliva was collected from 101 participants (51 IBD and 50 healthy) at Queen Elizabeth Hospital Birmingham, UK. All healthy volunteers had no previous health issues or known conditions. 5 mL of unstimulated saliva was collected from each participant in a 50 mL Falcon tube (Thermo Fisher) *via* the passive drool method. Subsequently, 5 µL of raw saliva was pipetted onto an aluminium slide and air dried in an airtight container for 30 minutes. Colon tissue biopsies were collected from 44 participants (32 IBD and 12 healthy) (Ethics Refs. ERN_22-0290 and 19/SW/0010) into specimen containers (Thermo Fisher) containing 2 mL PBS and analysed immediately. All candidate markers were purchased directly from Sigma-Aldrich and Miltenyi Biotec Cologne and subsequently diluted in PBS (pH 8) to a working concentration of 1µg/mL. 5 µL of each was then pipetted onto an aluminium foil covered glass slide and dried in an airtight container for 30 minutes.

Raman Spectroscopy. Raman spectra were acquired using a Renishaw *InVia* Qontor confocal Raman microscope equipped with a microscope Leica DMLM and 785 nm laser (Renishaw PLC) with a spectral resolution of 0.3cm^{-1}. Laser light was focussed using a x50 objective lens with a laser power of 5 mW. Spectral maps were acquired over an area of 50x50 μm^2 in the fingerprint region of 700-1700 cm^{-1} with a 5 µm step size, 10 accumulations and 1s exposure time per spectrum. Overall, 100 spectra per sample were collected and used for data processing and analyses.

Data Acquisition and Analyses. The data was collected using WiRE 5.1 and further employed for the polynomial background subtraction and the removal of cosmic rays. Normalisation was done using the standard normal variate (SNV) (Python 3.7), which was used

to derive the multi-chemical barcoding of the data. SNV is applied to each data point in the data set, whereby, the mean of all data points within the given dataset is subtracted from each individual data point thus centring the data around zero. Each centred data point is then divided by the standard deviation, scaling the data. Savitzky-Golay filter was applied to calculate the second derivative of each spectrum. Smoothing window was set to 21 with a polynomial order of 2. Maximum peak heights with absolute values over 40% were assigned a value of 1 with values below 40% assigned a value of 0. Values were overlaid on the averaged spectra with reference to the main peaks identified for the barcode generation. Statistical analysis was performed using a one-tailed students *t*-test, peaks identified as having a $p<0.001$ were determined as statistically significant.

Data Processing and Multivariate Analysis. Multi-variate analysis was performed using the self-optimising Kohonen index network based on self-organising maps (SOMs) with the accompanying Raman Toolkit web interface to build SOM models using training data and perform predictions against test data. SKiNET models were optimised by performing 10-fold cross validation on the training data, and tuning the number of neurons, initial learning rate and number of training steps. The final model used a 10x10 grid of neurons, 38,200 training steps (4 epochs of the data), with an initial learning rate of 0.1. The optimised model was subsequently used to classify the previously unused test data. Classification using the test data were repeated ten times from separate SOM initialisations. To achieve higher accuracy the SOM size, learning rate and number of epochs was empirically tested, with classification accuracy determined using a 10-fold cross validation. Stability of the model was further verified by running a repeat initialisation of the classification four times.

APPENDIX D

SUPPORTING INFORMATION FOR CHAPTER 7

~ Supplementary Information ~

Advanced Integrated Multipurpose Spectroscopic Lab-on-a-chip for the Timely Detection of Extracellular Vesicles as Key-Markers of Disease

Emma Buchan[1], Jonathan James Stanley Rickards[2], Mark Thomas[3], Pola Goldberg Oppenheimer[1,4]*

[1]School of Engineering, College of Engineering and Physical Science, University of Birmingham, Birmingham, B15 2TT, UK
[2]Department of Physics, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge, CB3 0HE, UK
[3]Institute of Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TH, UK
[4]Healthcare Technology Institute, Institute of Translational Medicine, Mindelsohn Way, Birmingham, B15 2TH, UK

S1. LoC Performance

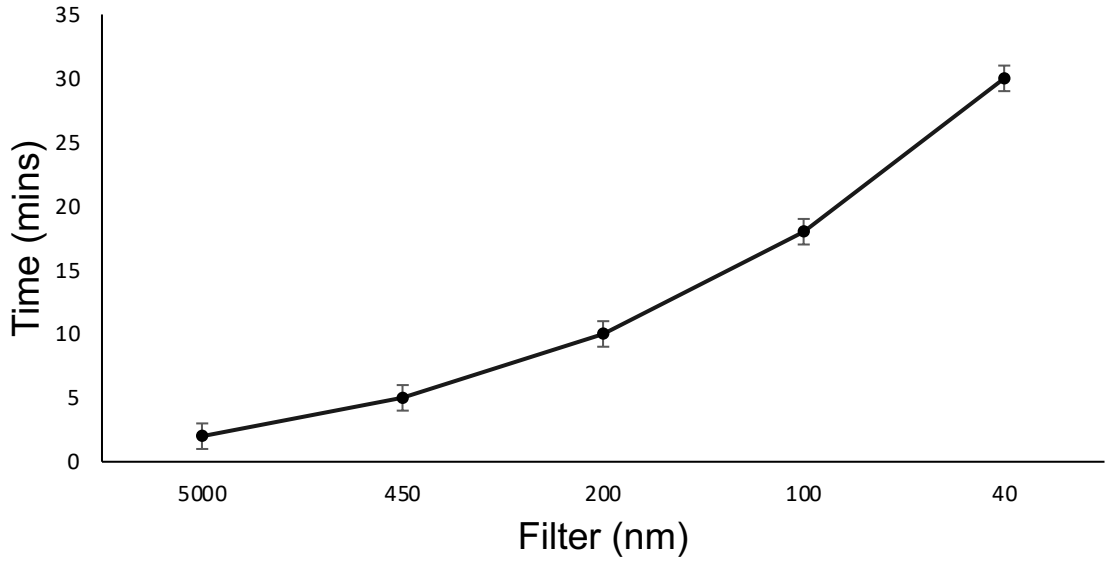

Figure S1. Overall time taken to filter saliva through the LoC. Based on 10 sample repeats.

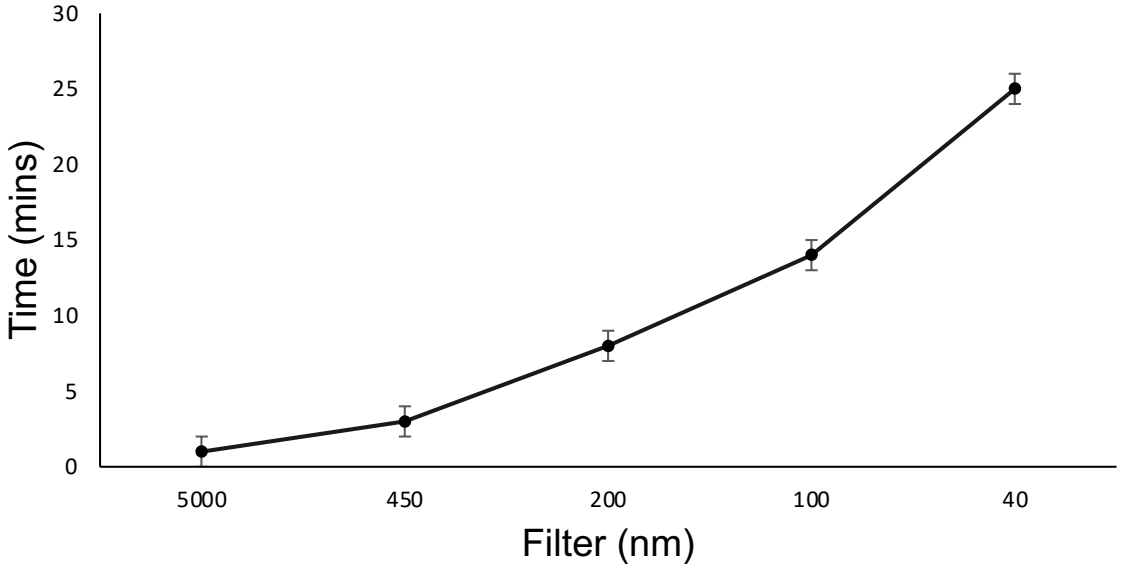

Figure S2. Overall time taken to filter blood plasma through the LoC. Based on 10 sample repeats

S2. EV Indicative Biomarkers

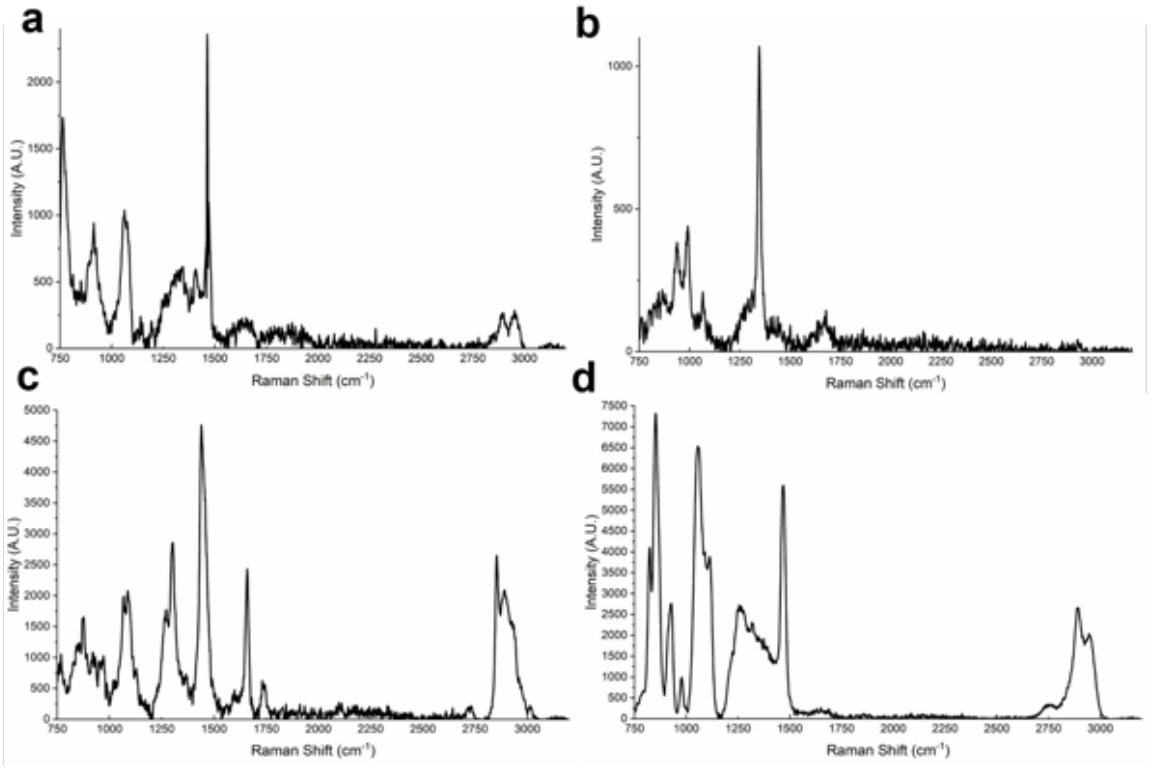

Figure S3. Raman extended scans of EV indicative biomarkers **(a)** AHA1, **(b)** CD81, **(c)** phosphatidylcholine and **(d)** ST13.

S3. Additional Box Plots

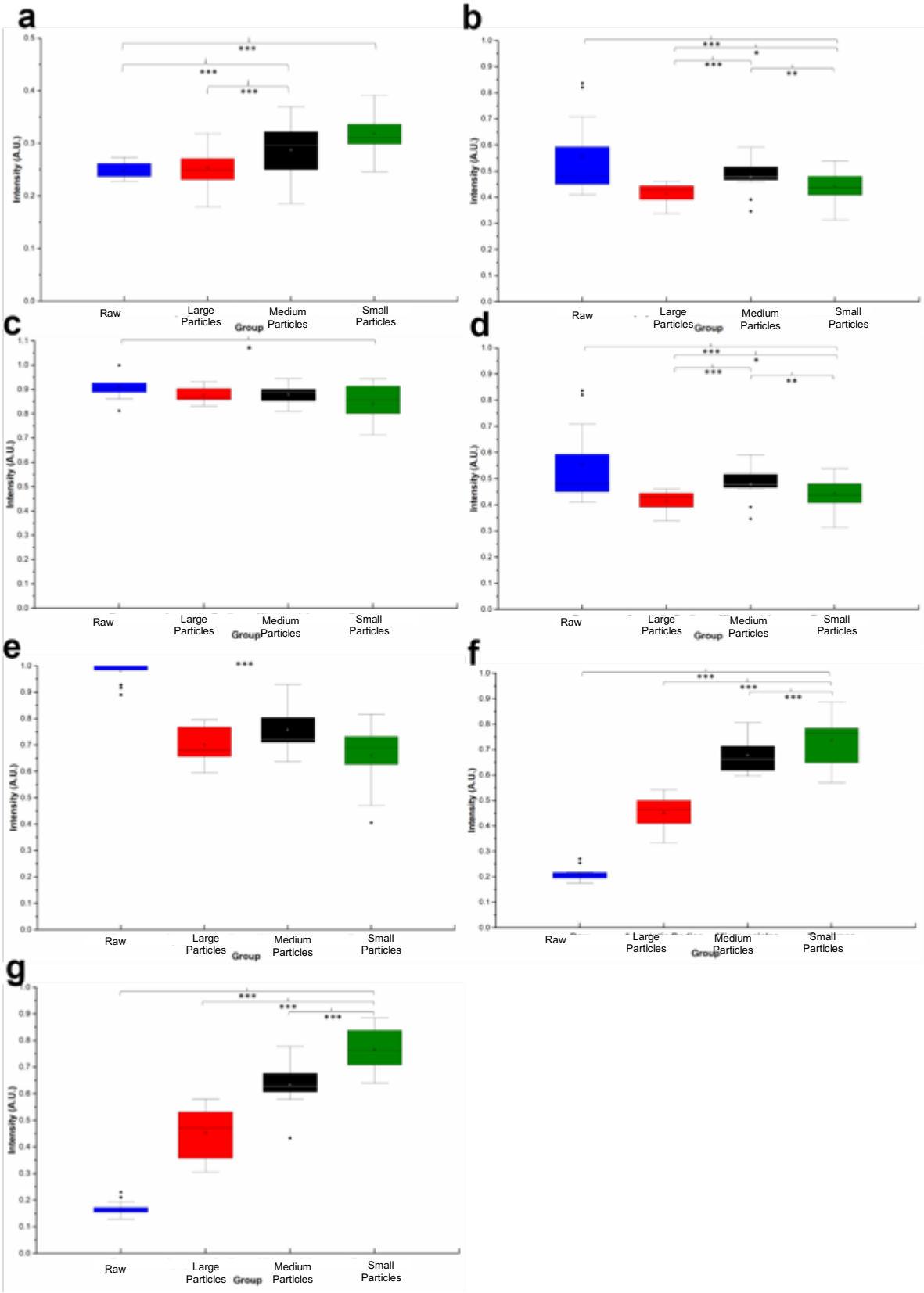

Figure S4. Boxplots illustrating average Raman peak intensity distributions at wavenumbers **(a)** 851 cm^{-1}, **(b)** 960 cm cm^{-1}, **(c)** 1003 cm^{-1}, **(d)** 1320 cm^{-1}, **(e)** 1336 cm^{-1}, **(f)** 1448 cm^{-1} and **(g)** 1657 cm^{-1} for

EV subgroups and raw blood plasma. The box represents the interquartile range (IQR) with the median indicated by a line inside the box. Whiskers extend to 1.5 times the IQR, and the diamonds denote outliers beyond the whiskers. Intensity levels were significantly higher between EV subgroups and raw plasma with * denoting $p < 0.05$, ** $p < 0.01$ and *** $p < 0.0001$, student's *t*-test.

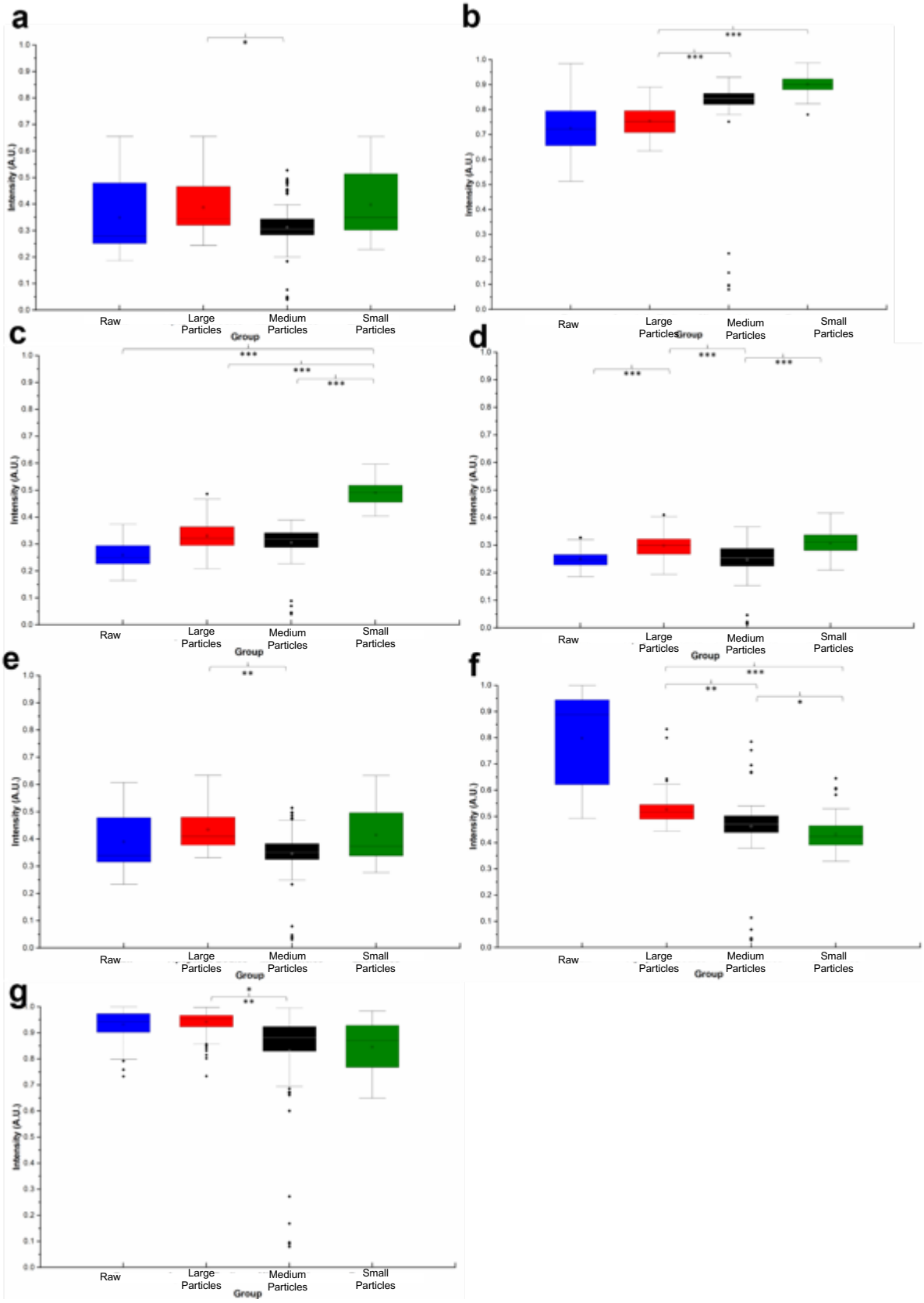

Figure S5. Boxplots illustrating average Raman peak intensity distributions at wavenumbers **(a)** 851 cm^{-1}, **(b)** 960 cm cm^{-1}, **(c)** 1003 cm^{-1}, **(d)** 1320 cm^{-1}, **(e)** 1336 cm^{-1}, **(f)** 1448 cm^{-1} and **(g)** 1657 cm^{-1} for

EV subgroups and raw blood plasma. The box represents the interquartile range (IQR) with the median indicated by a line inside the box. Whiskers extend to 1.5 times the IQR, and the diamonds denote outliers beyond the whiskers. Intensity levels were significantly higher between EV subgroups and raw plasma with * denoting $p < 0.05$, ** $p < 0.01$ and *** $p < 0.0001$, student's *t*-test.

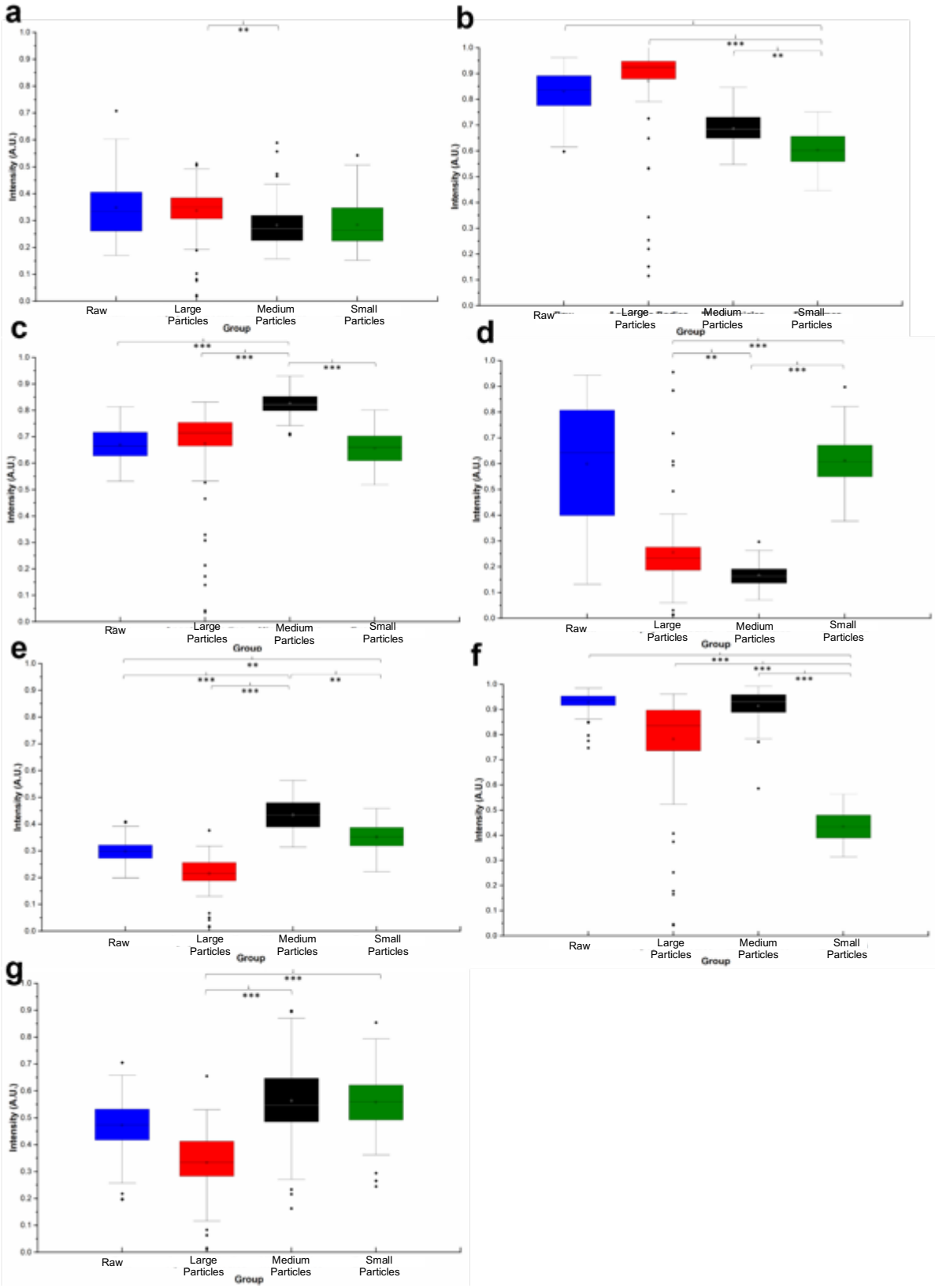

Figure S6. Boxplots illustrating average Raman peak intensity distributions at wavenumbers **(a)** 874 cm^{-1}, **(b)** 928 cm^{-1}, **(c)** 1003 cm^{-1}, **(d)** 1047 cm^{-1}, **(e)** 1296 cm^{-1}, **(f)** 1448 cm^{-1} and **(g)** 1656 cm^{-1} for EV subgroups and raw saliva. The box represents the interquartile range (IQR) with the median indicated

by a line inside the box. Whiskers extend to 1.5 times the IQR, and the diamonds denote outliers beyond the whiskers. Intensity levels were significantly higher between EV subgroups and raw plasma with * denoting $p < 0.05$, ** $p < 0.01$ and *** $p < 0.0001$, student's *t*-test.

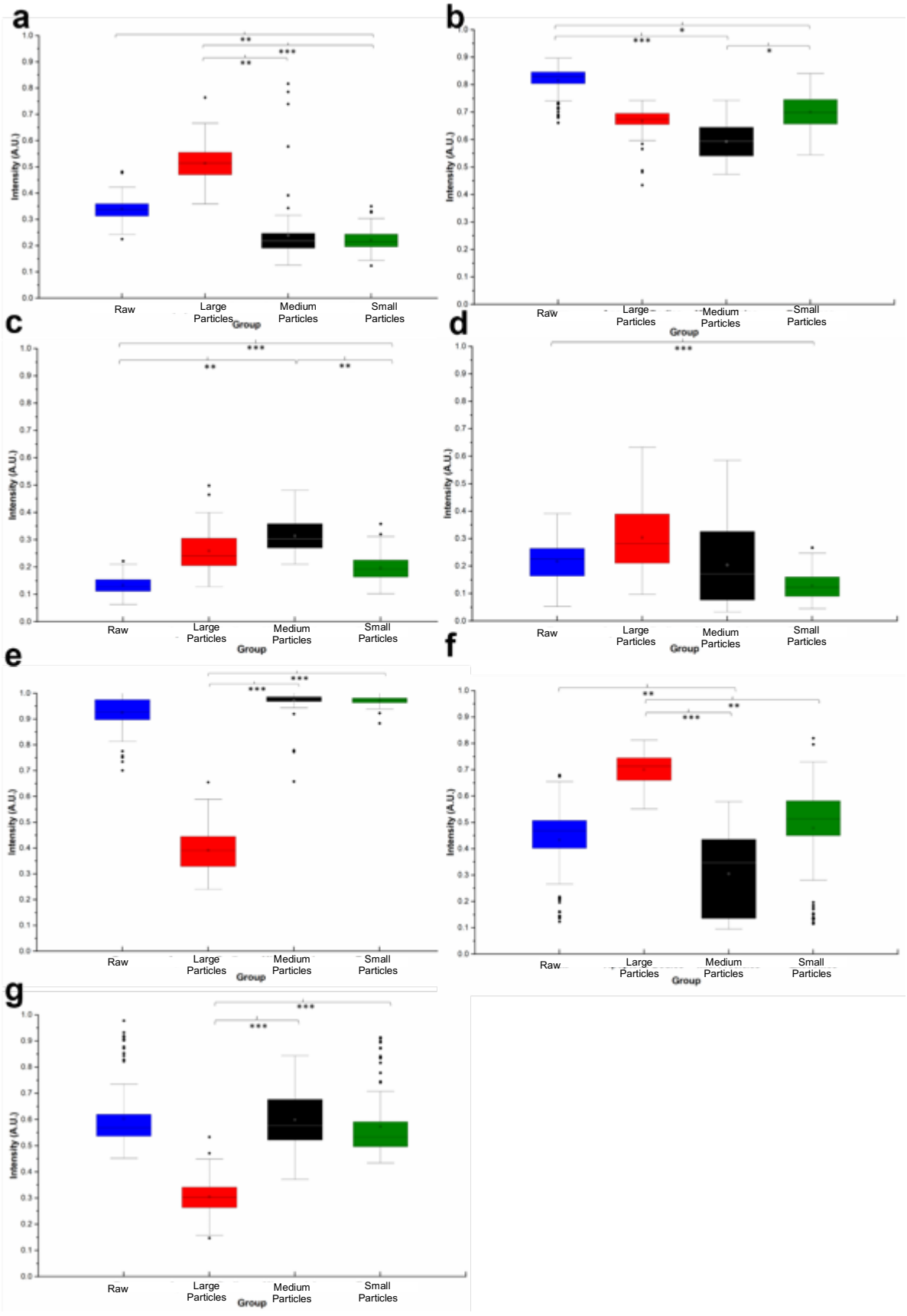
a
b
c
d
e
f
g
Intensity (A.U.)
Raw
Large Particles
Medium Particles
Small Particles
Group

Figure S7. Boxplots illustrating average Raman peak intensity distributions at wavenumbers **(a)** 874 cm^{-1}, **(b)** 928 cm^{-1}, **(c)** 1003 cm^{-1}, **(d)** 1047 cm^{-1}, **(e)** 1296 cm^{-1}, **(f)** 1448 cm^{-1} and **(g)** 1656 cm^{-1} for EV subgroups and raw saliva. The box represents the interquartile range (IQR) with the median indicated by a line inside the box. Whiskers extend to 1.5 times the IQR, and the diamonds denote outliers beyond the whiskers. Intensity levels were significantly higher between EV subgroups and raw plasma with * denoting $p < 0.05$, ** $p < 0.01$ and *** $p < 0.0001$, student's *t*-test.

S4. Additional RS Data – IBD

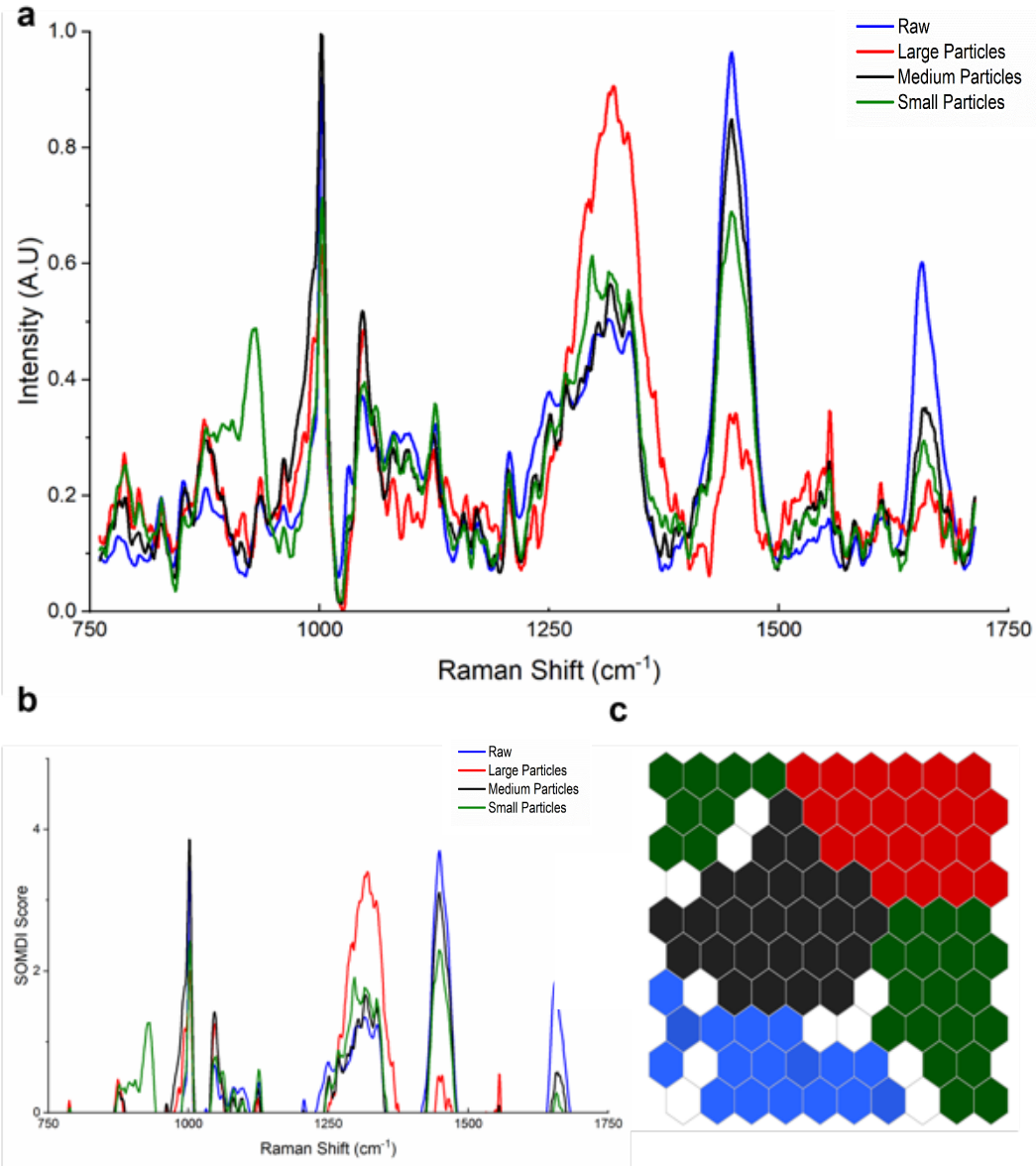

Figure S8. (a) Representative Raman spectra illustrating Raman spectral fingerprint of UC-indicative EVs. **(b)** SOMDI/SOM of UC patient saliva successfully classifying EVs according to size/subgroup with an accuracy of

98.2 ± 0.4%. Clustering of raw saliva is indicated in blue, large particles in red, medium particles in black and small particles in green.

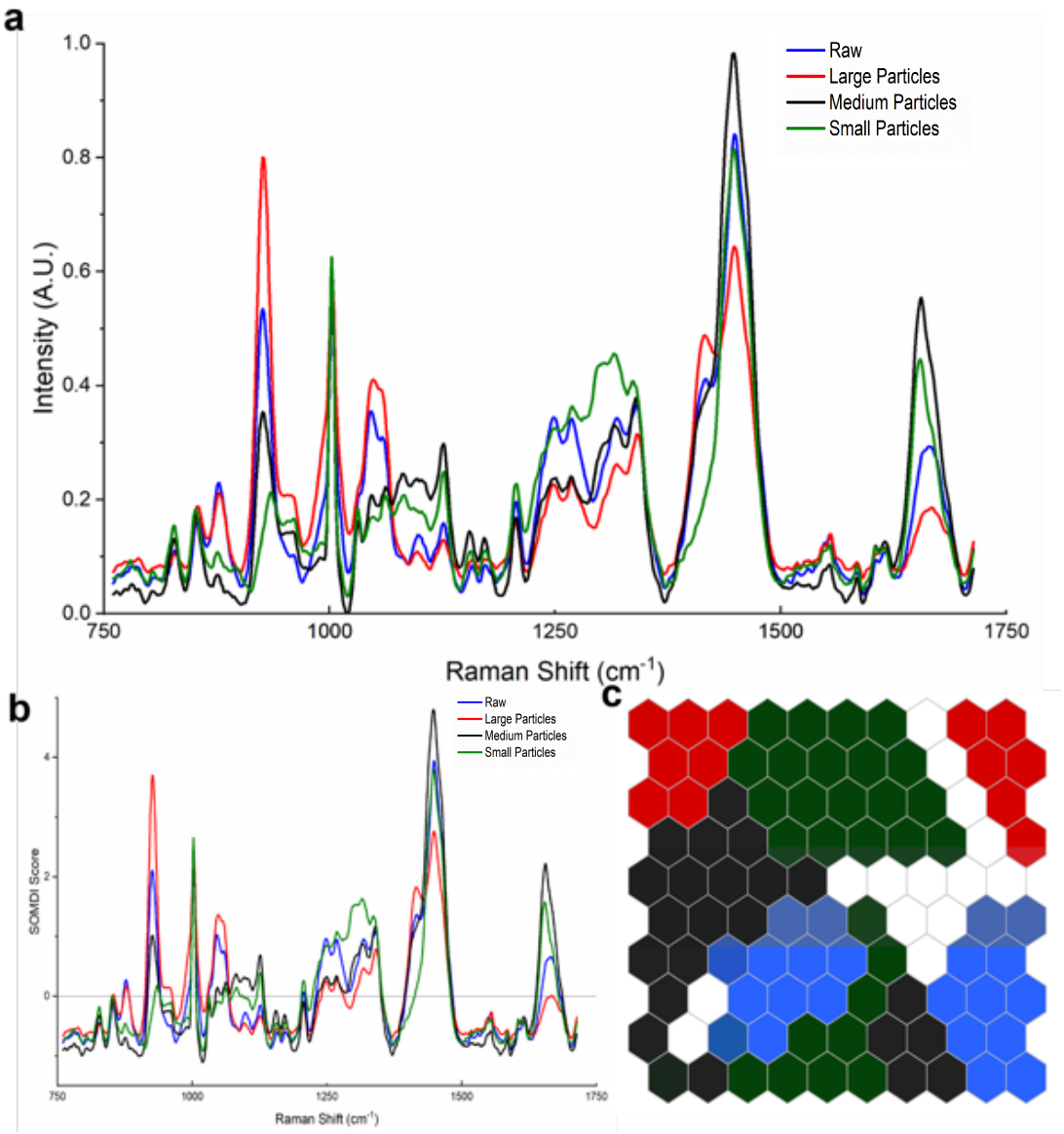

Figure S9. (a) Representative Raman spectra illustrating Raman spectral fingerprint of CD-indicative EVs. **(b)** SOMDI/SOM of UC patient saliva successfully classifying EVs according to size/subgroup with an accuracy of 92.1 ± 1.2%. Clustering of raw saliva is indicated in blue, large particles in red, medium particles in black and small particles in green.

Table S1. % change in CVD-indicative biomarkers in raw blood plasma at significant Raman shifts of interest.

Marker	Raw Plasma (Raman Shift cm^{-1})					
	812	**1003**	**1296**	**1315**	**1448**	**1656**
IL-9	34.1	13.4	16.7	72.2	271.4	23.5
ApoB	173.3	15.8	106.7	266.7	0.02	20.1
PCSK9	55.9	24	42.3	23.5	69.6	140
LpA	10.1	24	59.4	34.1	51	27.2
NT-ProBNP	60	0.03	54.5	42.6	10.3	11.6

Table S2. % change in CVD-indicative biomarkers in large particles derived from blood plasma at significant Raman shifts of interest.

Marker	Large Particles (Raman Shift cm^{-1})					
	812	**1003**	**1296**	**1315**	**1448**	**1656**
IL-9	17.4	37.3	0.06	66	338	0.03
ApoB	78.2	43.8	38.7	0.07	19.5	20.7
PCSK9	32.3	31.4	36.8	14.2	100	25
LpA	33.6	17.9	42.4	29.8	31.4	62.8
NT-ProBNP	82.6	17.5	24.6	29.8	31.4	62.8

Table S3. % change in CVD-indicative biomarkers in medium particles derived from blood plasma at significant Raman shifts of interest.

Marker	Medium Particles (Raman Shift cm^{-1})					
	812	**1003**	**1296**	**1315**	**1448**	**1656**
IL-9	18.5	41.8	0.06	66	338	0.03
ApoB	46.3	48.4	38.7	66.7	19.5	17.1
PCSK9	35.3	35.7	47.1	49.2	100	250
LpA	35.3	167	48.6	19.5	0.06	0.06
NT-ProBNP	47.6	21.8	42.4	29.8	31.4	62.8

Table S4. % change in CVD-indicative biomarkers in small particles derived from blood plasma at significant Raman shifts of interest.

Marker	Small Particles (Raman Shift cm^{-1})					
	812	**1003**	**1296**	**1315**	**1448**	**1656**
IL-9	0.04	46.3	38.9	58.8	338	23.5
ApoB	31.7	53.1	19.4	59.6	22.1	10.3
PCSK9	17.6	40	0.03	36.9	104	160
LpA	17.6	165	32.4	0.03	0.04	21.2
NT-ProBNP	33.3	25.6	24.2	14.9	34.3	20.9