INVESTIGATING THE CLINICAL APPLICATION OF VIBRATIONAL SPECTROSCOPY: A NOVEL TECHNIQUE FOR THE DIAGNOSIS OF COMMON VARIABLE IMMUNE DEFICIENCY

A thesis submitted to The University of Manchester for

the degree of

Doctor of Clinical Sciences

in the Faculty of Biology, Medicine and Health

2021

Emma L. Callery

School of Health Sciences

Table of Contents

Abstract....4

Declaration....5

Acknowledgements....6

Chapter 1: Introduction 7

1.1 Published Review Article: Vibrational spectroscopy and multivariate analysis techniques in the clinical immunology laboratory: a review of current applications and requirements for diagnostic use 8

1.2 Introduction (continued)....9

1.2.1 Common Variable Immunodeficiency 9

1.2.2 The need for a new approach to diagnosis and management 10

1.2.3 Aims and Objectives....15

1.2.4 Summary of Aims 16

1.3 Alternative thesis 16

References 22

Chapter 2: Method Evaluation....27

2.1 Background 28

2.1.1 Evaluation of pre-processing techniques and chemometric methods applicable to ATR-FTIR spectroscopy....28

2.1.2 Assessment of blood sample type suitability: Serum versus plasma 30

2.1.3 Evaluation of environmental influence....32

2.2 Materials and Methods....32

2.2.1 Population....32

2.2.2 Sample Collection and preparation 32

2.2.3 ATR-FTIR spectroscopy....33

2.2.4 Pre-processing techniques....35

2.2.5 Multivariate analysis 37

2.3 Results & Discussion 41

2.3.1 Verification of pre-processing techniques – Results....41

Figure 2.3.4. Examining the effect of pre-processing techniques on subsequent multivariate analysis techniques (unsupervised and supervised separation)....47

2.3.2 Verification of pre-processing techniques – Discussion 48

2.3.3 Investigating choice of biofluid - Results 51

2.3.4 Investigating choice of biofluid - Discussion 58
2.3.5 Examining environmental variance - Results 62
2.3.6 Examining environmental variance - Discussion 64
2.4 Summary of all Method Evaluation Results 66
References 67

Chapter 3: Research Study 69
Published Research Article: New approach to investigate Common Variable Immunodeficiency patients using spectrochemical analysis of blood 69

Chapter 4: Critical Appraisal of the Research Project 70
Abstract 71
4.1 Introduction 71
4.2 Critical appraisal of the study protocol and methodological approach 74
4.2.1 Common variable immune deficiency as the study population 74
4.2.2 ATR-FTIR as the biospectroscopy method 77
4.2.3 Sample collection and processing 79
4.2.4 Data Analysis and interpretation 82
4.3 Contribution to theory and clinical practice (including suggestions for further work) 84
4.4 Final thoughts and conclusion 88
References 93

APPENDIX A: INNOVATION PROPOSAL 988
APPENDIX B: ORAL PRESENTATION FEEDBACK..1100
APPENDIX C: IRAS AND HRA APPROVALS 1111
APPENDIX D: PATIENT INFORMATION AND CONSENT 1200
APPENDIX F: QUALIFICATIONS..1277
APPENDIX G: SUPPLEMENTARY INFORMATON TO CHAPTER 3 1322
APPENDIX H: ADDITIONAL RESEARCH CONDUCTED DURING HSST..1433

Word Count: 38,710 words

Abstract

Common variable immunodeficiency (CVID) is the most prevalent symptomatic primary immunodeficiency diagnosed in adults. CVID is considered to be a spectrum of different antibody deficiency disorders, with only 10-20% of cases attributed to monogenic mutations. Clinical features vary between patients. The majority have increased susceptibility to infection; however non-infectious complications (autoimmunity, inflammation and malignancy) have the greatest influence on morbidity and mortality. The laboratory features of low serum immunoglobulin G, A, and/or M, alongside poor vaccine responses contribute to the diagnosis of CVID; however the low specificity of these tests renders diagnosis one of exclusion. Diagnostic delay remains a significant problem worldwide. Immunophenotyping can be used to subgroup patients based on shared clinical features; however predictive biomarkers to identify patients at risk of severe disease have yet to be identified. In light of the limitations associated with current laboratory tests for CVID, a new diagnostic approach is needed.

Vibrational spectroscopy is a powerful technique that can be used to characterise the molecular composition of a sample. Within a biological sample, important molecules such as lipids, carbohydrates, nucleic acids and proteins are held together by chemical bonds; these bonds will vibrate following excitation with infrared light. By measuring the vibrational energy of each molecule present in a sample, a unique spectrum, known as the 'molecular fingerprint' is generated. Fourier Transform infrared (FTIR) spectroscopy is a mode of vibrational spectroscopy gaining wider application in the clinical setting over the past decade. In this method, bond vibrations cause a change in the dipole moment of molecules; these vibrations can be quantified as spectral peaks. As disease-related changes in biological samples will be reflected in the molecular fingerprint, FTIR spectroscopy is a well-placed candidate for the investigation of disease.

The experimental approach in this thesis uses FTIR spectroscopy to characterise the molecular fingerprint of blood serum and plasma samples of CVID patients. We examined two biologically relevant regions of the spectrum, Fingerprint (1800-900 cm^{-1}) and High (3700-2800 cm^{-1}). We determined the Fingerprint region to have superior performance for potential use as a diagnostic technique. Following the application of machine-learning algorithms, we successfully classified CVID patients from Healthy controls with sensitivities and specificities of 97% and 93%, respectively, for serum; and 94% and 95%, respectively, for plasma. Key spectral features capable of discriminating CVID patients from Healthy controls were identified for both serum and plasma samples; with a greater number of biomarkers associated with the Fingerprint region of the spectrum. Wavenumbers in regions indicative of nucleic acids (984 cm^{-1}, 1053 cm^{-1}, 1084 cm^{-1}, 1115 cm^{-1}, 1528 cm^{-1}, 1639 cm^{-1}), and a collagen-associated biomarker (1034 cm^{-1}) were found to have statistically significant ($p < 0.0005$) absorbance intensity differences in the CVID group compared to the healthy controls. Future studies to validate these findings are required prior to translation into the clinical setting.

The CVID-specific spectral variances identified in this study may present future candidate biomarkers and provide new knowledge into the aetiology of CVID, which remains largely undefined. Vibrational spectroscopy demonstrates a promising new approach which may improve the diagnosis and management of CVID patients. Earlier diagnosis of patients alongside the identification of prognostic markers may prevent development of severe complications, leading to better outcomes for patients.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, the University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in the University's policy on Presentation of Theses.

Acknowledgements

Firstly I would like to thank my supervisors Professor Anthony Rowbottom and Dr Peter Arkwright for their help, guidance and encouragement during the course of my project. The time and effort they have both provided has been greatly appreciated. Special thanks go to Anthony for his knowledge, advice, and enthusiasm, particularly regarding the experimental approach. His support and belief in me over the last five years has been invaluable.

Secondly I would like to thank Professor Frank Martin at The University of Central Lancashire and all past and present members of his research team. Without the guidance, expert knowledge and training provided by this group, my research project would not have been possible.

I would also like to thank the Immunology Department at Royal Preston Hospital, and I am grateful to Dr Pavaladurai Vijayadurai and Dr Ariharan Anantharachagan for their help in the recruitment of patients, and clinical input throughout. I am extremely thankful to all of the patients who provided additional samples for the purpose of this research.

Finally I would like to acknowledge the support provided by my friends and family. I would like to extend special thanks to my Mum, not only for proofreading my work on numerous occasions, but for her endless encouragement throughout all of my academic endeavours. I am especially grateful for the love and support of my partner, David. Thank you for having unwavering belief in me and for making me smile again whenever things became too serious.

Chapter 1: Introduction

Chapter 1: Introduction

1.1 Published Review Article: Vibrational spectroscopy and multivariate analysis techniques in the clinical immunology laboratory: a review of current applications and requirements for diagnostic use

Emma L. Callery & Anthony W. Rowbottom

Key words: Immunology, clinical, biological, infra-red, Raman, spectroscopy, molecular, diagnostics

Paper published in Applied Spectroscopy Reviews (Appl. Spectrosc. Rev, 2021)

DOI:10.1080/05704928.2021.1958337

APPLIED SPECTROSCOPY REVIEWS
https://doi.org/10.1080/05704928.2021.1958337

Vibrational spectroscopy and multivariate analysis techniques in the clinical immunology laboratory: a review of current applications and requirements for diagnostic use

Emma L. Callery[a] and Anthony W. Rowbottom[a,b]

[a]Department of Immunology, Lancashire Teaching Hospitals NHS Foundation, Preston, UK; [b]School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston, UK

ABSTRACT
Laboratory tests are essential for clinicians to reach an accurate diagnosis and informing appropriate treatments. The expansion in the use of immunotherapies has highlighted the gap between the knowledge of molecular pathways and targeted therapies with availability of laboratory tests. The translation of vibrational spectroscopic techniques such as Fourier-transform infrared (FTIR) spectroscopy and Raman spectroscopy into clinical practice offer rapid-, noninvasive and inexpensive methods to obtain information on the molecular composition of biological samples. Advances in instrumentation, data analysis and machine learning techniques are key developments that have permitted the availability of results to clinicians in an appropriate timescale. Immunological disorders are complex, often demonstrating interaction across multiple molecular pathways which results in delayed diagnosis. Vibrational spectroscopy is being applied in many fields and here we present a review of its potential use in clinical immunology. This review addresses the potential use of spectroscopy in clinical immunology. Potential benefits that these novel techniques offer, including enhanced definition of molecular process and its use in disease diagnosis, monitoring and treatment response is discussed. Whilst not covered extensively, an overview of the method principle, quality control processes, and the requirements for multivariate data analysis is included to provide the reader with sufficient understanding of its application in the clinical setting.

KEYWORDS
Immunology; clinical; biological; infra-red; Raman; spectroscopy; molecular; diagnostics

Introduction

The investigation of pathological disease

Medical diagnoses can be improved by rapid, highly-sensitive and quantitative analysis of biological samples. The diagnostic investigation of pathological diseases relies on the detection of disease-driven structural or functional changes that occur in cells, tissues and organs, or in identification of biomarkers present in biofluids such as blood, urine, or CSF.[1] Advancements over the last century have led to significant improvements in the delivery of healthcare and has ultimately resulted in an increasing life expectancy

CONTACT Emma L. Callery emma.callery@lthtr.nhs.uk; Anthony W. Rowbottom anthony.rowbottom@lthtr.nhs.uk Department of Immunology, Lancashire Teaching Hospitals NHS Foundation, Preston, UK.
© 2021 Taylor & Francis Group, LLC

across society.[2] However, despite continued scientific research and exponential technological advances, the diagnosis and treatment of many pathological disorders remain irremediable. Innovative approaches must therefore be sought to further our understanding of complex diseases, to uncover novel diagnostic platforms and to imminently provide superior, personalized treatment options. As a speciality, pathological investigations provide a valuable means for disease diagnoses, but perhaps more importantly, contribute to the understanding of disease-related mechanisms. Once the etiological processes of a disease have been elucidated it is then possible to transfer this scientific knowledge into clinical practice; explaining the basis of symptoms, improving diagnostic testing and discovering new treatment options to improve patient outcomes.[3]

The coming of age for immunology

There has been an unprecedented pace of change within the field of immunology in the 21st century, driven by both scientific research and medical application. During an explosion of discoveries in the 1900s, the term 'immunobiology' was coined and the foundations for both the innate and adaptive immune system were mapped out.[4] Following on from this, the success of the Human Genome Sequencing Project provided new opportunities for immunological research, igniting new ideas and innovative experimental approaches for the diagnosis and treatment of immune-mediated disorders that had not previously been possible.[5] As a consequence, the use of tools to manipulate the immune system (immunotherapy) now play a ground-breaking therapeutic role in the fields of malignancy, transplantation and immune dysregulation; with the production of monoclonal antibodies[6] and the use of checkpoint blockade and genetically modified T cells as cancer therapeutics[7] being awarded Nobel prizes in 1984 and 2018 respectively. The discipline of immunology has finally taken center stage and proves as vital as ever.

The immune system is comprised of a complex network of tissues and cells which is organized throughout the entire body, joined together by circulating leukocytes and soluble mediators. It is therefore not surprising that immune-mediated conditions can affect anyone at any age, and span almost every speciality of medicine, including pediatrics, infectious disease, rheumatology, allergy, oncology, respiratory, gastroenterology, nephrology, hematology, dermatology, and neurology. Major disorders of the immune system include primary and secondary immune deficiencies, which result in an increased susceptibility to infections, malignancy and autoimmunity; allergy, which results in an exaggerated, inappropriate immune response to a usually harmless substance, and autoimmunity, whereby a dysregulated immune response is mounted against self, causing damage to organs and tissue.[8]

Is the immunology laboratory equipped for the future?

In order to investigate individual components and function of the immune system in relation to disease, the diagnostic immunology laboratory makes use of a variety of analytical platforms and methodologies including flow cytometry, immunohistochemistry, enzyme-linked immunosorbent assay (ELISA) and related techniques, radial

immunodiffusion, hemolytic assays, nephelometry, turbidimetry, and radioimmunoassay. However, given the complexity of the immune system, our current test armory is challenged by only being able to investigate a finite area of the immune system with each individual test. It is a shortcoming to consider that the majority of diagnostic laboratories around the world will also have the same, rather limited repertoire of techniques available to them. Future immunology will involve taking a wider, systems biology approach to the study of the immune response in both physiological and pathological conditions. Through the large-scale analysis of the immune response on multiple levels (genome, transcriptome, proteome), a far more comprehensive view will be possible, uncovering new concepts and development in basic immunology including the intricate cross-talk of cytokine signaling, the molecular mechanisms underpinning the differentiation and plasticity of immune cells and the role of the regulatory network in preventing disease.[4,5]

The forward momentum generated from the Human Genome project was underpinned by advances in technical platforms of genomics, cell biology, molecular biology and bioinformatics.[5] Despite the availability and potential of these new tools, the routine clinical immunology laboratory continues to rely on conventional *in vitro* techniques for the investigation of disease. In order to keep up with the on-going advances in immunotherapies and personalized medicine, the diagnostic laboratory must adopt and translate novel analytical platforms into routine use. As evidenced in the field of genomics where an astonishing amount of data is generated from a single sample, bioinformatics is vital to extract the relevant information for clinical interpretation. As advances in the field have enabled 'big data' analysis to become more accessible and user-friendly, the time for the routine application of mathematical modeling and machine learning is here. Importantly within the specialty of immunology, these *in silico* techniques have already been applied to describe various aspects of the immune system, thus highlighting the potential clinical utility for use within the diagnostic laboratory.[9–14]

Vibrational spectroscopy, coupled with bioinformatics is a novel analytical platform that provides a rapid, low-cost molecular approach to pathological investigations. This makes it an ideal candidate for diagnostic medicine. Translation of vibrational spectroscopy into the immunology laboratory would provide a vital tool to keep up with the current pace of change and to transition into the future of molecular diagnostics.

Vibrational spectroscopy

Vibrational spectroscopy is an informative optical technique that is being increasingly applied in the field of diagnostic medicine. By providing a measure of how light interacts with matter, it enables the elucidation of unique biomolecular features of a given sample in a nondestructive, label-free manner.[15,16] Vibrational spectroscopy is a well-suited technique to study complex heterogeneous samples such as blood, due to the fact that important biomolecules (such as lipids, carbohydrates, nucleic acids and proteins) have characteristic chemical structures that can be specifically determined by their unique spectral signature. Given that the composition and structure of biomolecules in a blood sample will vary depending on the current physiological or pathological state,

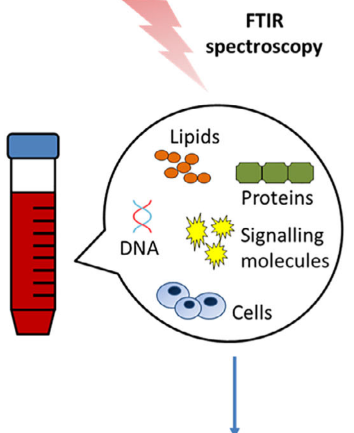
vibrational spectroscopy offers a potential diagnostic tool for the investigation of both health and disease.

Vibrational spectroscopy is an umbrella term to describe the techniques that have allowed scientists to observe the changes in the vibrational and rotational modes (transitions) of a given molecule.[1] Following irradiation of a sample with light, molecular transitions emit or absorb energy; this causes the molecules to move into new vibrational or electronic energy state, detectable by vibrational spectroscopy. Characteristic absorption spectra can be generated for individual samples due to the fact that each molecule present within the sample has a unique vibrational frequency (i.e., DNA, RNA, proteins, lipids, carbohydrates, collagen and glycogen). These spectra can be used to determine a sample's molecular identity following the application of computational analysis techniques (chemometrics).[16–18]

The absorption spectra of a sample will vary according to biochemical changes, some of which may be due to the presence of disease; vibrational spectroscopy is therefore a well-placed candidate for the study of pathological processes thus providing a potential novel diagnostic platform.[19] The most important optical techniques are infrared (IR) and Raman spectroscopy[19]; both of which are well established methods for studying sample types such as biofluids, tissues and cell cultures.[1] These methods are based on the two physical mechanisms involved in the excitation of molecular vibrations; inelastic scattering of photons (Raman spectroscopy) or the absorption of light energy (IR spectroscopy).[20] Raman and IR are fundamentally different techniques but can be used in a complementary fashion to increase the specificity of molecular identification.[19] In both Raman and FTIR spectroscopy methods, the detection and measurement of molecular excitations in the form of vibrational energy generates a unique pattern of spectral peaks (spectral bands) dependent on the biochemical composition of the sample, illustrated in Figure 1. Further analysis of each sample spectrum can be performed on a qualitative and quantitative basis, through the determination of individual peak positions (wavenumbers) and corresponding signal intensity (peak area), respectively. This enables vibrational spectroscopy to be used as a highly sensitive and specific analytical technique for the interrogation of complex sample types, and as a result is now being used extensively in biomedical research.[15]

The molecular fingerprint

The information generated from Raman and FTIR spectroscopy has been coined the 'molecular fingerprint', due to the unique spectral signature obtained from a given material.[21] As vibrational spectroscopy examines the entirety of a sample, rather than a single analyte, it provides an overall picture of the biological status at a given timepoint. When applied to biomedical samples, a unique pattern consisting of thousands of spectral peaks will be produced in a single experiment; this generates an enormous amount of data relating to the structural and functional properties of the sample.[22,23] In light of this, comparisons have been made between spectroscopy techniques and the 'omics' methodologies such as genomics, metabolomics, proteomics, and transcriptomics.[1]



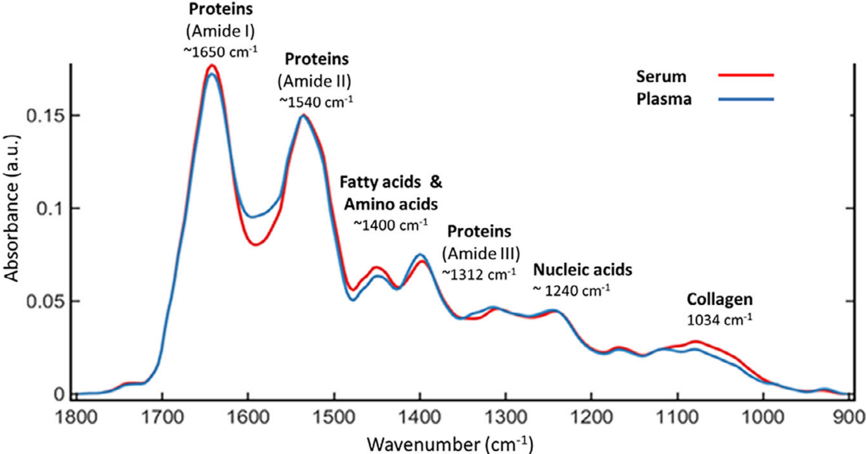


Figure 1. Generating unique FTIR spectra for biological samples. Following interrogation with IR light, the changes in light absorbance can be measured and a biological spectral fingerprint generated. The characteristic spectrum of blood serum (red) and plasma (blue) is illustrated, with six key wavenumber peaks highlighted alongside their associated biomolecular component assignments.

The multitude of potentially diagnostic features generated within the molecular fingerprint of a sample can only be harnessed through the use of intelligent information technology and appropriate chemometric tools.[1] Thus, technological advances in both spectroscopic data acquisition and chemometric processing techniques have been major contributing factors to the recent explosion of research studies in the clinical setting.[15] Nevertheless, the interpretation of the enormous dataset obtained from vibrational spectroscopy of biological materials can be challenging, particularly when attempting to assign subtle spectral features to discrete molecular compounds.[1] In this regard it is important to appreciate that the spectral variations observed between samples may not always be due to pathologically-recognized structural or metabolic modifications. In these cases, the detection of spectral differences can instead provide an early warning

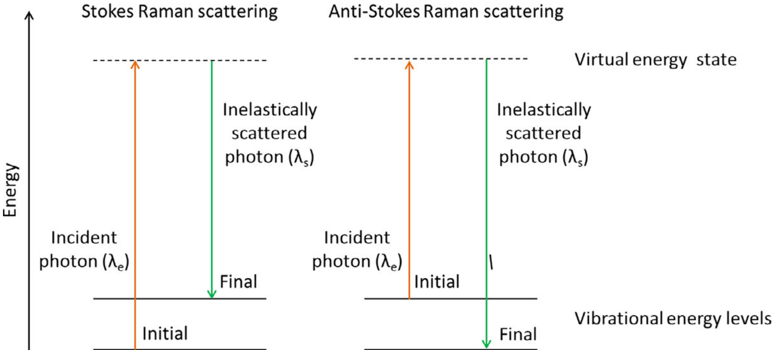


Figure 2. Stokes and anti-stokes Raman scattering. Following transient excitation of molecular bonds to the virtual energy state, the change in wavelength between the incident (λ e) and resultant (λ s) light is directly related to the amount of energy required to shift a particular molecular bond into a different vibrational energy level; this forms the basis of vibrational spectroscopy. Figure adapted from 32.

signal, highlighting samples that require further laboratory analysis, or alternatively, identifying novel disease-related spectral biomarkers.[1]

Principles of vibrational spectroscopy

Light energy is transported across space in electromagnetic (EM) waves. The energy is not continuously distributed but is packaged within individual photons. The energy of the photon can be expressed by the equation $E = h\ v$, where h is Planck's constant ($h = 6.626 \times 10^{-34}$ J s) and v is the frequency of the photon. The EM spectrum is made up of different wavelengths; the frequency (v) of each photon is inversely related to wavelength (λ), as described by the equation $v = c/\lambda$, where c is the speed of light.[24] In spectroscopy, the wavenumber ($\tilde{v}$), in cm^{-1}, of a particular bond interaction is usually referenced as a more manageable figure. Wavenumbers are widely cited throughout the literature and in reference libraries as a means for analysts and researchers to identify specific biomolecules.[17,18,25]

The interaction between light (photons) and matter can result in a change in the energy level of molecules present in the material, from this a number of outcomes are possible such as reflection, absorption, elastic scattering (Rayleigh), inelastic scattering (Raman) and fluorescence.[26] Photons that are not directly reflected at the surface of a sample will pass through it, resulting in events that can potentially change the frequency and energy of the photon. Through the absorption of energy, molecules become excited and are promoted into a higher electronic energy state; as the molecules return to their original ground state they can release absorbed energy (at a longer wavelength) as fluorescence.[27] If the incident energy of a photon remains unchanged, this is known as elastic- or Rayleigh scattering. If, however, the interactions of photons with molecules in the sample result in a transfer of energy (manifesting as the vibrational energy state of the molecule), this is known as inelastic- or Raman scattering.[19,28]

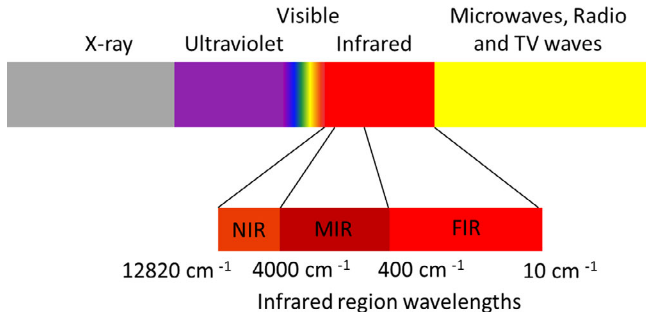


Figure 3. The electromagnetic spectrum; highlighting the wavelengths of the infrared spectrum (12,820 to 10 cm^{-1}) and its subdivisions.

Raman spectroscopy

Although most of the light scattered by a molecule results in elastic scattering, approximately one in a million photons will be inelastically scattered, altering the frequency of the incident wavelength to provide a quantifiable signal, without causing any damage to the sample. Following this brief excitation, the molecules return back to their ground state. The energy change is infinitely short following the scattering of photons therefore it is referred to as a 'virtual energy state' (and is different to the change in electronic energy state required for samples to fluorescence).[29] Raman scattering, known as the 'Raman shift' most often results in a transfer of energy from the photon to the molecule; the loss of energy from the photon increases its wavelength and is described as Stokes shift. In contrast to this, Anti-Stokes signals can also be detected which come from the energy shift as photons take up energy from already excited molecular vibrations (Figure 2).[15,30] As the signal intensity of the scattered light is proportional to the concentration of a molecule within the sample, the overall Raman spectrum generated from an unknown sample can provide information on both its molecular constituents and the concentration of these present in the sample.[31]

In the examination of complex biological samples with Raman spectroscopy, there will be molecules present that share some of the same chemical bonds within their structure (for example C-N bonds contained within proteins, or C = O bonds within fatty acids). Additionally, there will be molecules present containing different chemical bonds but sharing the same overall vibrational energy. Regardless of this, the detection of discrete Raman shifts (peaks) within the spectrum in conjunction with the signal intensity of each peak can provide a highly specific and unique biochemical read-out of a sample, capable of discriminating between even subtle differences between healthy and pathological specimens[32]

Fourier transform infrared spectroscopy

Fourier transform IR spectroscopy (FTIR) is a high resolution mode of IR spectroscopy that monitors the changes in the vibrational, rotational and stretching modes of individual molecular bonds following irradiation with polychromatic light. An IR absorption spectrum is generated by measuring how much incident light energy is absorbed at a particular wavelength as molecules in the sample to vibrate, absorb energy and move to

a higher energy level.[19,33] FTIR spectroscopy follows the Beer-Lambert law, thus IR-active vibrations can be quantitatively measured when the frequency of absorbed radiation matches the vibrational frequency of the molecule.[19,23] The absorbance frequency of a particular molecule can be determined due to changes in the dipole moment during the vibration; this is dependent on the polarity, size and shape of the molecule, and type of bond (single, double, triple) present.[34] If a molecule is not in a dipole condition it will not be IR active, thus will not give rise to spectral peaks when excited with IR light. As a general rule, covalently bonded molecules consisting of unalike atoms are IR active, thus virtually all organic molecules will give rise to an IR spectrum.[35]

FTIR has a broader range of application than Raman spectroscopy as almost all molecules can become excited to undergo vibrations after absorption of IR light energy.[19,23] The IR range of electromagnetic spectrum covers the wavelengths of light between 780 nm (12,820 cm^{-1}) to 1 mm (10 cm^{-1}) and can further be divided into near-, mid-, and far-IR regions (Figure 3). Fundamental to biological FTIR spectroscopy is the mid-IR (MIR) wavelength region (~350–4000 cm^{-1}) which includes the well described 'fingerprint' area (1800 to 900 cm^{-1}) of spectral peaks frequently used to characterize organic matter.[36]

There are three modes of sampling for FTIR spectroscopy; transmission, transflection and attenuated total reflection (ATR), relating to how the IR light is directed at, and detected following interaction with the sample. In transmission mode, IR light is passed directly through the sample and the resultant absorption frequencies are subsequently detected. In this case, expensive transparent substrates (e.g., calcium fluoride) are required for the mounting of biological specimens. In transflection modes analysis, the IR light is transmitted through the sample before being reflected off the substrate to pass back through the sample again; the absorbed radiation is then detected. Transflection sampling methods use inexpensive, highly reflective substrates (e.g., low-emission (Low-E) glass slides) for the loading of samples therefore may offer cost advantages over transmission analysis. The third method, ATR-FTIR, is considered the most suitable for the analysis of biological fluids and dried sample films as it is less subjective to unwanted spectral inferences from physical or environmental factors.[37,38] In ATR-FTIR, IR light is directed through an internal reflection element (IRE) with a high refractive index (e.g., diamond/germanium). This generates an evanescent wave which can penetrate a sample in contact with the surface of the IRE. The radiated light waves are absorbed by the sample and the resultant frequencies can be detected.[19,33] This method has demonstrated clinical application through the successful quantification of various serological components in blood[39] therefore would provide the most suitable sampling platform for translation into the clinical immunology laboratory.

Identifying biomarkers for clinical applications

Each peak in the spectrum relates to the absorption frequency of a particular molecule as it becomes exited to vibrate. There are several different kinds of molecular vibrations or oscillations that have been characterized by IR (and Raman) spectroscopy such as stretching (symmetric and asymmetric); in-plane bending (scissoring and out-of-plane bending (wagging and twisting). Many of these vibrations have now been associated with well-defined wavenumbers for the identification of certain functional groups within

Table 1. Assignment of wavenumbers and vibration modes to the major absorption bands observed in the IR spectrum of blood plasma, adapted from 66.

Wavenumber [cm^{-1}]	Assignment
3300	ν(N-H) of proteins (amide A band)
3055–3090	ν(=CH) of lipids and proteins
2950–2960	νas(CH_3) of lipids and proteins
2920–2930	νas(CH_2) of lipids and proteins
2865–2880	νs(CH_3) of lipids and proteins
2840–2860	νs(CH_2) of lipids and proteins
1730–1760	ν(C = O) of fatty acids
1660	ν(C = O) of proteins (amide I band)
1550	δ(N-H) of proteins (amide II band)
1400	ν(COO−) of amino acids
1240	νas(P = O) of nucleic acids
1170–1120	ν(C-O) and ν(C-O-C) of carbohydrates

ν, stretching vibrations; δ, bending vibrations; s, symmetric; as, asymmetric.

a molecule.[17,40] This has led to the extensive application of FTIR spectroscopy across the chemical industry (including biochemistry) to accurately identify the composition of unknown compounds by identifying which molecular bonds are present through their unique vibrational signatures.

It is key, and fundamental to the interpretation of spectral data, to note that an individual absorption peaks observed within the IR spectrum of a heterogeneous sample is not diagnostic of a single molecule, compound or biomarker within the sample. The spectrum produced, even for samples containing a single purified compound will often be made up of numerous absorptions relating to the vibrations from the combination of chemical components present in the molecule. For instance, organic compounds containing carbon-hydrogen (C-H) single bonds will produce characteristic wavenumber vibrations depending on the strength and polarity of the bond. Caution must be applied however when interpreting spectral absorptions relating to organic chemical functional groups (e.g., alcohol (–OH), alkene (C = C), amine (-NH_2)) particularly those with multiple bonds, as there can be variability even within individual vibrational modes depending on the properties of the individual molecule in which they are found.[41] Although it is analytically correct to consider the individual absorptions present in a molecule and assign them to particular molecular vibrations, this is not practical for complex samples containing a multitude of biological molecules. Therefore, in order to apply spectroscopic analysis to the investigation of biological specimens, it is the overall spectrum that must be considered. This is known as the 'molecular fingerprint' and serves as the characteristic identifier for each molecule present in the sample.[41]

Complex absorption spectra are generated from biological specimens due to the vast number of unique vibrations that can be measured in the sample.[42] The most important wavenumber regions relating to key biological components (proteins, lipids, nucleic acids and carbohydrates) are the fingerprint region (1800–900 cm^{-1}) and the high region (3700–2800 cm^{-1}), illustrated in Figure 1 and Table 1. Signal intensity at a given wavelength will vary depending on three main factors: (i) the concentration of the molecule in the sample, (ii) the chemistry of the functional group in the molecule (primary structure), and (iii) the conformation (secondary structure) of the molecule within the sample.[19] For the biological investigation of proteins, the amide I band (1600–1690 cm^{-1}) has received the most attention as its shape, position and intensity

can be used to predict different types of secondary structures.[43] Within this peak, wavenumber assignments for α-helixes (1650 cm^{-1}), β-sheets (1632 cm^{-1}), and random coils (1658 cm^{-1}) have been determined to elucidate different structural segments within a protein.[44] An increased confidence in peak assignment has been demonstrated in recent years as the use of vibrational spectroscopy has expanded and spectral wavenumber libraries for biological material have been generated for both Raman and FTIR.[18,45] Based on the evaluation of the literature describing numerous spectroscopy studies, wavenumber libraries have been constructed to assign distinct chemical bonds and functional groups to the most frequently seen peaks in biological materials. This has provided researchers with an accurate and reliable database to find the meanings of different unknown peak intensities, thus strengthening the evidence for the application of vibrational spectroscopy to both clinical and non-clinical aspects of biochemistry.[18]

Analytical considerations for the interrogation of spectral data

Quality control and pre-processing steps

Prior to the translation of vibrational spectroscopy data into a potential diagnostic tool, several specialized analytical techniques must be applied. Due to the high complexity of data from biological samples and the fact that subtle spectral variations are often undetectable by visual examination, chemometric and machine-based learning algorithms are required to extract the relevant information which could relate to the presence of disease.[42] Quality control (QC) approaches are employed in the first stages of data analysis to ensure unsuitable spectral measurements are removed from the data set. The application and integration of appropriate QC processes are crucial to all analytical methods used for patient diagnosis, monitoring and prognostic evaluation in the healthcare setting, underpinned by the UKAS 15189 accreditation requirements.[46] Similarly to the commercial auto-analysers currently employed within pathology laboratories, most commercial Raman- and infrared spectroscopy instruments have QC packages built in to the spectral acquisition software which provide the first line of QC. The main issues resulting in poor quality spectrum include excessive sample thickness, environmental variations (including temperature, humidity, and pressure), instrumentation errors, and excessive scattering, dispersion or noise detected during the analysis.[47] To reduce the impact of these known sources of error, steps must be taken to ensure there is standardization of specimen collection and storage, and that reproducible processes are followed during sample preparation and spectral acquisition.

Pre-processing steps form a second line of quality control, with several recommended protocols published in the literature.[16,47,48] Pre-processing is essential to standardize and correct unwanted signals (background noise) such as sloping baseline effects, fluorescence, scattering, variations in sample thickness, environmental and instrumental variations, and to remove contaminants (i.e., fixatives or residual water in dried sample films).[47,49] These unwanted signals cause oscillations in the spectrum, producing false absorption bands that do not represent the sample composition. The major pre-processing techniques are cutting, de-noising, baseline correction and normalization. As previously described, during FTIR spectroscopy of biological samples, absorbance occurs in the MIR wavelength region (~350–4000 cm^{-1}) inclusive of the well-described

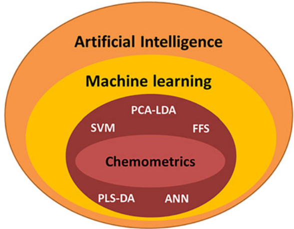


Figure 4. The use of artificial intelligence and machine learning for the analysis of 'big data'. Chemometrics is a speciality of machine learning; several mathematical and statistical techniques can be routinely applied for optimal analysis of complex spectral data. SVM—support vector machine; PLS-DA—partial least squares discriminant analysis; PCA-LDA—Principle component analysis–linear discriminant analysis: ANN—artificial neural networks; FFS—forward feature selection.

fingerprint region (1800–900 cm^{-1}). 'Cutting' of the FTIR spectra to select the region(s) of interest is the first stage of pre-processing, reducing the number of data points for subsequent explorative analysis. As the fingerprint region contains the wavenumber regions relating to biological components (proteins, lipids, nucleic acids and carbohydrates)[16] this region is most commonly selected for initial cutting of biological sample spectra. Cutting of the 'high region' which covers wavenumbers 3700–2800 cm^{-1} has gained increasing interest in recent years for examination of biological samples[17]; key biomolecules in this region include lipids, fatty acids, phospholipids and contribution from protein side chains, nucleic acids, cholesterol and creatinine.[50–53]

Subsequent to cutting, de-noising methods, such as smoothing, baseline correction and normalization are routinely applied pre-processing techniques in FTIR biospectroscopy. Smoothing techniques such as Savitzky-Golay (S-G) smoothing is one of the most common analytical procedures employed to digital data.[54] An SG filter can be applied to a dataset in order to increase the precision when detecting relevant signal peaks. To achieve this, high frequency noise and random error signals must be eliminated without degrading any genuine signal peaks. The SG digital filter is a moving window based on a chosen sets of integers such as polynomial degree, number of convolution coefficients and window size. Caution must be applied when using smoothing methods as they can potentially distort or smooth out relevant peaks, therefore the S-G algorithm tuning parameters must be appropriately selected to avoid any amplification of unwanted signals.[47] Various baseline correction techniques can be applied to FTIR data such as rubber-band correction, polynomial fitting and differentiation-based methods, often combined with SG smoothing to improve signal to noise ratio. These techniques will account for several causes of a skewed baseline including scattering, reflection, temperature fluctuation, sample thickness and instrumental issues.[49] The application and detail of each of these techniques is beyond the scope of this review but can be found elsewhere.[16,47,48,55]

Application of machine learning for analysis and interpretation of spectral data

On completion of spectral pre-processing, chemometrics must be applied in order to interpret any clinically-relevant variance within the data set. Chemometrics is a

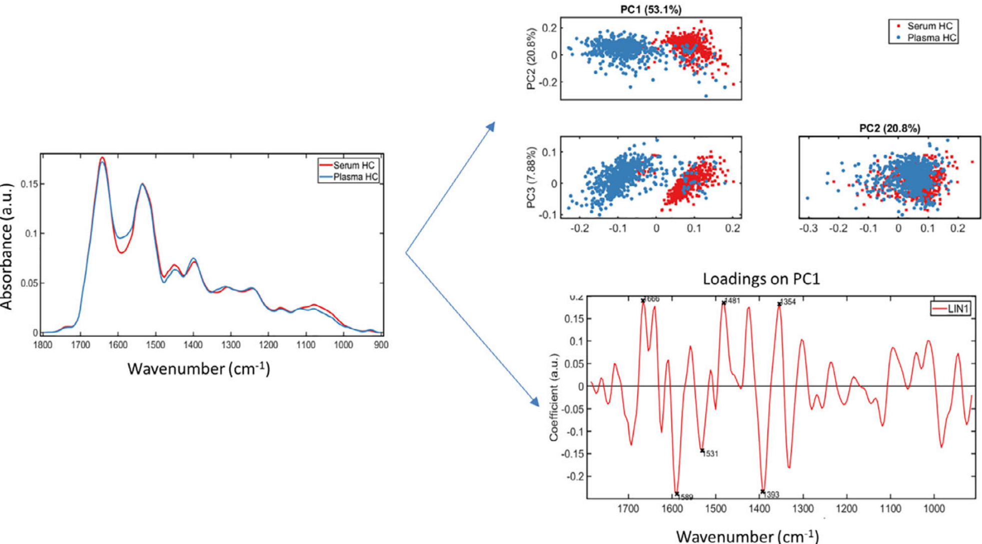


Figure 5. The application of PCA on ATR-FTIR spectral data to identify key variances responsible for class separation. Six hundred spectra (thirty healthy controls, 20 replicate point spectra) recorded on dried blood spots; either serum (red) or plasma (blue) and cut to the fingerprint region (1800–900 cm^{-1}). When PCA is applied to the dataset, scores plots can be obtained (top right) illustrating two discrete clusters (red and blue) from each sample type. This demonstrates that spectra from either serum or plasma can be clearly discriminated from one another, particularly when using Principle Component 1 (PC1) which contains 53.1% of the total variance. Further interpretation of discriminating wavenumbers within the dataset is possible by creating a loadings plot of PC1 (bottom right); by application of a 'peak detector' the wavenumbers associated with the greatest variance between the two sample types will be identified.

speciality within the field of Machine Learning (Figure 4). It applies mathematical and statistical methods to allow optimal analysis of multivariate data. Interrogation of spectral data is conducted over two phases; an exploratory analysis, during which the data set is interrogated for disease-associated patterns or biomarkers that may enable separation (clustering) of different classes (e.g., disease versus healthy controls), followed by diagnostic prediction analysis, whereby key spectral variances are evaluated using classifier-based mathematical algorithms. Diagnostic classification models must be built on a 'training dataset' using machine-learning algorithms, the success (correct classification rate) of the chosen model must then be verified on a new 'test' dataset to evaluate the diagnostic potential. Prior to translation into the clinical laboratory a number of refining steps can be conducted; this further training of the algorithm-based prediction system can be achieved through the blind testing of new unclassed data.[47]

During exploratory analysis, unsupervised chemometric methods can be used, which refer to the fact that class information (i.e., disease or non-disease) is omitted to prevent over-fitting of the data and introduction of bias. Principle component analysis (PCA) is a widely used unsupervised multivariate analysis technique for interpretation of complex data. The goal of PCA is to reduce the complexity of a dataset by producing a fewer number of independent variables (called loadings vectors), while retaining the maximum levels of variation present in the original data (Figure 5). In order to obtain the loading

vectors, commonly referred to as principle components (PCs), the original data is linearly transformed and ordered as such that the first PC retains the maximal variance and each succeeding component retains the next highest variance. Each successive PC is orthogonal to the others (i.e., forms an angle of 90 with the PCs). This enables the data to be visualized in a reduced dimensional space and any relationship between classes can be identified with less difficulty.[56,57] In the analysis of biological spectra, PCA can reduce the number of features from approximately 900 wavenumbers to 10–20 PCs. In biological samples, 10–20 PCs will often account for 99% of the total variance.[58] As an unsupervised multivariate method, PCA has weak discriminatory power, which limits its use in classification and prediction models. In the diagnostic classification phase, this can be overcome by combining PCA with more powerful supervised methods such as linear discriminant analysis (LDA). PCA is therefore applied as a further pre-processing step for several computational analysis techniques in machine learning and data mining algorithms.

Whilst unsupervised processing can reveal unique spectral features common to particular groups (or clusters) of samples this is often not sufficient for diagnostic application. Supervised data processing techniques are therefore required to build a robust diagnostic framework using machine-learning algorithms which can be applied to predict pathological disease.[59] In this case, a prediction model is built based on the assignment of subtle spectral differences from known classes (disease or non-disease) of samples within a training data set. The model is then verified using a new 'test' data set to establish the ability of the model to correctly classify unknown samples into disease- or non-disease groups.[16,60]

LDA is an example of a commonly used supervised chemometric technique. As opposed to unsupervised methods, supervised techniques include consideration of class information (i.e., disease or non-disease) during analysis. LDA, similar to PCA, is a linear transformation technique, both of which look for features in a reduced dimensional space to explain complex data variance. In contrast to PCA (which generates a loading vector associated with maximal variance), LDA generates a vector which will maximize between-class separation whilst minimizing within-class spreading of the data. This enables high accuracy as a prediction or classification model to be achieved if the data can be separated in a linear manner. A known limitation associated with LDA is the tendency to over-fit the data (given the fact that classes are considered during the analysis), particularly if insufficient number of spectra are acquired within the study. To overcome this, a general rule for application of LDA is to collect a number of spectra at least 5–10 times larger than the number of features. Application of PCA as a pre-processing step, thus reduction of the number of variables in the dataset, will also aid in overcoming this limitation.[48]

A variety of multivariate pattern recognition and classification techniques are available to facilitate interpretation of complex spectral data, the details of which are beyond the scope of this article but are covered in detail elsewhere.[47,55,61] Application of these chemometric tools during the exploratory analysis phase permits the extraction of biochemically-relevant spectral information and the development of classification models.[47] As described earlier, prior to the translation of these newly-constructed classification algorithms into clinically useful diagnostic tools, further validation of

classification performance is required; this is achieved by blind testing the classification model using spectra from a new 'test' population dataset. Once validated, the need for complex computational analysis is no longer required, permitting the use of biospectroscopy to become a 'plug-and-play' methodology either within the immunology laboratory or at the bedside/clinic consultation as a point-of-care device.

Clinical application of vibrational spectroscopy

Vibrational spectroscopy has been successfully applied across a wide area of clinical medicine; providing a new approach to detect molecular and structural changes caused by complex disorders such as Alzheimer's disease,[62,63] multiple sclerosis,[64] mental disorders,[65,66] HIV/AIDS,[67] diabetes[68] and carcinogenesis.[36,65,69–71] High diagnostic accuracy was demonstrated for classification of numerous cancer types and other biological applications.[15,18,33,48,52,72] Despite several promising proof-of-concept studies demonstrating impressive sensitivities and specificities for a wide range of pathologies, the translation of vibrational spectroscopy into the clinical setting has yet to gain momentum. This is likely due to a lack of clinical studies with the appropriate sample sizes to reflect the disease prevalence; thus most studies to date may be biased due to the use of relatively small sample sizes.[73] Furthermore, in order to fully validate this novel method for the detection of rare disease, further spectroscopy studies must be conducted alongside the current gold-standard laboratory techniques within large-scale screening programmes.[19]

Application of vibrational spectroscopy in clinical immunology

As highlighted earlier, immune dysregulation has interconnectivity with almost all specialities of medicine. There is an unmet need for an improved and holistic approach toward the investigation of immune-associated disorders, whereby simultaneous analysis of multiple immune-mediated components is performed in order to provide an earlier indication of disease. Following on from the completion of the 100,000 Genome Project and an ever increasing number of genomic sequencing studies, a multitude of immune-related loci and disease-associated alleles have been identified.[74–80] Unfortunately, for the majority of immunological disorders, the monogenetic link has not yet been identified, or quite possibly, does not exist. Research has been unable to explain these disorders using simple Mendelian logic and instead have found increasing numbers of 'risk' loci in immune-associated genes, this therefore limits the clinical utility of a genomic approach in routine diagnostics and an alternative strategy is required.

To do this, a combined use of 'omics'-like technologies and whole sample profiling may provide more clues in the search for diagnostic biomarkers and causative etiology. Through advances in molecular methods it is now understood that additional micro-environmental factors play a role in the varied clinical phenotypes observed within a disease entity. For example, in common variable immune deficiency (CVID), the most common symptomatic immune deficiency in adults, epigenetic alterations, such as DNA methylation[81] and histone modifications[82]; as well as transcriptional disturbances, such as single nucleotide polymorphisms (SNPs) impact on CVID susceptibility.[83,84] In

light of this, novel approaches to complement current molecular findings are warranted, not only to shed a new light on the pathogenic mechanisms associated with the disease, but to enable risk stratification and individualization of treatment. Toward this goal, the interrogation of whole unprocessed samples with vibrational spectroscopy will generate a new library of 'big data' at the molecular level, which can be examined in the context of complex disorders such as CVID.

Autoimmune disease

Autoimmunity is a term used to describe an immune response which is directed against 'self', a process famously coined 'horror autotoxicus' by Dr Paul Ehrlich. The etiology of autoimmune disease has yet to be fully elucidated but is considered to be multi-factorial, involving both genetic and environmental influences to ultimately result in a loss of tolerance to self-proteins of the body.

Both Raman and FTIR vibrational spectroscopy methods have been applied to the investigation of rheumatological conditions in several proof-of-concept studies. As early as 2003, the diagnostic and disease monitoring potential in rheumatoid arthritis (RA) was demonstrated by Canvin et al. using near infrared spectroscopy on specific joints alongside PCA-LDA multivariate analysis.[85] Through the identification of spectral features that correlated with metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joint damage, swelling and tenderness, they were able to differentiate between normal and RA patients with greater than 70% sensitivity and 80% specificity. Serological analysis followed later, when in 2011 Carvalho et al.[86] used Raman spectroscopy of blood sera and PCA-LDA for diagnostic discrimination of RA patients from controls to achieve even greater sensitivity and specificity, at 88% and 96%, respectively. Similar serological findings were obtained in 2016 by Lechowicz et al.[87] using FTIR and a K nearest neighbor (K-NN) mathematical modeling to achieve 85% sensitivity and 100% specificity in the discrimination of RA patients from controls. These superior sensitivities, compared to the current serological laboratory tests (rheumatoid factor (54% sensitivity) and c-reactive protein (58% sensitivity),[86] would therefore position vibrational spectroscopy as a strong candidate for a new screening test for RA in the clinical setting. Moreover both groups illustrated the capacity to detect early and late changes in patients with rheumatoid synovitis; thus if used in clinical practice would provide a means to identify and treat the right patients at the right time.

In 2019, a further vibrational spectroscopy study in the field of rheumatology was undertaken by Hackshaw et al.[88] Using FTIR and Raman spectroscopy of dried blood spots, followed by SIMCA® multivariate analysis they were able to successfully discriminate between patients with RA, systemic lupus erythematosus (SLE), osteoarthritis and fibromyalgia with no misclassifications. The unique spectral signatures derived from the study were capable of separating out the four patient groups due to vibrational variations in the protein and nucleic acid backbone.[88]

Raman spectroscopy of serum and plasma samples has recently been successfully applied to anti-neutrophil cytoplasmic antibody (ANCA) vasculitis,[89] a multisystem autoimmune disease in which inflammation and damage to small blood vessels occurs during active disease. Left untreated AAV can rapidly progress leading to organ

damage and death. Morbidity and mortality has improved dramatically following the use of immunosuppressive therapy to induce remission, however prompt identification of relapsing disease remains a diagnostic difficulty for clinicians due to a lack of reliable disease activity biomarkers.[90,91] In their study, Morris et al. illustrated for the first time, the use of Raman spectroscopy as a candidate biomarker for active disease in AAV. Using partial least squares discriminant analysis (PLS-DA) they achieve discrimination between patients with active disease and remission with 70% and 80% sensitivity, and 93% and 80% specificity, for serum and plasma, respectively.[89] Subsequently, the group applied ATR-FTIR and PLS-DA to distinguish between active and quiescent disease, which resulted in superior sensitivity and specificity, achieving 100% and 86%, respectively.[92] These findings have the potential to aid patient care and improve clinical outcomes through the earlier detection and management of relapsing disease.

ATR-FTIR has also been used to evaluate inflammatory glomerulonephritis (GN) using urine as an alternative biofluid. Yu et al. used a mouse model of progressive crescentic GN induced by anti-glomerular basement membrane antibodies to identify key spectral biomarkers.[93] Specifically, they observed an increased intensity at wavenumber $1545\,cm^{-1}$ with increased severity of disease, which subsequently decreased with response to corticosteroid treatment. This biomarker was also diagnostically applicable to the human cohort of 24 ANCA vasculitis patients included in the study. These findings have demonstrated the future potential for urine biospectroscopy to be used as a sensitive, cost-effective, and noninvasive test for diagnosis and monitoring of renal patients in the clinic.

FTIR and multivariate analysis techniques have been applied to CSF samples in the diagnosis multiple sclerosis (MS).[64] Immune-mediated destruction of the myelin sheath surrounding nerve fibers of the central nervous system characterizes the pathogenesis of MS, however the underlying cause remains unclear, as is the cause of disease progression from a clinically isolated incident (CIS) to relapsing-remitting MS (RRMS) and, of increased severity, to progressive MS. Despite availability of diagnostic criteria,[94] the diagnosis of MS is rendered difficult due to the overlapping clinical and analytical features shared with other neurodegenerative diseases. ATR-FTIR spectroscopy of CSF coupled with multivariate analysis (PCA and hierarchical cluster analysis) has proven a successful novel technique in MS capable of diagnosing and differentiating CIS and RRMS patients from control samples. Furthermore, they identified a unique disease-associated spectral biomarker at wavenumber $795\,cm^{-1}$, which could provide a laboratory tool to accurately diagnose MS. Their findings suggest that vibrational spectroscopy and multivariate analysis has high potential for use in the early detection of MS, and that the spectral changes identified may indicate disease-induced changes at an earlier subclinical time point.[64] This could have great clinical impact on patient outcome, as previous studies indicate that the risk of developing MS can be reduced by early treatment after the first clinical demyelinating event.[95]

Allergic and hypersensitivity disorders

In the field of clinical hypersensitivity, which includes allergy, asthma and atopic dermatitis, the application of vibrational spectroscopy has not been widely explored. However over the last decade both Raman and FTIR spectroscopy methods have started to

become increasingly recognized as a powerful new approach for the clinical investigation of allergy patients. Allergy can be defined as an exaggerated immune response (hypersensitivity) to a foreign substance that is usually harmless in non-allergic individuals. Currently, the diagnosis and management of severe allergic and hypersensitivity disorders requires referral to specialist services for clinical consultation in order to obtain a thorough allergy-focused history. In the clinical immunology laboratory, serological testing is limited to the detection of specific IgE (sIgE) directed against suspected allergens, and to the measurement of mast cell tryptase, which is released from mast cells during severe acute allergic reactions (anaphylaxis). There are several limitations associated with these tests and an unmet need to improve patient care,[96] toward which vibrational spectroscopy could offer a novel diagnostic platform, either as a laboratory-based assay or as a point-of-care device.

It should be noted that the detection of sIgE is not diagnostic for allergy, rather it indicates that a patient has been sensitized to that allergen. There is a need for an improved approach given the lack of standardization between these assays, and the fact that clinical sensitivities have not been well established, being particularly poor in the investigation of drug allergy (e.g., 19.3% for penicillin.[97] Whilst mast cell tryptase measurement can be useful in the investigation of anaphylaxis, full interpretation requires serial sample collection (immediately, within four hours and a baseline measurement 24 hours post-reaction. As the majority of patients present to the emergency department, a complete set of serial samples are rarely received, leaving patients with potentially severe allergy at risk of being missed and appropriately managed.

Biochemical mediators associated with an allergic phenotype could be detected through unique spectral variances present in biological samples. This would offer a new approach for allergy diagnosis in the clinical setting; improving patient care through eliminating the need for serial sample collection, increasing diagnostic sensitivities for allergy testing and alleviating NHS financial pressures associated with hospital-based patient management. As more is uncovered about the biochemical changes associated with an allergic phenotype (Th2-driven inflammation involving IL-4, IL-5, IL-13) and the response to immunotherapy (decreased Th2 response alongside increased immune regulation markers IL-10, TGF-β and T regulatory cells),[98] gaining an overview of the whole metabolic picture in a single test will be more informative in the assessment of allergic responses. As the spectral fingerprint is representative of all the biochemical constituents in a biological sample, vibrational spectroscopy paired with multivariate analysis provides the means to do this in a robust and economical way. In Asthma patients, assignment of disease severity category (mild, moderate, severe, very severe) is vital for risk stratification and provision of optimal treatment. Current assessment is largely based on clinical parameters however in 2013 Sahu et al. demonstrated promising results using Raman spectroscopy and PCA-LDA on serum as a novel laboratory test.[99] Spectral changes observed in the 44 asthma patients of varying disease severities enabled separation of all asthma patients from a reference group. Of note, an increase in Raman bands assigned to DNA and glycosaminoglycans were associated with increased asthma severity. Moreover, good separation of treated versus untreated severe asthma patients was achieved, indicating that treatment response could be monitored if used in future clinical practice.

Improvements in the treatment of asthma and chronic airway disease are also being influenced by successful research using FTIR spectroscopy.[100] Through the identification of unique spectral features within chronic allergic airway models, biochemical changes induced by anti-asthmatic therapies can be studied in a way that has not been possible before. The need for novel asthma therapeutics is high given the limitations associated with current treatments which include bronchial dilators and glucocorticoids. Whilst a number of promising new treatments have demonstrated clinical benefit using conventional methods,[101–103] FTIR spectroscopy may provide the tool that will expedite their use in clinical practice, through overcoming the challenge of elucidating the therapeutic mechanism of action at a detailed molecular level.

Research into the allergenicity of aeroallergens, specifically pollen, has been undertaken using both Raman and FTIR.[104,105] In 2014, Guedes et al.[105] use Raman spectroscopy to create a spectral library of 34 pollen species, demonstrating an innovative method to characterize airborne pollen. In an era in which the prevalence of respiratory allergies continues to rise worldwide,[106] this approach may provide a future aid to patient management, particularly during the pollen seasons. Subsequent to this, in 2017, Depciuch et al.[104] used FTIR spectroscopy to demonstrate that molecular alterations in hazel pollen due to the influence of air pollutants and urbanization may result in an increased allergenicity. These findings provide new evidence to further understand the epidemiology of allergic disease, and will enable clinicians to tailor future treatment and prevention strategies accordingly.

The *in vivo* application of Raman and FTIR spectroscopy to the study of skin barrier function and molecular composition has been demonstrated by several groups.[107–111] In the context of atopic dermatitis, vibrational spectroscopy may provide a novel, noninvasive assessment tool to monitor both the safety and clinical efficacy of new treatments. This is currently being explored in a clinical trial at Sheffield University (NCT04194814), in which FTIR spectroscopy is being used as a potential outcome measure for evaluating the safety of topical treatments in atopic dermatitis.[112] In addition to assessment of treatment response, the group aim to identify novel biomarkers of skin barrier disruption and atrophy that can be applied to future therapeutic drug trials to assess safety.

In the most recent piece of allergy-focused biospectroscopy research, Korb et al.[113] used machine learning-empowered FTIR spectroscopy to successfully discriminate between healthy, allergic and immunotherapy-treated allergy patients using both a mouse model and human patients. This study, conducted in 2020 demonstrated the clinical potential of using serum samples and vibrational spectroscopy for the rapid diagnosis of allergy and therapeutic monitoring of patients treated with allergen-specific immunotherapy (SIT). Building on the success of the studies to date, further larger-scale patient studies using FTIR and Raman spectroscopy methods are warranted in order to fully understand the potential impact this platform could have in the field of allergy and hypersensitivity disorders.

Immune deficiency

Immune deficiency disorders are a heterogeneous group of disorders often presenting with an increased susceptibility to infection due to defective immune system

development and/or function. These disorders can be divided into primary immunodeficiency disorders (PIDs), also known as inborn errors of immunity, and secondary immunodeficiency disorders, which are acquired as a result of disease or environmental factors. In the most recent 2019 update to the International Union of Immunological Societies (IUIS) classification of Inborn Errors of Immunity report, 430 distinct PIDs were characterized; a significant increase from the 180 previously report in 2009.[114,115] The increased availability and accessibility of molecular techniques such as next-generation sequencing has expedited the identification of novel genetic defects associated with PIDs, and led to greater understanding of immunological mechanisms of disease. Further contribution to the understanding of molecular and cellular pathways related to PIDs may soon be provided from novel platforms, such as the use of vibrational spectroscopy, which can provide detailed molecular information on a wide range of biofluids.

The laboratory investigation of immunodeficiency disorders consist of an initial work-up, including complete blood counts, lymphocyte subsets (T-, B-, and NK-cells), and serological measurement of immunoglobulin levels, vaccine titers and complement. Functional studies, cytokine assays and molecular tests can be subsequently performed to aid final diagnosis. Flow cytometry is used to determine the identity of the cells of the immune system, a method that requires using fluorescently-labelled antibodies that can potentially alter the behavior of the cells, or may also require chemical fixation, which destroys the cell so that further functional assays cannot be performed. The potential use of Raman and FTIR spectroscopy to provide a label-free method for cellular studies is being currently being explored, as this would enable further use and manipulation of identified immune cells that has previously been a challenge with flow cytometry.

In the first and only clinical application of vibrational spectroscopy to PID, Callery et al.[116] applied ATR-FTIR to the study of patients with Common Variable Immune deficiency (CVID). CVID is a clinically heterogeneous disorder characterized by hypogammaglobulinemia and recurrent infections with further complications such as autoimmunity and malignancy affecting approximately 1 in 5 patients.[117,118] Currently, there is no diagnostic clinical or laboratory test available for CVID. Using ATR-FTIR alongside PCA-SVM, the group demonstrated the ability to classify patients with Common Variable Immune deficiency (CVID) from healthy controls with sensitivities and specificities of 97% and 93%, respectively for serum, and 94% and 95%, respectively for plasma. Furthermore, using a combination of chemometric algorithms such as PCA-LDA, Student's T Test and Forward feature selection (FFS), several discriminating spectral bands were identified, including wavenumbers in regions indicative of nucleic acids (984 cm^{-1}, 1053 cm^{-1}, 1084 cm^{-1}, 1115 cm^{-1}, 1528 cm^{-1}, 1639 cm^{-1}), and a collagen-associated biomarker (1528 cm^{-1}), which may represent new candidate biomarkers in future diagnostics.

A number of groups have demonstrated the proof-of-concept for the Raman–based methods to classify immune cells in a label-free manner; however, as expected due to the similarities in shape, size and biochemical makeup of lymphocyte subsets, there are several areas of spectral overlap observed when characterizing immune cells. In 2014, Maguire et al.[119] successfully classified lymphocytes and monocytes from the blood of

a population of volunteers using Raman microspectroscopy and three different chemometric techniques. Their work highlighted the importance of understanding and applying the most appropriate multivariate analysis and modeling techniques to maximize robustness and accuracy.

Subsequently, in 2015, Chen et al.[120] published their research which demonstrated, for the first time, the ability of wavelength modulated Raman spectroscopy to identify unfixed and unperturbed lymphocyte populations from the peripheral blood of multiple healthy donors. Wavelength modulated Raman spectroscopy was used to enhance the standard Raman technique and allowed discrimination of closely related lymphocyte subsets, CD4+ T cells, CD8+ T cells and CD56+ NK cells which had previously not been achieved. This pioneering work provides significance for both in vitro and in vivo studies of the immune system, and as the cells are unaltered, Raman-identified cells could be subsequently analyzed for functionality, cytokine profiling, or transcriptome signatures.[120]

Hobro et al.[121] used Raman spectroscopy alongside PLS-DA to question how significant the spectral variations are between T cells and B cells, as well as characterizing key wavenumbers associated with different cell lines of B- and T lymphocytes. Whilst demonstrating the possibility of this approach to distinguish between single point-spectra of B- and T cells, and between individual B- and T cell lines, the misclassification rate was high at 75 in 1000. This highlights the need for further optimization studies prior to translation into the diagnostic laboratory.

Lastly, using an alternative multi-modal approach with the use of both Raman spectroscopy and digital holographic microscopy (DHM), Reynolds et al.[122] were able to characterize individual cells of the immune system through acquisition of single-point spectra. They demonstrated the ability to identify and discriminate between CD4+ T cells, B cells and monocytes at sensitivities in the range of 87–100% and specificities of 85–100%. Given the more rapid analysis capabilities of DHM but lower specificity than Raman spectroscopy, they propose using the two label-free techniques together for a more robust analysis. Translated clinically, DHM could be used as a fast initial screening test in which cells of interest can be further interrogated by Raman spectroscopy for a more detailed molecular analysis.

In summary, Raman spectroscopy is not currently in a position to replace flow cytometry and genetic analysis in the laboratory investigation of immune deficiency disorders. However, given the detailed biomolecular information obtained from spectral analysis of immune cells, it would provide a complementary methodology and molecular insight into the complex pathological mechanisms underpinning immunodeficiency disorders. Future research studies are required to fully unleash the diagnostic potential of vibrational spectroscopy in the cellular immunology laboratory.

Advantages and challenges of vibrational spectroscopy in biological investigations

Both IR and Raman spectroscopy have been extensively applied to investigate biological fluids for the detection of disease.[123] The majority of research has focused on blood serum and plasma, however saliva,[60] tears,[124] bile fluid,[125] urine[126] and cerebrospinal fluid[64] have also been used. FTIR spectroscopy offers many advantages over

conventional methods for laboratory analysis of pathological specimens. As previously described, almost all of the molecules present in a biological sample are IR-active, therefore can be quantitatively and qualitatively interrogated by IR spectroscopy.[60] Compared to using enzyme-based assays (such as ELISA) or fluorescently-tagged monoclonal antibody methods to probe particular analytes in a sample, IR-spectroscopy techniques can perform direct measurement on unprocessed samples and require very minimal sample preparation. Moreover, spectroscopy methods do not require expensive reagents or labeling components and permit the analysis of the entire sample constituents rather than a discrete number of known analytes.[60] Further instrumental advances in IR spectroscopy have led to the generation of increasingly sensitive, cost-effective, portable devices that are well-suited for point-of-care (POC) tools. Prototypic POC devices have already been built and trialed to accurately determine clinical parameters in an intensive care unit.[127]

Limitations facing IR spectroscopy for the quantification of biomolecules in clinical practice are largely related to a lack of analytical sensitivity and specificity, the latter is often due to the complexity of biological matrices, and can be improved to an appropriate level with the application of chemometric techniques for multivariate analysis. Sensitivity issues have been described as intrinsic to the IR technique which are further compounded by the strong IR absorption of water molecules present in biofluid samples. In light of this, only ATR-FTIR is deemed acceptable for interrogating unprocessed fluids such as whole blood; pre-processing steps such as drying the biofluids onto slides or directly into the ATR crystal are also well-practiced methods to improve quality of the spectral data.[60] Sample preparation can impact the biochemistry and structural properties of a sample therefore must be carefully considered prior to initiating vibrational spectroscopy studies. A well-described spectral sampling phenomenon known as the 'coffee-ring effect' can impact the homogeneity of a dried blood film as numerous biochemical components (proteins, lipids and nucleic acids) migrate toward the periphery of the blood spot as it dries.[128,129] This phenomenon has been shown to have a greater impact in diluted samples compared to non-diluted serum samples, and can be further avoided by the use of reduced size blood spots.[130] The use of robotic microfluidics for standardization of sample loading alongside vacuum drying techniques has produced spectral data with higher reproducibility than non-automated processes; these developments will aid the future implementation of a robust, high-throughput analytical platform required for large-scale laboratory investigations.[131]

Optimum sample preparation for FTIR spectroscopy on blood serum has been recently investigated to support the translation of IR spectroscopy to the clinical setting. Repeatability and reproducibility for both ATR- and high throughput-FTIR has been demonstrated, achieving spectral sample deviations of 0.0015 for both methods, and the ability to discriminate similar spectral variances in a patient study.[130] The importance of sample preparation methods for FTIR spectroscopic analysis of tissues have also been described following examination of spectral variance following different preservation techniques (drying, ethanol or formalin fixation).[132] The use of freshly excised tissue was deemed the optimum approach in order to avoid perturbing the biochemical composition of the tissue; nevertheless, the conduction of FTIR spectroscopy studies requiring sample preservation remain valuable provided consideration is given to the spectral alterations that may arise.[132]

Conclusion

Vibrational biospectroscopy is a tool that is being increasingly applied in the field of diagnostic medicine. The potential applications of FTIR and Raman spectroscopy have advanced in recent years due to the developments made in both instrumentation and crucially bioinformatics. This review describes its application to clinical immunology, highlighting the key findings from research studies across the major disorders of the immune system; autoimmunity, immune deficiency and allergy. These immunological disorders are complex, often demonstrating interaction across multiple molecular pathways. As increased knowledge of these signaling pathways expand, the use of targeted immunotherapies has become routinely available. However, advances in molecular medicine are not being translated into laboratory diagnostics. The limitations facing current laboratory methods, which detect only single or few discrete biomarkers are becoming increasingly problematic, highlighting the need for a new approach. The potential for vibrational spectroscopy to fill this gap and be translated into the clinical setting has been explored and promoted throughout this review. However, as widely discussed throughout the literature,[39,133,134] there remain a number of barriers to overcome before these methods become routinely applied. To accomplish this, increasing collaborations between partners based within a clinical setting and academia is vital, with particular emphasis placed upon the implementation of robust quality control processes and high-throughput automation.

In summary, the proof-of-concept for applying vibrational spectroscopy within the clinical immunology setting has proven successful across a number of immune-mediated disorders. However wide-scale longitudinal research studies are now required to fully unleash the diagnostic potential of vibrational spectroscopy in the immunology laboratory. As increasing interest in the field of biospectroscopy continues, it is hoped that consensus- and best-practice protocols will be published, further promoting the standardized approach required for the clinical setting. If successful, this methodology could prove transformative within healthcare, improving patient diagnosis, prognosis and treatment monitoring.

Authors' contribution

Emma L. Callery contributed to the design of the article, reviewed and interpreted the literature and wrote the manuscript.

Anthony W. Rowbottom contributed to the conception of the review, critically reviewed the manuscript and provided intellectual discussions.

Disclosure statement

The authors declare no competing interests.

ORCID

Emma L. Callery http://orcid.org/0000-0003-0786-9182

References

1. Naumann, D. Vibrational Spectroscopy in Microbiology and Medical Diagnostics. In *Biomedical Vibrational Spectroscopy* ; Lasch, P., Kneipp, J., Eds.; John Wiley & Sons, Inc.: Hoboken, N.J, United States 2008; pp 1–8.
2. Weatherall, D.; Greenwood, B.; Chee, H. L.; Wasi, P.; et al. Science and Technology for Disease Control: Past, Present, and Future. In *Disease Control Priorities in Developing Countries*; Jamison, D. T. eds.; World Bank: Washington DC, United States, 2006.
3. Kumar, V., Abbas, A. K., Fausto, N., Abbas, F. K., Robbins, S. L. *Robbins and Cotran Pathologic Basis of Disease*. Elsevier Saunders: Philadelphia, United States. 2005.
4. Kaufmann, S. H. E. Immunology's Coming of Age. *Front. Immunol.* 2019, *10*, 1–10. doi: 10.3389/fimmu.2019.00684
5. Li, H.; Li, J.; Zhao, G.; Wang, Y. Current Research Status of Immunology in the Genomic Era. *Sci. China C Life Sci.* 2009, *52*, 43–49. doi:10.1007/s11427-009-0006-7
6. The Nobel Prize in Physiology or Medicine. 1984. *NobelPrize.org* https://www.nobelprize.org/prizes/medicine/1984/summary/.
7. The Nobel Prize in Physiology or Medicine. 2018. *NobelPrize.org* https://www.nobelprize.org/prizes/medicine/2018/press-release/.
8. Tebo, A. E.; Detrick, B.; Hamilton, R. G.; Khanolkar, A.; O'Gorman, M. R. G.; Schmitz, J. L.; Abraham, R. S. Clinical Laboratory Immunology: An Indispensable Player in Laboratory Medicine. *Am. J. Clin. Pathol.* 2014, *142*, 437–444. doi:10.1309/AJCPX25MFWNEYRIG
9. Rapin, N.; Lund, O.; Bernaschi, M.; Castiglione, F. Computational Immunology Meets Bioinformatics: The Use of Prediction Tools for Molecular Binding in the Simulation of the Immune System. *PLOS One.* 2010, *5*, e9862. doi:10.1371/journal.pone.0009862
10. Lundegaard, C.; Lund, O.; Kesmir, C.; Brunak, S.; Nielsen, M. Modeling the Adaptive Immune System: Predictions and Simulations. *Bioinformatics* 2007, *23*, 3265–3275. doi:10.1093/bioinformatics/btm471
11. Pappalardo, F.; Russo, G.; Reche, P. A. Toward Computational Modelling on Immune System Function. *BMC Bioinform.* 2020, *21*, 546. doi:10.1186/s12859-020-03897-5
12. Pernice, S.; Follia, L.; Maglione, A.; Pennisi, M.; Pappalardo, F.; Novelli, F.; Clerico, M.; Beccuti, M.; Cordero, F.; Rolla, S. Computational Modeling of the Immune Response in Multiple Sclerosis Using Epimod Framework. *BMC Bioinform.* 2020, *21*, 550. doi:10.1186/s12859-020-03823-9
13. Gomez-Perosanz, M.; Ras-Carmona, A.; Lafuente, E. M.; Reche, P. A. Identification of CD8+ T Cell Epitopes through Proteasome Cleavage Site Predictions. *BMC Bioinform.* 2020, *21*, 484. doi:10.1186/s12859-020-03782-1
14. Russo, G.; Pennisi, M.; Fichera, E.; Motta, S.; Raciti, G.; Viceconti, M.; Pappalardo, F. In Silico Trial to Test COVID-19 Candidate Vaccines: A Case Study with UISS Platform. *BMC Bioinform.* 2020, *21*, 527. doi:10.1186/s12859-020-03872-0
15. Baker, M. J.; Hussain, S. R.; Lovergne, L.; Untereiner, V.; Hughes, C.; Lukaszewski, R. A.; Thiéfin, G.; Sockalingum, G. D. Developing and Understanding Biofluid Vibrational Spectroscopy: A Critical Review. *Chem. Soc. Rev.* 2016, *45*, 1803–1818. doi:10.1039/c5cs00585j
16. Baker, M. J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H. J.; Dorling, K. M.; Fielden, P. R.; Fogarty, S. W.; Fullwood, N. J.; Heys, K. A.; et al. Using Fourier Transform IR Spectroscopy to Analyze Biological Materials. *Nat. Protoc.* 2014, *9*, 1771–1791. doi:10.1038/nprot.2014.110
17. Movasaghi, Z.; Rehman, S.; Rehman, D. I. u. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev.* 2008, *43*, 134–179. doi:10.1080/05704920701829043
18. Talari, A. C. S.; Martinez, M. A. G.; Movasaghi, Z.; Rehman, S.; Rehman, I. U. Advances in Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev.* 2017, *52*, 456–506. doi:10.1080/05704928.2016.1230863

19. Bunaciu, A. A.; Aboul-Enein, H. Y.; Fleschin, Ş. Vibrational Spectroscopy in Clinical Analysis. *Appl. Spectrosc. Rev.* 2015, *50*, 176–191. doi:10.1080/05704928.2014.955582
20. Herzberg, G. *Infrared and Raman Spectra of Polyatomic Molecules*. Vol. II. Van Nostrand: New York, United States; 1945.
21. Martin, F. L. Shining a New Light into Molecular Workings. *Nat. Methods*. 2011, *8*, 385–387. doi:10.1038/nmeth.1594
22. Kendall, C.; Isabelle, M.; Bazant-Hegemark, F.; Hutchings, J.; Orr, L.; Babrah, J.; Baker, R.; Stone, N. Vibrational Spectroscopy: A Clinical Tool for Cancer Diagnostics. *Analyst*. 2009, *134*, 1029–1045. doi:10.1039/b822130h
23. Quaroni, L.; Zlateva, T. Infrared Spectromicroscopy of Biochemistry in Functional Single Cells. *Analyst*. 2011, *136*, 3219–3232. doi:10.1039/c1an15060j
24. Hof, M.; Macháň, R. Basics of Optical Spectroscopy. In *Handbook of Spectroscopy*; John Wiley & Sons, Ltd: Weinheim, Germany 2014; pp 31–38. doi:10.1002/9783527654703.ch3
25. Lacombe, C.; Untereiner, V.; Gobinet, C.; Zater, M.; Sockalingum, G. D.; Garnotel, R. Rapid Screening of Classic Galactosemia Patients: A Proof-of-Concept Study Using High-Throughput FTIR Analysis of Plasma. *Analyst*. 2015, *140*, 2280–2286. doi:10.1039/c4an01942c
26. Zeng, H.; McWilliams, A.; Lam, S. Optical Spectroscopy and Imaging for Early Lung Cancer Detection: A Review. *Photodiagnosis Photodyn. Ther.* 2004, *1*, 111–122. doi:10.1016/S1572-1000(04)00042-0
27. Ashwood, E. R.; Aldrich, J. E.; Tietz, N. W. *Tietz Fundamentals of Clinical Chemistry*; W.B. Saunders: Philadelphia, United States, 1996.
28. Banwell, C.; Mccash, E. *Fundamentals for Molecular Spectroscopy*; McGraw-Hill Education: London, UK, 1994.
29. Lewis, I. R.; Edwards, H. *Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line*; CRC Press: New York, United States, 2001.
30. McGregor, H.; Wang, W.; Short, M.; Zeng, H. Clinical Utility of Raman Spectroscopy: Current Applications and Ongoing Developments. *Adv. Health Care Technol.* 2016, *13*, 1–13. doi:10.2147/AHCT.S96486
31. Byrne, H. J.; Sockalingum, G. D.; Stone, N. Chapter 4: Raman Microscopy: Complement or Competitor? In *Biomedical Applications of Synchrotron Infrared Microspectroscopy: A Practical Approach, The Royal Society of Chemistry, Cambridge, UK*, 2010, pp 105–143.
32. McGregor, H. C.; Short, M. A.; McWilliams, A.; Shaipanich, T.; Ionescu, D. N.; Zhao, J.; Wang, W.; Chen, G.; Lam, S.; Zeng, H.; et al. Real-Time Endoscopic Raman Spectroscopy for In Vivo Early Lung Cancer Detection. *J. Biophotonics*. 2017, *10*, 98–110. doi:10.1002/jbio.201500204
33. Hands, J. R.; Clemens, G.; Stables, R.; Ashton, K.; Brodbelt, A.; Davis, C.; Dawson, T. P.; Jenkinson, M. D.; Lea, R. W.; Walker, C.; et al. Brain Tumour Differentiation: Rapid Stratified Serum Diagnostics via Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy. *J. Neurooncol.* 2016, *127*, 463–472. doi:10.1007/s11060-016-2060-x
34. Barth, A. Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta*. 2007, *1767*, 1073–1101. doi:10.1016/j.bbabio.2007.06.004
35. Rehman, I. u.; Movasaghi, Z.; Rehman, S. *Vibrational Spectroscopy for Tissue Analysis*; CRC Press:*Boca Raton*, Florida, United States, 2012.
36. Bunaciu, A. A.; Hoang, V. D.; Aboul-Enein, H. Y. Applications of FT-IR Spectrophotometry in Cancer Diagnostics. *Crit. Rev. Anal. Chem*. 2015, *45*, 156–165. doi:10.1080/10408347.2014.904733
37. Shai, Y. ATR-FTIR Studies in Pore Forming and Membrane Induced Fusion Peptides. *Biochim. Biophys. Acta*. 2013, *1828*, 2306–2313. doi:10.1016/j.bbamem.2012.11.027
38. Tatulian, S. A. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy: A Method of Choice for Studying Membrane Proteins and Lipids. *Biochemistry*. 2003, *42*, 11898–11907. doi:10.1021/bi034235+
39. Cameron, J. M.; Butler, H. J.; Anderson, D. J.; Christie, L.; Confield, L.; Spalding, K. E.; Finlayson, D.; Murray, S.; Panni, Z.; Rinaldi, C.; et al. Exploring Pre-Analytical Factors for

the Optimisation of Serum Diagnostics: Progressing the Clinical Utility of ATR-FTIR Spectroscopy. *Vib. Spectrosc.* 2020, *109*, 103092. doi:10.1016/j.vibspec.2020.103092

40. Nawrocka, A. & Lamorska, J. *Chapter 14: Determination of Food Quality by Using Spectroscopic Methods in Advances in Agrophysical Research*. 2013. InTech, Rijeka, Croatia
41. Housecroft, C. E. *Chemistry [Electronic Resource]*; Prentice Hall/Pearson Education: Harlow, Essex, UK, 2010.
42. Old, O. J.; Fullwood, L. M.; Scott, R.; Lloyd, G. R.; Almond, L. M.; Shepherd, N. A.; Stone, N.; Barr, H.; Kendall, C. Vibrational Spectroscopy for Cancer Diagnostics. *Anal. Methods.* 2014, *6*, 3901–3917. doi:10.1039/c3ay42235f
43. Ganim, Z.; Chung, H. S.; Smith, A. W.; Deflores, L. P.; Jones, K. C.; Tokmakoff, A. Amide I Two-Dimensional Infrared Spectroscopy of Proteins. *Acc. Chem. Res*. 2008, *41*, 432–441. doi:10.1021/ar700188n
44. Stuart, B.; Ando, D. J. *Biological Applications of Infrared Spectroscopy*; Published on behalf of ACOL (University of Greenwich) by John Wiley: Chichester, New York, United States, 1997.
45. Talari, A. C. S.; Movasaghi, Z.; Rehman, S.; Rehman, I. u. Raman Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev*. 2015, *50*, 46–111. doi:10.1080/05704928.2014.923902
46. ISO 15189:2012 Medical Laboratories - Requirements for Quality and Competence - Westgard. https://www.westgard.com/iso-15189-2012-requirements-1.htm.
47. Trevisan, J.; Angelov, P. P.; Carmichael, P. L.; Scott, A. D.; Martin, F. L. Extracting Biological Information with Computational Analysis of Fourier-Transform Infrared (FTIR) Biospectroscopy Datasets: Current Practices to Future Perspectives. *Analyst.* 2012, *137*, 3202–3215. doi:10.1039/c2an16300d
48. Martin, F. L.; Kelly, J. G.; Llabjani, V.; Martin-Hirsch, P. L.; Patel, I. I.; Trevisan, J.; Fullwood, N. J.; Walsh, M. J. Distinguishing Cell Types or Populations Based on the Computational Analysis of Their Infrared Spectra. *Nat. Protoc*. 2010, *5*, 1748–1760. doi:10.1038/nprot.2010.133
49. Kelly, J. G.; Trevisan, J.; Scott, A. D.; Carmichael, P. L.; Pollock, H. M.; Martin-Hirsch, P. L.; Martin, F. L. Biospectroscopy to Metabolically Profile Biomolecular Structure: A Multistage Approach Linking Computational Analysis with Biomarkers. *J. Proteome Res.* 2011, *10*, 1437–1448. doi:10.1021/pr101067u
50. Lucassen, G. W.; Veen, G. N. A. V.; Jansen, J. A. J. Band Analysis of Hydrated Human Skin Stratum Corneum Attenuated Total Reflectance Fourier Transform Infrared Spectra In Vivo. *J. Biomed. Opt*. 1998, *3*, 267–281. doi:10.1117/1.429890
51. Mostaço-Guidolin, L. B.; Murakami, L. S.; Batistuti, M. R.; Nomizo, A.; Bachmann, L. Molecular and Chemical Characterization by Fourier Transform Infrared Spectroscopy of Human Breast Cancer Cells with Estrogen Receptor Expressed and Not Expressed. *J. Spectro*. 2010, *24*, 501–510. doi:10.3233/SPE-2010-0466
52. Khanmohammadi, M.; Garmarudi, A. B. Infrared Spectroscopy Provides a Green Analytical Chemistry Tool for Direct Diagnosis of Cancer. *TrAC Trends Anal. Chem.* 2011, *30*, 864–874. doi:10.1016/j.trac.2011.02.009
53. Wong, P. T. T.; Papavassiliou, E. D.; Rigas, B. Phosphodiester Stretching Bands in the Infrared Spectra of Human Tissues and Cultured Cells. *Appl. Spectrosc*. 1991, *45*, 1563–1567. doi:10.1366/0003702914335580
54. Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 1964, *36*, 1627–1639. doi:10.1021/ac60214a047
55. Morais, C. L. M.; Lima, K. M. G.; Singh, M.; Martin, F. L. Tutorial: Multivariate Classification for Vibrational Spectroscopy in Biological Samples. *Nat. Protoc*. 2020, *15*, 2143–2162. doi:10.1038/s41596-020-0322-8
56. Bro, R.; Smilde, A. K. Principal Component Analysis. *Anal. Methods.* 2014, *6*, 2812–2831. doi:10.1039/C3AY41907J
57. Sablinskas, V., Steiner, G. & Hof, M. Chapter 6: Applications. In Handbook of spectroscopy. in *Handbook of Spectroscopy* 89–168, John Wiley & Sons, Ltd, 2003. Weinheim, Germany doi:10.1002/3527602305.ch6.

58. Fearn, T. Discriminant Analysis. In *Handbook of Vibrational Spectroscopy*; Chalmers, J. M., Griffiths, P. R., Eds.; John Wiley & Sons, Ltd: New York, United States, 2006; p s4302. doi:10.1002/0470027320.s4302
59. Dubois, J.; Shaw, R. A. Peer Reviewed: IR Spectroscopy in Clinical and Diagnostic Applications. *Anal. Chem.* 2004, *76*, 360A–367A. doi:10.1021/ac041647v
60. Perez-Guaita, D.; Garrigues, S.; de la, M.; Guardia. Infrared-Based Quantification of Clinical Parameters. *TrAC Trends Anal. Chem.* 2014, *62*, 93–105. doi:10.1016/j.trac.2014.06.012
61. Biancolillo, A.; Marini, F. Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. *Front. Chem.* 2018, *6*, 576. doi:10.3389/fchem.2018.00576
62. Paraskevaidi, M.; Morais, C. L. M.; Lima, K. M. G.; Snowden, J. S.; Saxon, J. A.; Richardson, A. M. T.; Jones, M.; Mann, D. M. A.; Allsop, D.; Martin-Hirsch, P. L.; et al. Differential Diagnosis of Alzheimer's Disease Using Spectrochemical Analysis of Blood. *Proc. Natl. Acad. Sci. U S A*. 2017, *114*, E7929–E7938. doi:10.1073/pnas.1701517114
63. Peuchant, E.; Richard-Harston, S.; Bourdel-Marchasson, I.; Dartigues, J.-F.; Letenneur, L.; Barberger-Gateau, P.; Arnaud-Dabernat, S.; Daniel, J.-Y. Infrared Spectroscopy: A Reagent-Free Method to Distinguish Alzheimer's Disease Patients from Normal-Aging Subjects. *Transl. Res*. 2008, *152*, 103–112. doi:10.1016/j.trsl.2008.05.009
64. Yonar, D.; Ocek, L.; Tiftikcioglu, B. I.; Zorlu, Y.; Severcan, F. Relapsing-Remitting Multiple Sclerosis Diagnosis from Cerebrospinal Fluids via Fourier Transform Infrared Spectroscopy Coupled with Multivariate Analysis. *Sci. Rep*. 2018, *8*, 1025. doi:10.1038/s41598-018-19303-3
65. Abbas, S.; Simsek Ozek, N.; Emri, S.; Koksal, D.; Severcan, M.; Severcan, F. Diagnosis of Malignant Pleural Mesothelioma from Pleural Fluid by Fourier Transform-Infrared Spectroscopy Coupled with Chemometrics. *J. Biomed. Opt*. 2018, *23*, 1–14. doi:10.1117/1.JBO.23.10.105003
66. Depciuch, J.; Sowa-Kućma, M.; Nowak, G.; Dudek, D.; Siwek, M.; Styczeń, K.; Parlińska-Wojtan, M. Phospholipid-Protein Balance in Affective Disorders: Analysis of Human Blood Serum Using Raman and FTIR Spectroscopy. A Pilot Study. *J. Pharm. Biomed. Anal*. 2016, *131*, 287–296. doi:10.1016/j.jpba.2016.08.037
67. Sitole, L.; Steffens, F.; Krüger, T. P. J.; Meyer, D. Mid-ATR-FTIR Spectroscopic Profiling of HIV/AIDS Sera for Novel Systems Diagnostics in Global Health. *OMICS J. Integr. Biol.* 2014, *18*, 513–523. doi:10.1089/omi.2013.0157
68. Firdous, S.; Nawaz, M.; Ahmed, M.; Anwar, S.; Rehman, A.; Rashid, R.; Mahmood, A. Measurement of Diabetic Sugar Concentration in Human Blood Using Raman Spectroscopy. *Laser Phys.* 2012, *22*, 1090–1094. doi:10.1134/S1054660X12060023
69. Backhaus, J.; Mueller, R.; Formanski, N.; Szlama, N.; Meerpohl, H.-G.; Eidt, M.; Bugert, P. Diagnosis of Breast Cancer with Infrared Spectroscopy from Serum Samples. *Vib. Spectrosc.* 2010, *52*, 173–177. doi:10.1016/j.vibspec.2010.01.013
70. Paraskevaidi, M.; Morais, C. L. M.; Lima, K. M. G.; Ashton, K. M.; Stringfellow, H. F.; Martin-Hirsch, P. L.; Martin, F. L. Potential of Mid-Infrared Spectroscopy as a Non-Invasive Diagnostic Test in Urine for Endometrial or Ovarian Cancer. *Analyst.* 2018, *143*, 3156–3163. doi:10.1039/c8an00027a
71. Wang, X.; Shen, X.; Sheng, D.; Chen, X.; Liu, X. FTIR Spectroscopic Comparison of Serum from Lung Cancer Patients and Healthy Persons. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 2014, *122*, 193–197. doi:10.1016/j.saa.2013.11.049
72. Theophilou, G.; Lima, K. M. G.; Martin-Hirsch, P. L.; Stringfellow, H. F.; Martin, F. L. ATR-FTIR Spectroscopy Coupled with Chemometric Analysis Discriminates Normal, Borderline and Malignant Ovarian Tissue: Classifying Subtypes of Human Cancer. *Analyst.* 2016, *141*, 585–594. doi:10.1039/c5an00939a
73. Mitchell, A. L.; Gajjar, K. B.; Theophilou, G.; Martin, F. L.; Martin-Hirsch, P. L. Vibrational Spectroscopy of Biofluids for Disease Screening or Diagnosis: Translation from the Laboratory to a Clinical Setting. *J. Biophotonics*. 2014, *7*, 153–165. doi:10.1002/jbio.201400018

74. Pickrell, J. K.; Coop, G.; Novembre, J.; Kudaravalli, S.; Li, J. Z.; Absher, D.; Srinivasan, B. S.; Barsh, G. S.; Myers, R. M.; Feldman, M. W.; et al. Signals of Recent Positive Selection in a Worldwide Sample of Human Populations. *Genome Res.* 2009, *19*, 826–837. doi:10.1101/gr.087577.108
75. Barreiro, L. B.; Quintana-Murci, L. From Evolutionary Genetics to Human Immunology: How Selection Shapes Host Defence Genes. *Nat. Rev. Genet.* 2010, *11*, 17–30. doi:10.1038/nrg2698
76. Barreiro, L. B.; Laval, G.; Quach, H.; Patin, E.; Quintana-Murci, L. Natural Selection Has Driven Population Differentiation in Modern Humans. *Nat. Genet.* 2008, *40*, 340–345. doi:10.1038/ng.78
77. Voight, B. F.; Kudaravalli, S.; Wen, X.; Pritchard, J. K. A Map of Recent Positive Selection in the Human Genome. *PLoS Biol.* 2006, *4*, e72. doi:10.1371/journal.pbio.0040072
78. Sabeti, P. C.; Varilly, P.; Fry, B.; Lohmueller, J.; Hostetter, E.; Cotsapas, C.; Xie, X.; Byrne, E. H.; McCarroll, S. A.; Gaudet, R.; International HapMap Consortium; et al. Genome-Wide Detection and Characterization of Positive Selection in Human Populations. *Nature.* 2007, *449*, 913–918. doi:10.1038/nature06250
79. Ramos, P. S.; Shedlock, A. M.; Langefeld, C. D. Genetics of Autoimmune Diseases: Insights from Population Genetics. *J. Hum. Genet.* 2015, *60*, 657–664. doi:10.1038/jhg.2015.94
80. The Immunological Genome Project Consortium; et al. The Immunological Genome Project: Networks of Gene Expression in Immune Cells. *Nat. Immunol.* 2008, *9*, 1091–1094.
81. Rodríguez-Cortez, V. C.; Del Pino-Molina, L.; Rodríguez-Ubreva, J.; Ciudad, L.; Gómez-Cabrero, D.; Company, C.; Urquiza, J. M.; Tegnér, J.; Rodríguez-Gallego, C.; López-Granados, E.; et al. Monozygotic Twins Discordant for Common Variable Immunodeficiency Reveal Impaired DNA Demethylation during Naïve-to-Memory B-Cell Transition. *Nat. Commun.* 2015, *6*, 7335. doi:10.1038/ncomms8335
82. Kienzler, A.-K.; Rizzi, M.; Reith, M.; Nutt, S. L.; Eibel, H. Inhibition of Human B-Cell Development into Plasmablasts by Histone Deacetylase Inhibitor Valproic Acid. *J. Allergy Clin. Immunol.* 2013, *131*, 1695–1699.e9. doi:10.1016/j.jaci.2013.01.018
83. Maggadottir, S. M.; Li, J.; Glessner, J. T.; Li, Y. R.; Wei, Z.; Chang, X.; Mentch, F. D.; Thomas, K. A.; Kim, C. E.; Zhao, Y.; et al. Rare Variants at 16p11.2 Are Associated with Common Variable Immunodeficiency. *J. Allergy Clin. Immunol.* 2015, *135*, 1569–1577. doi:10.1016/j.jaci.2014.12.1939
84. Park, J.; Munagala, I.; Xu, H.; Blankenship, D.; Maffucci, P.; Chaussabel, D.; Banchereau, J.; Pascual, V.; Cunningham-Rundles, C. Interferon Signature in the Blood in Inflammatory Common Variable Immune Deficiency. *PLOS One.* 2013, *8*, e74893. doi:10.1371/journal.pone.0074893
85. Canvin, J. M. G.; Bernatsky, S.; Hitchon, C. A.; Jackson, M.; Sowa, M. G.; Mansfield, J. R.; Eysel, H. H.; Mantsch, H. H.; El-Gabalawy, H. S. Infrared Spectroscopy: Shedding Light on Synovitis in Patients with Rheumatoid Arthritis. *Rheumatology (Oxford).* 2003, *42*, 76–82. doi:10.1093/rheumatology/keg034
86. Carvalho, C. S.; Martin, A. A.; Santo, A. M. E.; Andrade, L. E. C.; Pinheiro, M. M.; Cardoso, M. A. G.; Raniero, L. A Rheumatoid Arthritis Study Using Raman Spectroscopy. *Theor. Chem. Acc.* 2011, *130*, 1211–1220. doi:10.1007/s00214-011-0905-0
87. Lechowicz, L.; Chrapek, M.; Gaweda, J.; Urbaniak, M.; Konieczna, I. Use of Fourier-Transform Infrared Spectroscopy in the Diagnosis of Rheumatoid Arthritis: A Pilot Study. *Mol. Biol. Rep.* 2016, *43*, 1321–1326. doi:10.1007/s11033-016-4079-7
88. Hackshaw, K. V.; Aykas, D. P.; Sigurdson, G. T.; Plans, M.; Madiai, F.; Yu, L.; Buffington, C. A. T.; Giusti, M. M.; Rodriguez-Saona, L. Metabolic Fingerprinting for Diagnosis of Fibromyalgia and Other Rheumatologic Disorders. *J. Biol. Chem.* 2019, *294*, 2555–2568. doi:10.1074/jbc.RA118.005816
89. Morris, A. D.; Morais, C. L. M.; Lima, K. M. G.; Freitas, D. L. D.; Brady, M. E.; Dhaygude, A. P.; Rowbottom, A. W.; Martin, F. L. A Comparative Analysis of Different

Biofluids Using Raman Spectroscopy to Determine Disease Activity in ANCA-Associated Vasculitis. *J. Biophotonics.* 2021, *14*, e202000426. doi:10.1002/jbio.202000426
90. Watts, R. A.; Mahr, A.; Mohammad, A. J.; Gatenby, P.; Basu, N.; Flores-Suárez, L. F. Classification, Epidemiology and Clinical Subgrouping of Antineutrophil Cytoplasmic Antibody (ANCA)-Associated Vasculitis. *Nephrol. Dial. Transplant.* 2015, *30*, i14–22. doi:10.1093/ndt/gfv022
91. Morris, A. D.; Rowbottom, A. W.; Martin, F. L.; Woywodt, A.; Dhaygude, A. P. Biomarkers in ANCA-Associated Vasculitis: Potential Pitfalls and Future Prospects. *Kidney360.* 2021, *2*, 586–597. doi:10.34067/KID.0006432020
92. Morris, A. D.; Morais, C. L. M.; Lima, K. M. G.; Freitas, D. L. D.; Brady, M. E.; Dhaygude, A. P.; Rowbottom, A. W.; Martin, F. L. Distinguishing Active from Quiescent Disease in ANCA-Associated Vasculitis Using Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy. *Sci. Rep.* 2021, *11*, 9981. doi:10.1038/s41598-021-89344-8
93. Yu, M.-C.; et al. Label Free Detection of Sensitive Mid-Infrared Biomarkers of Glomerulonephritis in Urine Using Fourier Transform Infrared Spectroscopy. *Sci. Rep.* 2017, *7*, 1–12.
94. Thompson, A. J.; Banwell, B. L.; Barkhof, F.; Carroll, W. M.; Coetzee, T.; Comi, G.; Correale, J.; Fazekas, F.; Filippi, M.; Freedman, M. S.; et al. Diagnosis of Multiple Sclerosis: 2017 Revisions of the McDonald Criteria. *Lancet Neurol.* 2018, *17*, 162–173. doi:10.1016/S1474-4422(17)30470-2
95. Freedman, M. S.; Comi, G.; De Stefano, N.; Barkhof, F.; Polman, C. H.; Uitdehaag, B. M. J.; Lehr, L.; Stubinski, B.; Kappos, L. Moving toward Earlier Treatment of Multiple Sclerosis: Findings from a Decade of Clinical Trials and Implications for Clinical Practice. *Mult. Scler. Relat. Disord.* 2014, *3*, 147–155. doi:10.1016/j.msard.2013.07.001
96. 1Royal College of Physicians Working Party. *Allergy: the unmet need: a blueprint for better patient care.* Royal College of Physicians of London. 2003
97. Sousa-Pinto, B.; Tarrio, I.; Blumenthal, K. G.; Araújo, L.; Azevedo, L. F.; Delgado, L.; Fonseca, J. A. Accuracy of Penicillin Allergy Diagnostic Tests: A systematic Review and Meta-analysis. *J. Allergy Clin. Immunol.* 2021, *147*, 296–308. doi:10.1016/j.jaci.2020.04.058
98. Yu, W.; Freeland, D. M. H.; Nadeau, K. C. Food Allergy: Immune Mechanisms, Diagnosis and Immunotherapy. *Nat. Rev. Immunol.* 2016, *16*, 751–765. doi:10.1038/nri.2016.111
99. Sahu, A.; Dalal, K.; Naglot, S.; Aggarwal, P.; Krishna, C. M. Serum Based Diagnosis of Asthma Using Raman Spectroscopy: An Early Phase Pilot Study. *PLOS One.* 2013, *8*, e78921. doi:10.1371/journal.pone.0078921
100. Mazarakis, N.; Vongsvivut, J.; Bambery, K. R.; Ververis, K.; Tobin, M. J.; Royce, S. G.; Samuel, C. S.; Snibson, K. J.; Licciardi, P. V.; Karagiannis, T. C.; et al. Investigation of Molecular Mechanisms of Experimental Compounds in Murine Models of Chronic Allergic Airways Disease Using Synchrotron Fourier-Transform Infrared Microspectroscopy. *Sci. Rep.* 2020, *10*, 11713. doi:10.1038/s41598-020-68671-2
101. Banerjee, A.; Trivedi, C. M.; Damera, G.; Jiang, M.; Jester, W.; Hoshi, T.; Epstein, J. A.; Panettieri, R. A. Trichostatin a Abrogates Airway Constriction, but Not Inflammation, in Murine and Human Asthma Models. *Am. J. Respir. Cell Mol. Biol.* 2012, *46*, 132–138. doi:10.1165/rcmb.2010-0276OC
102. Lee, H. Y.; Kim, I. K.; Yoon, H. K.; Kwon, S. S.; Rhee, C. K.; Lee, S. Y. Inhibitory Effects of Resveratrol on Airway Remodeling by Transforming Growth Factor-β/Smad Signaling Pathway in Chronic Asthma Model. *Allergy Asthma Immunol. Res.* 2017, *9*, 25–34. doi:10.4168/aair.2017.9.1.25
103. Han, S.; Zhao, H.; Lü, L.; Wang, X.; Wang, X. Effects of H_2 Relaxin on the Expression of Epac in a Murine Model of Chronic Asthma. *Zhonghua Yi Xue Za Zhi.* 2012, *92*, 3305–3309.
104. Depciuch, J.; Kasprzyk, I.; Sadik, O.; Parlińska-Wojtan, M. FTIR Analysis of Molecular Composition Changes in Hazel Pollen from Unpolluted and Urbanized Areas. *Aerobiologia (Bologna).* 2017, *33*, 1–12. doi:10.1007/s10453-016-9445-3

105. Guedes, A.; Ribeiro, H.; Fernández-González, M.; Aira, M. J.; Abreu, I. Pollen Raman Spectra Database: Application to the Identification of Airborne Pollen. *Talanta.* 2014, *119*, 473–478. doi:10.1016/j.talanta.2013.11.046
106. To, T.; Stanojevic, S.; Moores, G.; Gershon, A. S.; Bateman, E. D.; Cruz, A. A.; Boulet, L.-P. Global Asthma Prevalence in Adults: Findings from the Cross-Sectional World Health Survey. *BMC Public Health.* 2012, *12*, 204. doi:10.1186/1471-2458-12-204
107. Caussin, J.; Gooris, G. S.; Bouwstra, J. A. FTIR Studies Show Lipophilic Moisturizers to Interact with Stratum Corneum Lipids, Rendering the More Densely Packed. *Biochim. Biophys. Acta*. 2008, *1778*, 1517–1524. doi:10.1016/j.bbamem.2008.03.006
108. de Sousa Neto, D.; Gooris, G.; Bouwstra, J. Effect of the ω-Acylceramides on the Lipid Organization of Stratum Corneum Model Membranes Evaluated by X-Ray Diffraction and FTIR Studies (Part I). *Chem. Phys. Lipids*. 2011, *164*, 184–195. doi:10.1016/j.chemphyslip.2010.12.007
109. Boncheva, M.; Damien, F.; Normand, V. Molecular Organization of the Lipid Matrix in Intact Stratum Corneum Using ATR-FTIR Spectroscopy. *Biochim. Biophys. Acta*. 2008, *1778*, 1344–1355. doi:10.1016/j.bbamem.2008.01.022
110. Berkers, T.; Boiten, W. A.; Absalah, S.; van Smeden, J.; Lavrijsen, A. P. M.; Bouwstra, J. A. Compromising Human Skin In Vivo and Ex Vivo to Study Skin Barrier Repair. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids.* 2019, *1864*, 1103–1108. doi:10.1016/j.bbalip.2019.04.005
111. Caspers, P. J.; Lucassen, G. W.; Wolthuis, R.; Bruining, H. A.; Puppels, G. J. In Vitro and In Vivo Raman Spectroscopy of Human Skin. *Biospectroscopy.* 1998, *4*, S31–S39. doi:10.1002/(SICI)1520-6343(1998)4:5+<S31::AID-BSPY4>3.0.CO;2-M
112. Sheffield Teaching Hospitals NHS Foundation Trust. *Validation of a Novel Composite of Skin Biomarkers as a Primary Outcome Measure for Evaluating the Safety of Treatments for Atopic Dermatitis: A Randomized Controlled Trial (Phase 2) Comparing the Effects of Crisaborole 2% Ointment to Betamethasone Valerate 0.1% Cream on Skin Structure and Function in Participants With Atopic Dermatitis*. 2020. https://clinicaltrials.gov/ct2/show/NCT04194814.
113. Korb, E.; Bağcı oğlu, M.; Garner-Spitzer, E.; Wiedermann, U.; Ehling-Schulz, M.; Schabussova, I. Machine Learning-Empowered FTIR Spectroscopy Serum Analysis Stratifies Healthy, Allergic, and SIT-Treated Mice and Humans. *Biomolecules.* 2020, *10*, 1058. doi:10.3390/biom10071058
114. Notarangelo, L. D.; Fischer, A.; Geha, R. S.; Casanova, J.-L.; Chapel, H.; Conley, M. E.; Cunningham-Rundles, C.; Etzioni, A.; Hammartröm, L.; Nonoyama, S.; International Union of Immunological Societies Expert Committee on Primary Immunodeficiencies; et al. Primary Immunodeficiencies: 2009 Update. *J. Allergy Clin. Immunol.* 2009, *124*, 1161–1178. doi:10.1016/j.jaci.2009.10.013
115. Tangye, S. G.; Al-Herz, W.; Bousfiha, A.; Chatila, T.; Cunningham-Rundles, C.; Etzioni, A.; Franco, J. L.; Holland, S. M.; Klein, C.; Morio, T.; et al. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* 2020, *40*, 24–64. doi:10.1007/s10875-019-00737-x
116. Callery, E. L.; Morais, C. L. M.; Paraskevaidi, M.; Brusic, V.; Vijayadurai, P.; Anantharachagan, A.; Martin, F. L.; Rowbottom, A. W. New Approach to Investigate Common Variable Immunodeficiency Patients Using Spectrochemical Analysis of Blood. *Sci. Rep*. 2019, *9*, 7239. doi:10.1038/s41598-019-43196-5
117. Cunningham-Rundles, C. The Many Faces of Common Variable Immunodeficiency. *Hematology. Am. Soc. Hematol. Educ. Program*. 2012, *2012*, 301–305. doi:10.1182/asheducation.V2012.1.301.3798316
118. Bonilla, F. A.; Barlan, I.; Chapel, H.; Costa-Carvalho, B. T.; Cunningham-Rundles, C.; de la Morena, M. T.; Espinosa-Rosales, F. J.; Hammarström, L.; Nonoyama, S.; Quinti, I.; et al. International Consensus Document (ICON): Common Variable Immunodeficiency Disorders. *J. Allergy Clin. Immunol. Pract.* 2016, *4*, 38–59. doi:10.1016/j.jaip.2015.07.025

119. Maguire, A.; Vega-Carrascal, I.; Bryant, J.; White, L.; Howe, O.; Lyng, F. M.; Meade, A. D. Competitive Evaluation of Data Mining Algorithms for Use in Classification of Leukocyte Subtypes with Raman Microspectroscopy. *Analyst.* 2015, *140*, 2473–2481. doi:10.1039/c4an01887g
120. Chen, M.; McReynolds, N.; Campbell, E. C.; Mazilu, M.; Barbosa, J.; Dholakia, K.; Powis, S. J. The Use of Wavelength Modulated Raman Spectroscopy in Label-Free Identification of T Lymphocyte Subsets, Natural Killer Cells and Dendritic Cells. *PLOS One.* 2015, *10*, e0125158. doi:10.1371/journal.pone.0125158
121. Hobro, A. J.; Kumagai, Y.; Akira, S.; Smith, N. I. Raman Spectroscopy as a Tool for Label-Free Lymphocyte Cell Line Discrimination. *Analyst.* 2016, *141*, 3756–3764. doi:10.1039/c6an00181e
122. McReynolds, N.; Cooke, F. G. M.; Chen, M.; Powis, S. J.; Dholakia, K. Multimodal Discrimination of Immune Cells Using a Combination of Raman Spectroscopy and Digital Holographic Microscopy. *Sci. Rep*. 2017, *7*, 43631. doi:10.1038/srep43631
123. Shaw, R. A.; Mantsch, H. H. Infrared Spectroscopy of Biological Fluids in Clinical and Diagnostic Analysis. In *Encyclopedia of Analytical Chemistry*; John Wiley & Sons, Ltd: Chichester, New York, United States, 2006.
124. Lin, C.-C.; Kuo, M.-T.; Chang, H.-C. Review: Raman Spectroscopy—a Novel Tool for Noninvasive Analysis of Ocular Surface Fluid. *J. Med. Biol. Eng.* 2010, *30*, 343–354.
125. Untereiner, V.; Sockalingum, G. D.; Garnotel, R.; Gobinet, C.; Ramaholimihaso, F.; Ehrhard, F.; Diebold, M.-D.; Thiéfin, G. Bile Analysis Using High-Throughput FTIR Spectroscopy for the Diagnosis of Malignant Biliary Strictures: A Pilot Study in 57 Patients: Spectral Diagnosis of Malignant Biliary Strictures. *J. Biophotonics*. 2014, *7*, 241–253. doi:10.1002/jbio.201300166
126. Shaw, R. A.; Kotowich, S.; Mantsch, H. H.; Leroux, M. Quantitation of Protein, Creatinine, and Urea in Urine by Near-Infrared Spectroscopy. *Clin. Biochem*. 1996, *29*, 11–19. doi:10.1016/0009-9120(95)02011-X
127. Brandstetter, M.; Sumalowitsch, T.; Genner, A.; Posch, A. E.; Herwig, C.; Drolz, A.; Fuhrmann, V.; Perkmann, T.; Lendl, B. Reagent-Free Monitoring of Multiple Clinically Relevant Parameters in Human Blood Plasma Using a Mid-Infrared Quantum Cascade Laser Based Sensor System. *Analyst.* 2013, *138*, 4022–4028. doi:10.1039/c3an00300k
128. Deegan, R. D.; Bakajin, O.; Dupont, T. F.; Huber, G.; Nagel, S. R.; Witten, T. A. Capillary Flow as the Cause of Ring Stains from Dried Liquid Drops. *Nature.* 1997, *389*, 827–829. doi:10.1038/39827
129. Yunker, P. J.; Still, T.; Lohr, M. A.; Yodh, A. G. Suppression of the Coffee-Ring Effect by Shape-Dependent Capillary Interactions. *Nature.* 2011, *476*, 308–311. doi:10.1038/nature10344
130. Lovergne, L.; Clemens, G.; Untereiner, V.; Lukaszweski, R. A.; Sockalingum, G. D.; Baker, M. J. Investigating Optimum Sample Preparation for Infrared Spectroscopic Serum Diagnostics. *Anal. Methods.* 2015, *7*, 7140–7149. doi:10.1039/C5AY00502G
131. Ollesch, J.; Drees, S. L.; Heise, H. M.; Behrens, T.; Brüning, T.; Gerwert, K. FTIR Spectroscopy of Biofluids Revisited: An Automated Approach to Spectral Biomarker Identification. *Analyst.* 2013, *138*, 4092–4102. doi:10.1039/c3an00337j
132. Zohdi, V.; Whelan, D. R.; Wood, B. R.; Pearson, J. T.; Bambery, K. R.; Black, M. J. Importance of Tissue Preparation Methods in FTIR Micro-Spectroscopical Analysis of Biological Tissues: 'Traps for New Users'. *PloS One.* 2015, *10*, e0116491. doi:10.1371/journal.pone.0116491
133. Finlayson, D.; Rinaldi, C.; Baker, M. J. Is Infrared Spectroscopy Ready for the Clinic? *Anal. Chem*. 2019, *91*, 12117–12128. doi:10.1021/acs.analchem.9b02280
134. Baker, M. J. et al. Clinical applications of infrared and Raman spectroscopy: state of play and future challenges. *The Analyst*. 2018, *143*, 1735–1757.

1.2 Introduction (continued)

This section has been added to supplement the review article in section 1.1[1]. This will provide an introduction to the immunological disorder, common variable immunodeficiency deficiency (CVID), the clinical setting in which this project will focus. The project aims and objectives have been included out at the end of this section, bringing together aspects from both sections of the introduction to set out the line of enquiry within this research project.

Further studies to investigate CVID using omics-based methods (transcriptomic and proteomics) have been published since the experimental work for this project has been completed; the findings and relationship to biospectroscopy analysis will be discussed in this section.

1.2.1 Common Variable Immunodeficiency

Common variable immunodeficiency deficiency (CVID) is the most frequent life-threatening and symptomatic primary immune deficiency (PID) diagnosed in adulthood[2]. Precise prevalence data is lacking, but is estimated to be at between 1:10,000 and 1:100,000 of the population, with two peak ages of onset, one before the age of ten and another between 30-40 years of age[3,4].

CVID is a heterogeneous group of polygenic disorders for which an exact pathogenesis remains poorly understood[5–7]. The majority (>80%) of CVID cases are thought to be sporadic and therefore diagnosis remains one of exclusion[8]. Due to the molecular and symptomatic diversity found amongst PIDs, discrimination between primary and secondary antibody deficiencies can be challenging in the clinical setting; evidenced by an average diagnostic delay of around 4-5 years[9]. Correct and prompt classification is vital to improve patient outcomes; ensuring appropriate patient pathways are followed and further disease-associated complications are recognised and well-managed.

In CVID, 90% of patients suffer recurrent infections resulting from a failure to produce protective immunoglobulins, there is also an increased risk of autoimmune disorders (22% of patients) and malignancy (16% of patients)[3,10–12]. These non-infectious

manifestations related to underlying immune dysregulation are associated with an increased morbidity and mortality within the CVID patient group[13,14].

In the management of CVID patients, it is generally accepted that the only therapeutic option available is replacement immunoglobulin. Evidence for its use in CVID is substantial; with adequate replacement therapy resulting in improvements in the overall health and quality of life of patients, increasing life expectancy and preventing and major infections[15–17].

Replacement immunoglobulin can be given intravenously or subcutaneously, and for most patients treatment is a lifelong requirement. This can be supplemented with antibiotics as clinically required in certain cases, and to treat breakthrough infections. Dosing is patient-dependent and clinically targeted however a general aim is to maintain trough levels of IgG within the normal adult range (6–16 g/L)[18,19]. Whilst immunoglobulin replacement therapy can prevent infectious complications, there is limited evidence to guide clinicians on the treatment and management of patients with inflammatory complications resulting from immune dysregulation[20,21].

1.2.2 The need for a new approach to diagnosis and management

There are no clinical or laboratory features that are pathognomonic for CVID therefore diagnostic criteria have been developed which require sequential application of both clinical and laboratory findings in order to increase the specificity for the diagnosis. Current diagnostic criteria[10,22] define hypogammaglobulinemia as a major requirement for the diagnosis of CVID, but only when used in conjunction with further clinical or laboratory criteria. This is due to the low specificity of hypogammaglobulinemia for CVID; reduced serological levels of immunoglobulin are associated with a vast array of other primary and secondary immune disorders.

The incidence and prevalence rates of hypogammaglobulinemia are not clearly defined, however secondary causes are more common; Perez et al.[16] review the incidence of hypogammaglobulinemia in several disorders. The requirement to exclude secondary causes of hypogammaglobulinemia limits the use of immunoglobulins as a diagnostic marker however given the hallmark of CVID is a reduced serum level of IgG, IgA and/or IgM; it remains a fundamental test for this disorder.

The definition of local or regional population reference ranges will impact the diagnostic utility of results, as will the laboratory choice of analytical method (with nephelometry and turbidimetry being most widely used). An absolute lower limit value of IgG at 4.5 g/L for adults has been proposed; as despite the wide range of IgG levels observed in CVID patients, Chapel and Cunningham-Rundles[3] described the majority of their 334 patients (94%) as having initial IgG levels <4.5 g/l at diagnosis.

Further laboratory testing such as measuring specific antibodies to vaccine responses, enumeration of lymphocyte subsets (B, T and NK cells) and class-switched memory B cells by flow cytometry can provide additional evidence to suggest defective antibody production. The application of reference ranges for peripheral blood immunophenotyping remains a challenge, particularly within CVID, due to the disease heterogeneity. Findings are variable across the disease group and can also vary within individual patients. Repeat testing has been suggested to confirm any sub-normal findings, these limitations have been discussed in both of the recent diagnostic criteria[10,22].

Efforts to categorise clinical subgroups within cohorts of established CVID patients have further demonstrated the complex diversity of this disorder. Analysis of B cell phenotype using flow cytometry has provided additional classification protocols for this heterogeneous disease; examining the population sizes of class-switched memory- and transitional B cells in correlation with clinical aspects have demonstrated that defects at various stages of B cell differentiation occur in different subgroups of CVID patients[23–25].

Known phenotypic B cell populations associated with increased risk of immune dysregulation in CVID include a severe reduction in class-switched memory B cells (splenomegaly and granulomatous disease); expansion of $CD21^{low}$ B cells (splenomegaly); and expansion of transitional B cells (lymphadenopathy)[24]. Whilst these studies confirm that B-cell homeostasis is a pathogenic and clinically meaningful parameter for classification, low numbers of class-switched or memory B cells are not specific to CVID therefore diagnostic utility of these tests in isolation is limited[3]. Other risk factors for the development of immune dysregulation include a reduction of naïve CD4+ T cells and IgA deficiency[26,27].

Immunophenotyping can provide a useful tool towards clinically subgrouping CVID patients; however it provides only limited information on a patient's current inflammatory profile. Further to this, the peripheral blood compartment may not be the most appropriate to evaluate when searching for diagnostic or prognostic markers in CVID. The lack of protective, class-switched B cells suggests defective germinal centre function; however, it would not be suitable to undertake lymph node biopsies and immunophenotypic analysis of germinal centres as a routine front-line test, therefore investigations are limited to peripheral blood.

Given that non-infectious complications are associated with increased morbidity and mortality, there is an unmet need for the identification of new biomarkers associated with immune dysregulation, and further, a means to monitor patients for disease progression.

Similar to other complex diseases the aetiology of CVID is likely multifactorial, with genetic and environmental factors both contributing to the development of the immune dysregulation[7]. In recent years a number of monogenic disease-causing defects have been discovered, these include genes for ICOS, TACI, CD19, Msh5, CD81, CD20, CD21 and BAFF-R, which contribute to key immune regulatory pathways[8,28–32]. These mutations are only found in a small minority of cases (approximately 10-20%); nevertheless they provide a further diagnostic test and add to the knowledge on the pathogenic mechanisms of antibody deficiency[33].

In the remaining 80% of patients, the lack of an identified genetic defect not only renders the diagnosis one of exclusion, it also limits the development of novel treatment approaches and personalised management plans as the immune components and signalling pathways involved remain unknown. In order to improve understanding of the CVID, and gain a greater insight into the pathogenesis of antibody deficiencies, further detailed investigations and immunological studies are required.

To explore the likely polygenic cause of the majority of CVID patients, the combined use of 'omics'-like technologies and whole sample profiling is beginning to provide more clues in the search for diagnostic biomarkers and causative aetiology. Through advances in molecular methods it is now understood that additional micro-environmental factors play

a role in the varied clinical phenotypes observed within a disease entity. In CVID, these include post-translational epigenetic alterations, such as DNA methylation[34] and histone modifications[35]; as well as transcriptional disturbances, such as single nucleotide polymorphisms (SNPs) impacting on CVID susceptibility[36–38].

Several research groups have reported disturbed cytokine and chemokine profiles in CVID patients[39–45], however the development of a diagnostic or disease-monitoring profile has not been successful due to lack of agreement across the studies. CVID-associated profiles include increased serum levels of cytokines: IL-4 and IL-10[40]; IL-2 and IL-10[41].; IL-6, IL-8 and TNF-α[42,46]; and increased IL-10, IL-RA, and TNF-α[43]. In contrast to majority of studies, IL-10 levels were reported to be decreased by Polito et al.,[46] with levels of IL-2, IL-4, and INF-γ comparable between healthy controls and CVID patients.

The mixed findings likely reflect the clinical heterogeneity between CVID patients, but will also relate to the current inflammatory state of a patient, for example, CVID patients with immune dysregulation and active inflammation will likely have a different cytokine profile to CVID patients with an infection-only phenotype. To address this, recent studies have focussed on segregating patients into CVID with immune dysregulation-, and CVID with infection-only groups. Through the use of ‘omics-based technologies coupled with machine-learning, signature immune profiles for different subgroups of CVID patients are beginning to emerge[37,47–49].

In 2020, Hultberg et el.[49] applied proteomic technology, immunophenotyping and multivariate analysis for the first time to the study of CVID. In comparing all CVID patients from healthy controls they demonstrated differences between 72 of the 145 immune-related proteins detected in the plasma samples. Further to this, they determined two distinct plasma protein profiles capable of segregating CVID patients with immune dysregulation from those with infections only. In the immune dysregulation patients, 24 proteins related to INFγ signalling were found to be elevated. The immune profile associated with immune dysregulation included increased CD4+ T-helper-1 (Th1) cytokines, increased T cell activation (increased numbers of CD8+ T cells, increased CD8+HLADR+ T cells; increased CD4+HLA-DR+ T cells and increased CD4+PD-1+ T cells), and an increased concentration of the specific chemokine, CXCL10, which was identified as a potential predictive biomarker for this group.

In 2021, Berbers et al.[48] measured 180 immune-related markers using targeting proteomics; following the application of multivariate analysis and machine-learning algorithms they successfully classified infection-only CVID patients from patients with immune dysregulation. They identified MILR1, LILRB4, IL10, IL12RB1, and CD83 and key markers to discriminate between groups with a sensitivity of 83%, and specificity of 75%. This study corroborated the finding of Hultberg et al. in reporting increased Th1 signalling in CVID patients with immune dysregulation, in addition, they demonstrated increased IL-10, LAG3, TNFRSF9, CD83 and Th17 signalling in this group. An ongoing inflammatory response resulting in immune exhaustion has been postulated as an underlying mechanism driving immune dysregulation[48,49].

These studies highlight the potential use of 'omics-based profiling as a diagnostic tool to identify CVID patients at risk of developing immune dysregulation. Further studies to validate recently identified biomarkers and to determine their role in driving immunopathology are required. This will not only advance our understanding of CVID but could enable risk stratification and individualisation of treatment. However, 'omics-based technologies are not widely used in the clinical setting; there remains a need to reduce costs, overcome technical challenges and to alleviate complexities associated with 'big data' processing before these platforms can be standardised and subsequently enter the routine clinical setting[50,50,51].

In light of this, novel approaches to complement current molecular findings are warranted. One potential and innovative candidate that could provide molecular level analysis for CVID patient blood samples is vibrational spectroscopy; specifically attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy. As discussed in detail in Section 1.1[1], ATR-FTIR spectroscopy describes the technique used to produce a unique spectrum, or molecular 'fingerprint' of a sample following excitation with light. The molecular fingerprint of a given sample relates to its biomolecular constituents (i.e. proteins, lipids, nucleic acids, carbohydrates) and is generated from the vibrations of the chemical bonds present in these molecules.

The molecular composition of a sample can be altered in the presence of disease; these changes will be reflected in the spectrum obtained using vibrational spectroscopy, making it an ideal choice for the study of pathological processes and the development of a novel

diagnostic platform[52]. Towards this goal, the interrogation of whole unprocessed samples with vibrational spectroscopy will provide a new approach to investigate CVID, a disorder that remains largely undefined, despite four decades of research[12]. The use of this platform not only has diagnostic potential in the clinical setting, but will generate a new library of 'big data' for CVID (at an mRNA, DNA and protein level). This data can subsequently be examined in a 'systems biology' approach to add scientific knowledge and further understanding of the pathological mechanisms underpinning CVID.

Through the use of multi-omics technologies, the challenges once faced using traditional, and even single 'omics' approaches, could be overcome, enabling the multifactorial pathways of complex immunological disorders to be unravelled[53,54].

1.2.3 Aims and Objectives

The aims of this research project will be to seek out disease-related changes in the 'molecular fingerprint' of CVID patients compared to a control group using the vibrational spectroscopy method ATR-FTIR. This will be a proof-of-concept study which will be initially piloted in a small cohort due to the rarity of the disease. Given that the clinical phenotype and pathological features associated with CVID are likely to manifest from underlying defects and alterations in cellular mechanisms; these changes may be reflected in the biological composition of blood samples therefore be detectable by vibrational spectroscopy. We hypothesise that pathological changes in CVID patients produce characteristic FTIR spectra that distinguish them from healthy controls.

The spectral data generated from blood serum and plasma samples collected from a cohort of CVID patients and a control group will be analysed to investigate whether reproducible variances exist within the well-described fingerprint (1800-900 cm^{-1}) and high (3700-2800 cm^{-1}) regions of the IR spectrum. These variances will be subject to a blind testing and verification process to assess whether the method can correctly classify individuals into the correct category (disease versus non-disease). Postulations will be made as to the molecular composition attributing to any variances identified.

The results from conventional serological investigations and clinical notes will be used to subgroup CVID patients with the aim to assess whether spectral variances can be detected between patients with- and without further disease-associated complications. In

the application of vibrational spectroscopy to CVID, we anticipate that similar analytical advantages will be observed to those previously described in the literature for complex pathologies such as malignancy and neurological disorders. Through the detection of characteristic spectral features relating to CVID, this may lead to the development of a novel diagnostic technique and a means to risk-stratify patients, in addition to uncovering new molecular data to further elucidate pathological processes in CVID. We hypothesise that FTIR spectroscopy will provide an improved description of CVID; the work conducted within this thesis will address this through the aims and objectives set out within this section.

1.2.4 Summary of Aims

- Identify the spectral pattern (band wavelengths and intensities) known as the 'molecular fingerprint' for each sample of the CVID patient cohort and control group.
- Compare spectral features from CVID patients against normal controls by performing computational analysis on each unique spectrum to investigate variances.
- Develop and verify classification methods to allow for correct assignment of unknown samples into the correct groups (CVID versus non-CVID).
- Correlate spectral variances with clinical presentation and results from conventional immunological investigations in CVID patients.
- Postulate molecular bond assignments in order to tentatively identify biological constituents responsible for heterogeneous clinical presentations of CVID.

1.3 Alternative thesis

This research project and thesis has been completed and written up in journal format towards the University of Manchester qualification for Doctorate of Clinical Sciences (module C). The publication-style thesis was chosen by the author as the most appropriate format to present the research findings; this enabled the author to gain experience of submitting original work for peer review, and resulted in the successful publication of two journal articles. A summary of each of the thesis chapters has been provided below, including the level of contribution made by the author, and any co-

authors listed on the two published articles. As first author on each of the published papers, I played a major role and contributed to all aspects of the work, including conception of research question, obtaining ethical approval, undertaking the practical work for data acquisition, analysis of results and the writing of the papers. A series of appendices are also included at the end of this thesis, which refer to additional scientific training and qualifications gained, academic components completed, and evidence of any research-related experiences gained during the course of the higher specialist scientific training programme.

Chapter 1 – (Introduction: review article) Vibrational spectroscopy and multivariate analysis techniques in the clinical immunology laboratory: a review of current applications and requirements for diagnostic use

Authors: Emma L. Callery & Anthony W. Rowbottom

Publishing journal: Applied Spectroscopy Reviews

This paper provides an overview of vibrational spectroscopy in the clinical setting, with a focus on clinical immunology; bringing together both underpinning knowledge of vibrational spectroscopy methods and the impact that advances in these platforms may have on future immunology investigations. The review article was constructed from an iteration of the assessed literature review, completed as part of module C of the DClinSci Higher Special Scientific Training Programme, expanded to include a review of current applications in clinical immunology. Building upon the recent expansion of interest for using vibrational spectroscopy in the clinical setting (predominantly in cancer diagnostics), this review covers the application of this methodology within three major areas of immunology; Autoimmune disease, Allergic and hypersensitivity disorders and Immune deficiency.

This review was published in Applied Spectroscopy Reviews in 2021 [1]. As first author I am solely responsible for writing the text for this paper. The co-author contributed to the conception of the review, critically reviewed the final draft of the manuscript and provided intellectual discussions toward the successful publication of this article.

Chapter 2 – Method Evaluation

This chapter considered three areas of methodological and analytical process associated with biospectroscopy. The findings from this chapter have informed the final processes applied to produce the results in the main study (Chapter 3). This included the choice of pre-processing technique applied to the spectral data prior to multivariate analysis, the suitability and variability between blood sample types (serum versus plasma) and the potential impact of environmental factors on analysis of spectra collected over multiple different days.

I am responsible for writing the entirety of this text with the exception of a sub-section of the methods section, 'multivariate analysis', which has been contributed to by Camilo L. M. Morais, a chemometrician and co-author listed on the main study (Chapter 3); The complete chapter has been reviewed by my supervisors, comments and suggestions following review were included in the final text. I am responsible for the computational (chemometric) and statistical analysis of the data with the sections examining pre-processing techniques and blood sample-type suitability. In the section examining the impact of environmental factors, Camilo L. M. Morais performed chemometric analysis of the spectral data to produce the results. I am solely responsible for the interpretation and further discussions related to the results obtained throughout all sections of this method evaluation.

Chapter 3 – New Approach to Investigate Common Variable Immunodeficiency Patients Using Spectrochemical Analysis of Blood.

Authors: Callery, E. L.; Morais, C. L. M.; Paraskevaidi, M.; Brusic, V.; Vijayadurai, P.; Anantharachagan, A.; Martin, F. L.; Rowbottom, A. W.

Publishing Journal: Nature Scientific Reports

This manuscript is representative of the main body of work undertaken towards this research project. I was solely responsible for all experimental laboratory work and acquisition of spectral data. Dr Camilo Morais, a chemometrics expert at University of Central Lancashire provided advice and guidance on all computational analysis undertaken within this work. I performed the pre-processing, principal component

analysis (PCA) and PCA-linear discriminant analysis (PCA-LDA) for the acquired spectral data. Complex chemometric techniques including splitting of the datasets, support vector machine (SVM) machine-learning algorithms (applied to classify data) and feature selection methods (applied for biomarker identification) were performed by Dr Morais. My supervisors Professor Anthony Rowbottom and Professor Frank Martin provided advice and guidance on the conception of the study and all experimental work. As first author I am solely responsible for writing the text for this manuscript. The co-authors provided intellectual discussions and critically reviewed the final draft of the manuscript. Comments received from the co-authors, and journal reviewers were incorporated into the final text to produce the final published manuscript.

Chapter 4 – Critical Appraisal of Project and suggestions for further work

This chapter provides a critical appraisal of the research project, focussing on the study design, the methods applied and the impact of the results obtained. The major strengths and weaknesses associated with the project have been highlighted, with alternative approaches considered where applicable. Further experimental work and scope for development into a new laboratory test have also been included. I am solely responsible for writing this chapter. My supervisors have reviewed this chapter and provided comments, which have been incorporated in the final version.

List of Appendices

Appendix A – Innovation project proposal

The structure of module C of the HSST was amended for Cohort 3 students onwards. The innovation project (formerly C1 module) was no longer required to be submitted as formative assessment but was instead to be included as an appendix in the thesis. The written innovation project piece included as Appendix A puts forward a case for implementing vibrational spectroscopy into the clinical immunology laboratory. This piece of work has been assessed by both my workplace and academic supervisors.

Appendix B – Layman's oral presentation feedback

A formal assessment of the research project proposal, presented to a Layman's audience is required as a part of module C of the HSST. The feedback and assessment outcome has been included as Appendix B.

Appendix C – Confirmation of ethical approval

This study was approved by the ethics committee of the NHS Research Ethics Committee, Health Research Authority (HRA) (IRAS No. 212518). Appendix C is a copy of the approval letter to commence the study.

Appendix D – Research study documentation

This appendix includes copies of the approved study documentation (patient information sheets, consent form, letters of invitation) given to the eligible participants.

Appendix E – Summary of results from A modules of DClinSci

This research project and DClinSci Thesis forms module C of the HSST programme provided by the University of Manchester. Results for the taught module A, and summary of previously assessed module C components for this programme are included as Appendix E.

Appendix F – Qualifications

This appendix includes evidence of the formal qualifications and work-based components required for the completion of the HSST. This includes passing The Royal College of Pathologists (RCPath) examinations in Clinical Immunology (Part 1 - Written, Part 2 - Practical and Part 2 - Oral) and completion of the five domains of the Standards of Proficiency. This thesis also forms the research component required for completion of the RCPath Part 2 examinations.

Appendix G – Supplementary Information for published manuscript (Chapter 3)

This appendix contains the supplementary information associated with Chapter 3. This document was submitted to the journal Nature Scientific Reports, alongside the research

manuscript ‘New Approach to Investigate Common Variable Immunodeficiency Patients Using Spectrochemical Analysis of Blood’. References to the Figures and Tables within this document are included in the text of the published manuscript (Chapter 3).

Appendix H – Additional Research undertaken during HSST

This appendix contains a summary of a COVID-19 research project ‘EXCOVIR’ (Exploring COVID-19 specific immune responses in acute and convalescent phases of infection) to which I have provided a major contribution to during the final 18 months of HSST. As a Clinical Scientist completing the Life Sciences pathway, research and development will continue to be an important aspect of my role going forward.

References

1. Callery, E. L. & Rowbottom, A. W. Vibrational spectroscopy and multivariate analysis techniques in the clinical immunology laboratory: a review of current applications and requirements for diagnostic use. *Applied Spectroscopy Reviews* 1–30 (2021) doi:10.1080/05704928.2021.1958337.

2. Boyle, J. M. & Buckley, R. H. Population prevalence of diagnosed primary immunodeficiency diseases in the United States. *J. Clin. Immunol.* **27**, 497–502 (2007).

3. Chapel, H. & Cunningham-Rundles, C. Update in understanding Common Variable Immunodeficiency Disorders (CVIDs) and the management of patients with these conditions. *Br J Haematol* **145**, 709–727 (2009).

4. Gunten, S. von *et al.* IVIG pluripotency and the concept of Fc-sialylation: challenges to the scientist. *Nature Reviews Immunology* **14**, 349 (2014).

5. Bacchelli, C., Buckridge, S., Thrasher, A. J. & Gaspar, H. B. Translational Mini-Review Series on Immunodeficiency: Molecular defects in common variable immunodeficiency. *Clinical & Experimental Immunology* **149**, 401–409 (2007).

6. Salzer, U., Warnatz, K. & Peter, H. H. Common variable immunodeficiency - an update. *Arthritis Research & Therapy* **14**, 223 (2012).

7. Li, J. *et al.* Understanding the genetic and epigenetic basis of common variable immunodeficiency disorder through omics approaches. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1860**, 2656–2663 (2016).

8. Picard, C. *et al.* Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J Clin Immunol* **35**, 696–726 (2015).

9. Gathmann, B. *et al.* Clinical picture and treatment of 2212 patients with common variable immunodeficiency. *Journal of Allergy and Clinical Immunology* **134**, 116-126.e11 (2014).

10. Bonilla, F. A. *et al.* International Consensus Document (ICON): Common Variable Immunodeficiency Disorders. *J Allergy Clin Immunol Pract* **4**, 38–59 (2016).

11. Cunningham-Rundles, C. & Bodian, C. Common Variable Immunodeficiency: Clinical and Immunological Features of 248 Patients. *Clinical Immunology* **92**, 34–48 (1999).

12. Cunningham-Rundles, C. The many faces of common variable immunodeficiency. *Hematology Am Soc Hematol Educ Program* **2012**, 301–305 (2012).

13. Resnick, E. S., Moshier, E. L., Godbold, J. H. & Cunningham-Rundles, C. Morbidity and mortality in common variable immune deficiency over 4 decades. *Blood* **119**, 1650–1657 (2012).

14. Chapel, H. *et al.* Common variable immunodeficiency disorders: division into distinct clinical phenotypes. *Blood* **112**, 277–286 (2008).

15. Orange, J. S. *et al.* Use of intravenous immunoglobulin in human disease: a review of evidence by members of the Primary Immunodeficiency Committee of the American Academy of Allergy, Asthma and Immunology. *J Allergy Clin Immunol* **117**, S525-553 (2006).

16. Perez, E. E. *et al.* Update on the use of immunoglobulin in human disease: A review of evidence. *Journal of Allergy and Clinical Immunology* **139**, S1–S46 (2017).

17. Wilder, P. van, Odnoletkova, I., Mouline, M. & Vries, E. de. Immunoglobulin Replacement Therapy is critical and cost-effective in increasing life expectancy and quality of life in patients suffering from Common Variable Immunodeficiency Disorders (CVID): A health-economic assessment. *PLOS ONE* **16**, e0247941 (2021).

18. Bonilla, F. A. *et al.* Practice parameter for the diagnosis and management of primary immunodeficiency. *Journal of Allergy and Clinical Immunology* **136**, 1186-1205.e78 (2015).

19. Orange, J. S., Grossman, W. J., Navickis, R. J. & Wilkes, M. M. Impact of trough IgG on pneumonia incidence in primary immunodeficiency: A meta-analysis of clinical studies. *Clinical Immunology* **137**, 21–30 (2010).

20. Hurst, J. R. *et al.* British Lung Foundation/United Kingdom Primary Immunodeficiency Network Consensus Statement on the Definition, Diagnosis, and Management of Granulomatous-Lymphocytic Interstitial Lung Disease in Common Variable Immunodeficiency Disorders. *The Journal of Allergy and Clinical Immunology: In Practice* **5**, 938–945 (2017).

21. Chase, N. M. *et al.* Use of combination chemotherapy for treatment of granulomatous and lymphocytic interstitial lung disease (GLILD) in patients with common variable immunodeficiency (CVID). *J Clin Immunol* **33**, 30–39 (2013).

22. Ameratunga, R., Woon, S.-T., Gillis, D., Koopmans, W. & Steele, R. New diagnostic criteria for common variable immune deficiency (CVID), which may assist with decisions to treat with intravenous or subcutaneous immunoglobulin. *Clin Exp Immunol* **174**, 203–211 (2013).

23. Piqueras, B. *et al.* Common variable immunodeficiency patient classification based on impaired B cell memory differentiation correlates with clinical aspects. *J Clin Immunol* **23**, 385–400 (2003).

24. Wehr, C. *et al.* The EUROclass trial: defining subgroups in common variable immunodeficiency. *Blood* **111**, 77–85 (2008).

25. Warnatz, K. *et al.* Severe deficiency of switched memory B cells (CD27+IgM−IgD−) in subgroups of patients with common variable immunodeficiency: a new approach to classify a heterogeneous disease. *Blood* **99**, 1544–1551 (2002).

26. Mannina, A. *et al.* Clinical Predictors of a Diagnosis of Common Variable Immunodeficiency-related Granulomatous-Lymphocytic Interstitial Lung Disease. *Ann Am Thorac Soc* **13**, 1042–1049 (2016).

27. Giovannetti, A. *et al.* Unravelling the complexity of T cell abnormalities in common variable immunodeficiency. *J Immunol* **178**, 3932–3943 (2007).

28. Bonilla, F. A. & Geha, R. S. Common Variable Immunodeficiency. *Pediatr Res* **65**, 13R-19R (2009).

29. Castigli, E. & Geha, R. Molecular basis of common variable immunodeficiency. *Journal of Allergy and Clinical Immunology* **117**, 740–746 (2006).

30. Kopecký, O. & Lukešová, Š. Genetic defects in common variable immunodeficiency. *International Journal of Immunogenetics* **34**, 225–229 (2007).

31. Park, M. A., Li, J. T., Hagan, J. B., Maddox, D. E. & Abraham, R. S. Common variable immunodeficiency: a new look at an old disease. *The Lancet* **372**, 489–502 (2008).

32. van de Ven, A. A. J. M., Compeer, E. B., van Montfrans, J. M. & Boes, M. B-cell defects in common variable immunodeficiency: BCR signaling, protein clustering and hardwired gene mutations. *Crit. Rev. Immunol.* **31**, 85–98 (2011).

33. Bergbreiter, A. & Salzer, U. Common variable immunodeficiency: a multifaceted and puzzling disorder. *Expert Review of Clinical Immunology* **5**, 167–180 (2009).

34. Rodríguez-Cortez, V. C. *et al.* Monozygotic twins discordant for common variable immunodeficiency reveal impaired DNA demethylation during naive-to-memory B-cell transition. *Nat Commun* **6**, 7335 (2015).

35. Kienzler, A.-K., Rizzi, M., Reith, M., Nutt, S. L. & Eibel, H. Inhibition of human B-cell development into plasmablasts by histone deacetylase inhibitor valproic acid. *Journal of Allergy and Clinical Immunology* **131**, 1695-1699.e9 (2013).

36. Maggadottir, S. M. *et al.* Rare variants at 16p11.2 are associated with common variable immunodeficiency. *J Allergy Clin Immunol* **135**, 1569–1577 (2015).

37. Park, J. *et al.* Interferon Signature in the Blood in Inflammatory Common Variable Immune Deficiency. *PLOS ONE* **8**, e74893 (2013).

38. Orange, J. S. *et al.* Genome-wide association identifies diverse causes of common variable immunodeficiency. *J Allergy Clin Immunol* **127**, 1360–7.e6 (2011).

39. Paquin-Proulx, D. & Sandberg, J. K. Persistent Immune Activation in CVID and the Role of IVIg in Its Suppression. *Front Immunol* **5**, (2014).

40. Rezaei, N. *et al.* B-cell–T-cell activation and interaction in common variable immunodeficiency. *Human Immunology* **71**, 355–362 (2010).

41. Kasztalska, K., Ciebiada, M., Cebula-Obrzut, B. & Górski, P. Intravenous immunoglobulin replacement therapy in the treatment of patients with common variable immunodeficiency disease: an open-label prospective study. *Clin Drug Investig* **31**, 299–307 (2011).

42. Ibáñez, C. *et al.* Modulating effects of intravenous immunoglobulins on serum cytokine levels in patients with primary hypogammaglobulinemia. *BioDrugs* **19**, 59–65 (2005).

43. Hel, Z. *et al.* Altered serum cytokine signature in common variable immunodeficiency. *J Clin Immunol* **34**, 971–978 (2014).

44. Litzman, J. *et al.* Chronic immune activation in common variable immunodeficiency (CVID) is associated with elevated serum levels of soluble CD14 and CD25 but not endotoxaemia. *Clin Exp Immunol* **170**, 321–332 (2012).

45. North, M. E., Spickett, G. P., Webster, A. D. & Farrant, J. Raised serum levels of CD8, CD25 and beta 2-microglobulin in common variable immunodeficiency. *Clin. Exp. Immunol.* **86**, 252–255 (1991).

46. Polito, R. *et al.* Adiponectin Receptors and Pro-inflammatory Cytokines Are Modulated in Common Variable Immunodeficiency Patients: Correlation With Ig Replacement Therapy. *Frontiers in Immunology* **10**, 2812 (2019).

47. Cols, M. *et al.* Expansion of inflammatory innate lymphoid cells in patients with common variable immune deficiency. *Journal of Allergy and Clinical Immunology* **137**, 1206-1215.e6 (2016).

48. Berbers, R.-M. *et al.* Targeted Proteomics Reveals Inflammatory Pathways that Classify Immune Dysregulation in Common Variable Immunodeficiency. *Journal of Clinical Immunology* **41**, 1–12 (2021).

49. Hultberg, J., Ernerudh, J., Larsson, M., Nilsdotter-Augustinsson, Å. & Nyström, S. Plasma protein profiling reflects TH1-driven immune dysregulation in common variable immunodeficiency. *Journal of Allergy and Clinical Immunology* **146**, 417–428 (2020).

50. Chandramouli, K. & Qian, P.-Y. Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Hum Genomics Proteomics* **2009**, 239204 (2009).

51. Eicher, T. *et al.* Challenges in proteogenomics: a comparison of analysis methods with the case study of the DREAM proteogenomics sub-challenge. *BMC Bioinformatics* **20**, 669 (2019).

52. Bunaciu, A. A., Aboul-Enein, H. Y. & Fleschin, Ş. Vibrational Spectroscopy in Clinical Analysis. *Applied Spectroscopy Reviews* **50**, 176–191 (2015).

53. Martino, D., Ben-Othman, R., Harbeson, D. & Bosco, A. Multiomics and Systems Biology Are Needed to Unravel the Complex Origins of Chronic Disease. *Challenges* **10**, 23 (2019).

54. Chu, X., Zhang, B., Koeken, V. A. C. M., Gupta, M. K. & Li, Y. Multi-Omics Approaches in Immunological Research. *Frontiers in Immunology* **12**, 2312 (2021).

Chapter 2: Method Evaluation

Chapter 2: Method Evaluation

2.1 Background

2.1.1 Evaluation of pre-processing techniques and chemometric methods applicable to ATR-FTIR spectroscopy

ATR-FTIR spectroscopy requires minimal sample preparation, is relatively low-cost, and can be applied to a vast range of sample types (blood, tissue, CSF, urine etc.). However a number of pre-analytical, analytical, and post-analytical factors must be considered prior to translation into the clinical laboratory setting. In order to obtain robust, unbiased, high quality data from vibrational spectroscopy studies towards the development of a new diagnostic platform, a series of method optimisation steps and computational data analysis processes must be undertaken.

A standardised and reproducible approach for sample collection and acquisition of spectral data is paramount; subsequent to this, pre-processing is the next most important step that must be performed prior to further analysis[1]. There are several protocols published in the literature to describe the application of pre-processing steps to biospectroscopy data[2–4]. The minimum recommended steps include cutting, baseline correction and normalisation, for which there are various methods available.

If ATR-FTIR spectroscopy is to become a routine diagnostic technique, it is important for researchers and healthcare professionals to understand the mathematical techniques applied in the conversion of sample absorption spectra to a numerical value. The various potential methods of pre-processing and data reduction algorithms must be assessed during a laboratory's method development phase of ATR-FTIR biospectroscopy, prior to clinical evaluation. On completion of pre-processing method development, further application of multivariate analysis and classification tools can be applied to assess the performance characteristics of the method (sensitivity, specificity, misclassification rate) for the purpose of answering pathological questions.

The unique spectral read-out or 'molecular fingerprint' of a sample relates to its biomolecular constituents (i.e. proteins, lipids, nucleic acids, carbohydrates) and is generated from the vibrations of the chemical bonds in these molecules. Each biological sample analysed using ATR-FTIR spectroscopy can generate a spectrum consisting of

hundreds or thousands of individual data points, depending on the acquired spectral range and resolution. Within this complex data set it is extremely difficult to identify relevant and disease-specific spectral features from the background of high variance (noise)[2]. De-noising methods, such as smoothing, baseline correction and normalisation are extensively applied pre-processing technique in FTIR biospectroscopy.

Several spectral features contribute to the enormous total variance between biological samples; these include wavenumber (peak) distribution, absorbance intensity and shape (area under the peak). The high levels of background variance observed between biological spectra are in part due to the heterogeneous sample matrices within physiological samples.

As almost all of the biological components present in human serum and plasma samples are consistent between individuals, the FTIR spectra will contain numerous shared spectral features (both within and between disease and non-disease groups) which may mask the presence of any unique disease-associated features. In addition to this, relevant spectral differences occurring between healthy and pathological samples are very small, and rarely observable by eye in the raw spectral data[3,5]. In order to visualise and interpret any subtle spectral variances between samples which may have biological or clinical relevance both univariate and multivariate (chemometric) methods are required. Samples with unique (potentially disease-associated) spectral features can then be grouped together in order to build powerful prediction models for subsequent diagnostic or quantitative analysis[6,7].

During the first part of this method evaluation, I will explore the suitability and strength of a number of well-recognised biospectroscopy pre-processing techniques in order to determine the optimum conditions for the research study. There are a range of computational methods that can be applied to biospectroscopy data analysis; however a consensus on a standardised approach has yet to be established. A critical appraisal of the machine learning and statistical techniques available for FTIR biospectroscopy have been published in a review by Trevisan et al.[2], and provides a means for researchers to understand the mathematical algorithms and chemometric processes required for spectral data analysis.

Based on the evidence in the literature and the experience of the research group, the pre-processing techniques evaluated within this study include spectral region extraction (‘cutting’) and three different modes of baseline correction with normalisation (rubberband correction with amide I normalisation, rubberband correction with amide II normalisation and Savitzky-Golay (SG) second-order differentiation baseline correction followed by vector normalisation). The principle of these techniques is described in the methods section.

As biospectroscopy becomes more widely used, standard protocols are being developed however the final choice of data analysis should be guided by the individual parameters and requirements of the study.

2.1.2 Assessment of blood sample type suitability: Serum versus plasma

Blood serum and plasma are ideal sample types for clinical investigations, being easily accessible biofluids easily collected by relatively non-invasive procedures. Whilst serum and plasma are both derived from the liquid compartment of whole blood once the cells have been removed, there are distinct differences in the molecular compositions of each biofluid. Most notably, clotting factors and fibrinogen are removed from blood serum in the process of separating the cells (clot) from the liquid. In plasma these proteins are not removed as the clotting cascade is prevented through the addition of anti-coagulant Ethylenediaminetetraacetic acid (EDTA), centrifugation is therefore routinely used to separate the liquid from the cellular component.

In light of this, plasma has a higher total protein content than serum, however both biofluids contain approximately 60–80 mg/mL, the majority of which is made up of albumins and globulins (approximately 50-60% and 40%, respectively)[8,9]. Prevention of the clotting cascade in EDTA-anticoagulated plasma is due to chelation of metallic ions such as calcium, magnesium, zinc and potassium; this results in serum having higher concentrations of these analytes compared to plasma[10].

Routine laboratory analysis techniques such as enzyme-linked immunosorbent assay (ELISA), nephelometry and turbidimetry often permit the interchangeable use of either blood serum or blood plasma for analysis, measuring discrete biochemical analytes within the sample media which are largely unaffected by the sample matrix of either serum or plasma. However, in a recent study using UHPLC-mass spectrometry, Liu et al.[11] reported

significant differences in the metabolomics profiles of serum and plasma; with 46% of the 216 identified metabolites showing different levels between the biofluids, 44% of which were at higher levels in serum. The authors suggest that these metabolomics alterations are most likely a result of clotting-associated processes (for serum collection) such as the activation of platelets and conversion of fibrinogen in the fibrin clot. This study further reported there to be no significant differences in the majority of amino acids, bile acids and phospholipids. Similar findings were observed in an earlier study by Yu et al.[12], with higher concentrations of metabolites reported in serum. The research group postulated that this could be attributed to the removal of the clotting protein fraction and therefore result in a smaller volume remaining for distribution of the remaining metabolites.

The second aspect of the method evaluation will explore the suitability of using blood serum and blood plasma for ATR-FTIR analysis. As vibrational spectroscopy shares similarities with the omics-based studies in that during the analysis, a spectrum of the complete sample is collected which will be representative of all of the biochemical constituents present within the biofluid. Any chemical changes or modifications that occur during the collection and separation of blood serum or blood plasma will be reflected in the spectrum. In order to examine samples for disease-specific spectral differences, the variances attributed to serum or plasma must be considered.

In this sample type evaluation I will compare the spectra of blood serum and blood plasma to confirm or exclude the presence of statistically significant variance between the sample types, and to determine if either sample type is superior in its ability to separate disease from healthy control samples using both unsupervised and supervised methods of multivariate analysis. I will also assess whether any major spectral variances observed between the two sample types are independent of disease class (i.e. observed in both healthy controls and immune deficiency patients).

The findings from these explorative preliminary studies will be used to validate the choice of materials, methods and data analysis techniques applied to the main chapter of this research, in which ATR-FTIR spectroscopy will be used as a novel analytical platform for the investigation of common variable immunodeficiency (CVID).

2.1.3 Evaluation of environmental influence

Environmental variations (including temperature, humidity, and pressure) and any day-to-day fluctuations relating to instrumentation or user input can have an impact the quality of the spectra acquired. To reduce the impact of these known sources of error, standardised processes for sample collection, sample storage, slide preparation and spectral acquisition were performed throughout the research project. In order to assess the reproducibility of these processes, the final aspect of the method evaluation will investigate whether any inter-assay spectral variance relating to environmental factors can be observed.

2.2 Materials and Methods

2.2.1 Population

This study was approved by the ethics committee of the NHS Research Ethics Committee, Health Research Authority (HRA) (IRAS No. 212518). All participants were recruited at Royal Preston Hospital and samples were collected with informed written consent; all methods were carried out in accordance with relevant guidelines and regulations. The study population included 21 adult (>18 years old) CVID patients and 30 healthy controls (HCs).

CVID patients were clinically diagnosed according to the European Society for Immunodeficiencies and the Pan American Group for Immunodeficiency (ESID/PAGID) (1999) diagnostic criteria[13]. All HCs completed a brief questionnaire to document any past medical history relating to immune disorders or recurrent infections, any current medications and the occurrence of any infections requiring medical attention in the past six months. Individual patient characteristics were not considered during this method development aspect of the study.

2.2.2 Sample Collection and preparation

Whole-blood samples were collected into EDTA-treated or serum gel-separator tubes and centrifuged at 110 x g for 5; extracted supernatant samples were then stored as 0.3 mL aliquots at -80°C until required. Prior to spectroscopic analysis, individual aliquots were thawed; mixed and 50 µL from each aliquot was deposited onto IR-reflective glass slides

(MirrIR Low-E slides; Kevley Technologies). For each study subject (n = 51), 50 µL dried blood spots were produced in duplicate (Figure 2.2.1).

Slides were left to air dry for up to 8 hours before being placed into a desiccator overnight. Once generated, dried blood spot slides were analysed the subsequent day. Double-blind unbiased acquisition of spectra was performed on all 51 samples following the allocation of a randomised unique study number to each subject at the point of recruitment. This process was undertaken for both serum and plasma samples.

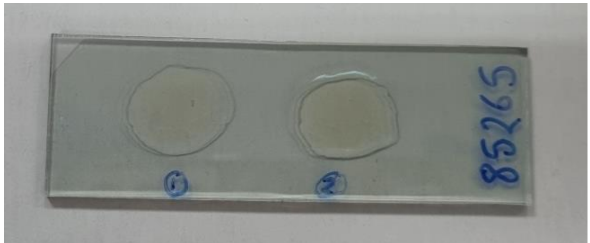


Figure 2.2.1. The generation of dried blood spot slides prior to spectral acquisition. In brief, sample aliquots stored at -80 °C were thawed and mixed; 50 ul of sample was pipetted in duplicate onto IR-reflective glass slides (MirrIR Low-E slides; Kevley Technologies). The samples were then spread out using the pipette tip in a circular motion to achieve two similar sized blood spots and left to air dry on the bench prior to being placed in a desiccator overnight.

2.2.3 ATR-FTIR spectroscopy

Spectra of undiluted serum and plasma samples were obtained using a Tensor 27 FTIR spectrometer with Helios attenuated total reflection (ATR) attachment (Bruker Optics Ltd) operated by OPUS 5.5 software. Spectral acquisition was collected from an approximate sampling area 250 × 250 µm (defined by the diamond crystal surface of the internal reflective element) at a resolution of 8 cm^{-1} with two times zero-filling, giving a data spacing of 4 cm^{-1} over the range 4,000–400 cm^{-1}. In order to reduce the level of interference from fluctuations in atmospheric conditions, a background spectrum was collected prior to each new sample measurement (study subject) and automatically removed from the sample spectra via OPUS 5.5 software. The diamond crystal was cleaned with distilled water and dried between each patient sample and between replicate blood spots. Single point spectral measurements were obtained for each study subject, collecting 10 spectra per 50 µL dried blood spot replicate (total of 20 spectra per biofluid). In order to minimize bias and account for variations in sample thickness or

sample homogeneity, point spectra were collected from 10 different locations across each blood spot.

2.2.3.1 Spectral Acquisition Detail

OPUS 5.5 software was used for all spectral data collection. The Bruker Tensor 27 spectrometer was equipped with a room temperature Deuterated Lanthanum α Alanine doped TriGlycine Sulphate (DLaTGS) detector, mid-IR source (4000 to 400 cm-1) and a potassium bromide (KBr) beamsplitter.

The instrument test, called the Performance Qualification Test (PQ test) was performed every day to prove that the spectrometer was measuring correctly. An advanced measurement file was set up for the serum and plasma samples, with a specified file name and location for storage of the spectral data. The following parameters were saved within the experiment settings: Resolution – 4 cm^{-1}; Sample Scan Time – 32 Scans; Background Scan Time – 32 Scans; Save Data from - 4000 cm^{-1} to 400 cm^{-1}; Result Spectrum – ATR spectrum; Data blocks to be saved – ATR Spectrum, Single Channel, Background.

Prior to each experiment, the 'Check Signal' tab was used to confirm and record the peak position and amplitude of the interferogram. The purpose of checking the signal is to either achieve a maximum interferogram signal or to optimize the shape of the single-channel spectrum.

A single beam background spectrum was collected prior to each sample replicate (every 10 spectra collected). Using the 'Basic' tab, 'Background Single Channel' was selected, ensuring that there was no sample slide in the measurement beam path underneath the ATR crystal. A background spectrum was also collected if there were any changes in atmospheric conditions, for example a door opening or a temperature change.

Once checks were completed, the dried blood spot slides were added to the beam path and visualised using the instrument digital camera, this aided in locating a region of interest (avoiding the edges of the blood spot) prior to bringing into direct contact with the ATR crystal. Once the region was selected, the sample stage was moved up to the crystal and the pressure was increased until a suitable preview image of the spectra could be observed on the screen. Using the 'Basic' tab, a spectrum was acquired by clicking 'Sample Single Channel'. Each spectrum took approximately 2 minutes to acquire. Once

complete, the sample was moved away from the crystal; the crystal was cleaned with distilled water and dried with absorbent cloth prior to moving the sample stage into a new location to repeat the spectral acquisition 10 times. Consistent pressure was applied when collecting the spectra from each location. Files were saved in OPUS format prior to loading into the analysis software, MATLAB R2017a (Mathworks).

2.2.4 Pre-processing techniques

Analysis of the spectral datasets was performed using the IRootLab toolbox (trevisanj.github.io/irootlab/), within MATLAB R2017a software (MathWorks), unless stated otherwise. The major objectives from the application of computational methods (pre-processing and multivariate statistical analysis) are to determine class similarities and differences (i.e. HC vs CVID, or serum vs plasma) within the population dataset, in addition to identifying specific wavenumber peaks that are attributed to the most relevant differences[4].

The Fingerprint region, between wavenumbers 1800-900 cm^{-1}, which includes known wavenumbers of interest for biological samples, was initially extracted ('cut') from the dataset. Subsequently, pre-processing steps were performed in order to improve the quality of the data and optimise further analysis. The output of major pre-processing techniques assessed during this method development stage were; i) rubberband baseline correction followed by either amide I, amide II or vector normalisation; and ii) Savitzky-Golay (SG)[14] second-order differentiation baseline correction followed by vector normalisation.

Baseline corrected spectra are obtained by subtracting the baselines from the original spectra. Rubberband correction is a popular technique applied to FTIR data that stretches the spectra down by simulating a rubber band. In brief, the rubberband baseline is drawn by finding the convex polygonal line whose edges trace the minimal (trough) signals throughout the region of interest. This is then subtracted from the original spectrum to eliminate slopes.

Normalisation to amide I or amide II applies a scaling factor so that all spectra have the same absorbance intensity at the amide I or amide II peak, respectively. The disadvantage of these normalisation methods are that they eliminate these peaks as biomarkers, as any

spectral variance between samples will be eliminated at either the amide I- or amide II-associated wavenumbers.

Vector normalization is typically applied after differentiation-based baseline correction methods but can be applied more widely. As this normalisation technique does not require a reference peak as amide I/II normalisation does, it can be advantageous if the amide I/II peaks are to be used as potential biomarkers[3].

Smoothing techniques are commonly applied during the analysis of digital data. The use of Savitzky-Golay (SG) smoothing[14] is a well described technique within spectral data analysis as a means to increase precision when detecting relevant wavenumber peaks. The SG filter is an improved version of a moving average filter. This is achieved by performing least-squares fit of a small consecutive set of data points to a polynomial of a given degree, and taking the central point (filter coefficient) of the fitted polynomial curve as the new smoothed data point. The aim of the SG digital filter is to improve the signal-to-noise ratio by eliminating random error within the spectrum. The limitation of SG smoothing includes the potential to smooth out relevant peaks or amplify unwanted signals therefore care must be taken when selecting the tuning parameters [2].

Polynomial fitting and differentiation-based baseline correction methods are often combined with Savitzky-Golay (SG) smoothing to improve signal-to-noise ratio. These techniques will account for several causes of a skewed baseline including scattering, reflection, temperature fluctuation, sample thickness and instrumental issues[15]. Lower order polynomials are well suited for background correction in FTIR spectroscopy.

Applied in combination with SG smoothing, second order derivatives can efficiently correct the background signal present in the FTIR spectrum that can be accentuated from taking the derivative of the spectral data points[2]. Whilst this method can successfully eliminate slopes and resolve overlapping bands, applying a second-order differentiation algorithm adds complexity to the visual interpretation of spectral data as both positive and negative values are plotted on the y-axis (as compared to rubberband baseline correction which has only positive values on the y-axis).

2.2.5 Multivariate analysis

Principle component analysis (PCA) is a widely used unsupervised multivariate technique which provides a mathematical and statistical means of identifying maximally relevant biochemical information, whilst retaining the majority of the original variance in the dataset.

Each PC is composed of scores and loadings, the scores are used to identify clustering patterns among the samples and the loadings to identify the main wavenumbers responsible for class differentiation. In order to obtain the loading vectors, commonly referred to as principle components (PCs), the original data is linearly transformed and ordered such that the first PC retains the maximal variance and each succeeding component retains the next highest variance. Each successive PC is orthogonal to the others (i.e. forms an angle of 90 with the PCs). This enables the data to be visualised in a reduced dimensional space and any relationship between classes can be identified with less difficulty[16].

Application of this mathematical algorithm can reduce a complex spectral data matrix ($\mathbf{X}$) (potentially containing over thousands of data points), to a much fewer number of relevant variables termed principal components (PCs), composed of scores ($\mathbf{T}$) and loadings ($\mathbf{P}$) for final interpretation. This decomposition is performed as follows [16]:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}$$

(Equation 1)

where $\mathbf{E}$ represents the residuals (non-explained spectral variability), and the superscript T represents the matrix-transpose operation.

PCA has weak discriminatory power as it is an unsupervised method of multivariate analysis; however it can also be used as a pre-processing step prior to supervised methods of multivariate analysis, such as linear discriminant analysis (LDA). Throughout this method development study, exploration of any spectral variance between classes was initially assessed by comparing the scores after PCA. As a means of supervised multivariate analysis, principal component analysis linear discriminant analysis (PCA-LDA) was used to observe inter-group differences by means of a linear discriminant function applied to the PCA scores[17].

The PCA-LDA scores [$cf(\mathbf{t}_i)$] are calculated as follows [17]:

$$cf(\mathbf{t}_i) = (\mathbf{t}_i - \bar{\mathbf{t}}_k)^{\mathrm{T}}\mathbf{C}_{\mathrm{pooled}}^{-1}(\mathbf{t}_i - \bar{\mathbf{t}}_k)$$

(Equation 2)

where $\mathbf{t}_i$ is the PCA scores for a given sample i; $\bar{\mathbf{t}}_k$ is the mean scores vector for class k; $\mathbf{C}_{\mathrm{pooled}}$ is the pooled covariance matrix; and T denotes the matrix transpose operation.

As recommended by Kelly et al.[15] in the literature, 10 PCs were chosen for inclusion in the PCA-LDA analysis. The suitability of using 10 PCs was further confirmed by looking at the Pareto charts; whilst choosing too few PCs may omit important spectral information, the inclusion of increasing numbers of principle components (past the point at which the amount of explained variance begins to plateau) will increase the amount of noise and may lead to overfitting.

Visualisation of supervised multivariate analysis output was done using scores plots and cluster vector plots. Scores plots are scatter diagrams used to assess class separation, whereas cluster vectors plots provide a means to identify which wavenumbers are responsible for the variance in the datasets and thus the observed separation[18].

To construct a PCA-LDA scores plot, the calculated PCA-LDA projected scores (equation 2) from the spectral data within each class are plotted on the scatter diagrams. The scores plots display each spectrum as a point in multidimensional space, with the number of dimensions (D) determined by the number of classes being compared (n) minus 1, (*n*-1). In this evaluation only two classes were compared during each analysis (HC vs CVID; or Serum vs plasma) therefore the scores plots were always displayed in 1D.

The spectral variation between the two classes can be interpreted visually by how far apart the points contributing to each class are positioned in the scores plot; overlapping scatter plots indicate spectral similarities between classes. A one-way ANOVA was used to calculate p-values for PCA-LDA scores where statistical significance was evaluated at a 95% confidence level ($P < 0.05$).

The cluster vector plots are constructed following PCA-LDA data reduction, and the creation of loadings vectors for each class. Each vector points from the origin of the PCA-LDA factor space to the centre of its corresponding class .The cluster vector plots create

pseudo-spectra, in that whilst they are associated with wavenumbers, the loadings vectors that are constructed for each class are in fact linear combinations of variables from LDA that 'pass through' respective data points (wavenumbers) instead of pointing towards void space[15].

The loadings plots and cluster vector plots allow the user to identify which wavenumbers are responsible for the major spectral variances between classes. The absorbance intensity associated with the identified wavenumbers can subsequently be measured on the original FTIR spectrum; this enables the absorbance intensity change to be used as a potential diagnostic biomarker, i.e. either increased or decreased in the disease population compared to the control population or reference absorbance intensity.

In this method development section, construction of the classic cluster vectors was done using the IRootLab toolbox, for which the index of the class to be the origin of the PCA-LDA factors space was left as zero (ignored)[18], an alternative approach is to use the centre of one of the classes as the origin, resulting in its corresponding cluster vector being a flat line in the centre. The cluster vectors generated from the 'test' classes can then be plotted from this index, allowing key wavenumber variance and potential biochemical changes to be identified.

Verification of pre-processing techniques

In this section of the method development study, the strength of each pre-processing technique in its ability to reduce noise and optimise the spectral data quality was evaluated by calculating and comparing the standard deviation of the mean spectra for each class (CVID patients versus healthy controls; and serum versus plasma). Subsequent to this, each pre-processing method was then put through multivariate analysis (unsupervised and supervised) to evaluate whether the choice of pre-processing technique could improve or reduce the ability to identify key spectral variances and/or discriminate between classes. A one-way ANOVA was used to calculate p-values for PCA-LDA scores where statistical significance was evaluated at a 95% confidence level ($P < 0.05$).

Assessment of blood sample type suitability: Serum versus Plasma

Suitability of blood sample type was assessed through performing supervised and unsupervised methods of analysis on the spectral data. Pre-processing methods selected during previous method evaluation section were applied as described above. PCA scores and PCA-LDA projected scores were calculated as described in multivariate analysis methods section, and assessed for significance using a one-way ANOVA ($P < 0.05$), as above. The discriminating biomarkers extracted following PCA-LDA analysis of serum versus plasma samples were obtained from the cluster vector analysis. A peak detection algorithm was applied to identify the 12 most segregating peaks.

Evaluation of environmental stability

Assessment of inter-assay variability associated with environmental factors and acquiring spectral data across several different days was performed by grouping samples into classes based on the date of spectral acquisition, and examining for significant spectral variance between the classes. This was done independently for both HC and CVID patient samples. Pre-processing and multivariate analysis was performed using the MATLAB R2014b software (MathWorks, Inc., USA) through the PLS Toolbox version 7.9.3 (Eigenvector Research, Inc., USA) and lab-made routines. Firstly, pre-processing of the acquired spectra was performed to remove physical interferences and improve the signal-to-noise ratio. This was made by 'cutting' the fingerprint region (1800–900 cm^{-1}), followed by Savitzky-Golay 2nd derivative (window of 5 points, 2nd order polynomial fitting) and vector normalisation. The PCA scores for each spectrum were used to assess environmental stability. P-values were calculated based on a MANOVA test of the scores within PC1, PC2 and PC3. Statistical significance was evaluated at a 95% confidence level ($P < 0.05$).

2.3 Results & Discussion

2.3.1 Verification of pre-processing techniques – Results

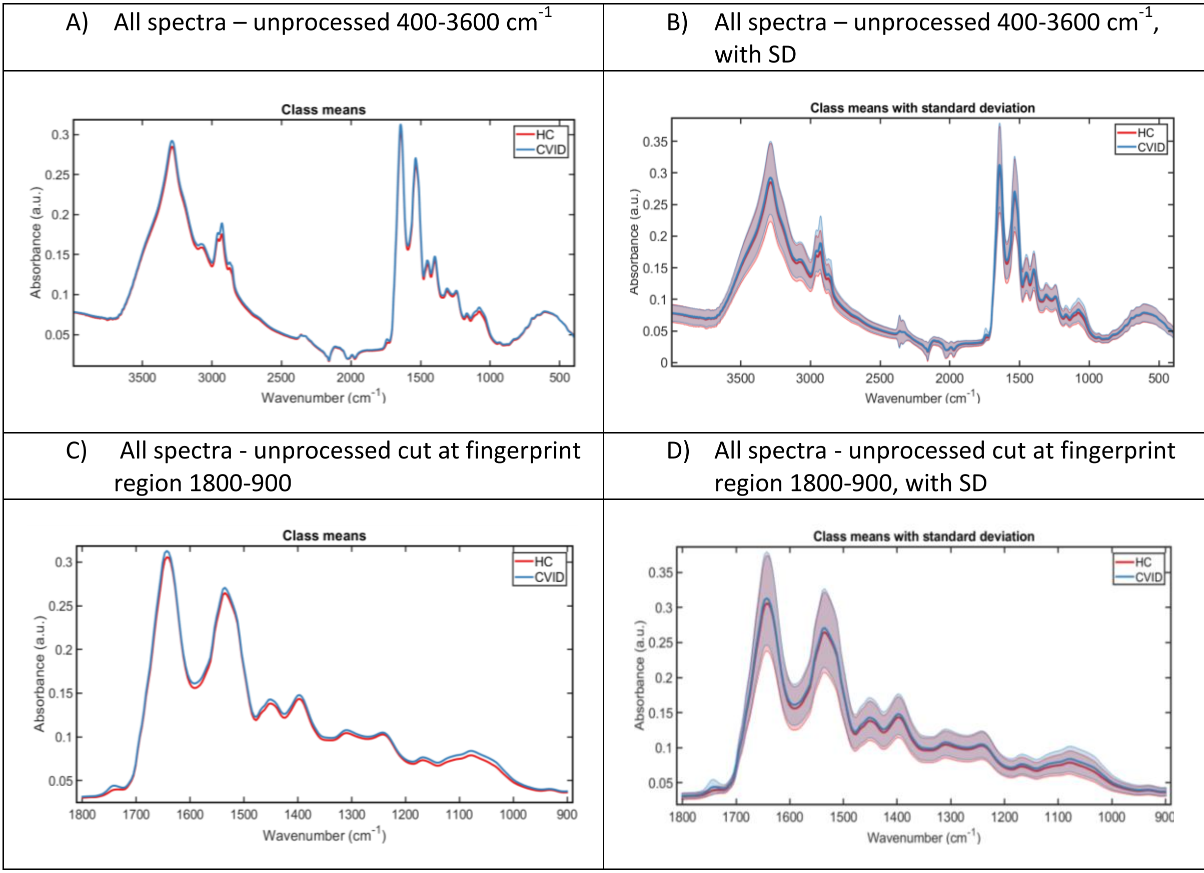


Figure 2.3.1. Visualisation of the raw serum spectra for CVID patients and HCs. Unprocessed serum spectra for HC and CVID patients (mean spectra) with and without standard deviation (SD) displayed.

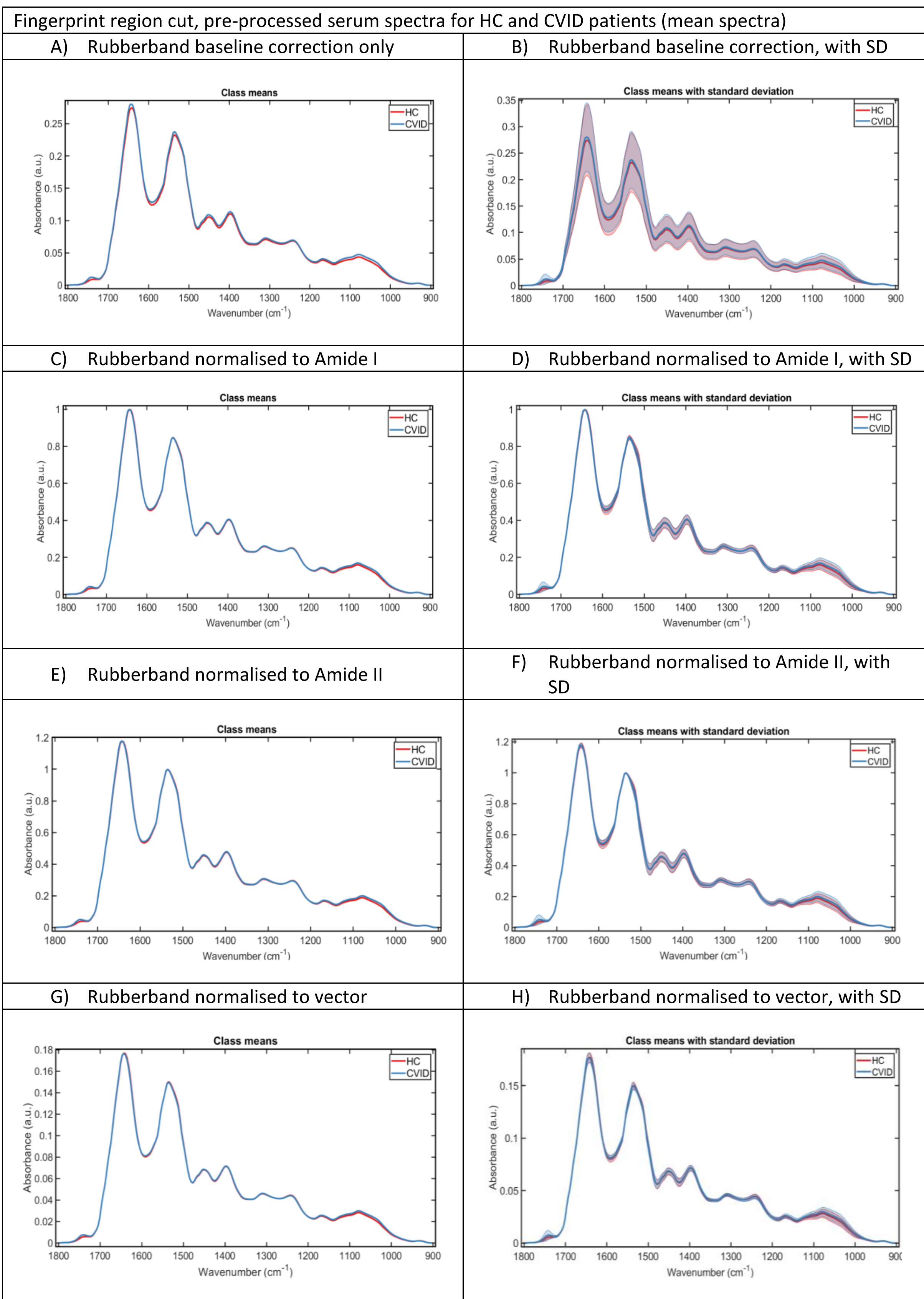


Figure 2.3.2. Visual examination of rubberband baseline corrected mean spectra and associated standard deviations following Amide I, Amide II and vector normalisation techniques with and without standard deviation (SD) displayed.

In Figure 2.3.1A and B, the raw spectra of serum samples from 30 HC and 21 CVID patients are illustrated (plotting the class mean spectra for the wavenumber region 3600-400 cm-1). These spectra were initially cut at the fingerprint region, 900-1800 cm^{-1}, (Figure 2.3.1C and D), reducing the size of dataset for the evaluation of pre-processing techniques.

Visually, the spectra from each class appear very similar therefore the application of multivariate statistical analysis will be required to further investigate whether FTIR spectroscopy can be used to discriminate between serum samples from each class. Analysis of the variance of absorbance intensities within the spectral dataset was performed (using standard deviations (SD)). As expected for unprocessed data, higher SD was observed, depicted by the grey shaded regions across the spectra in Figures 1B and D, and the first two columns of Table 1.

Table 2.3.1. Standard deviations of mean spectra for each group following rubberband and second-differential (SG smoothed) pre-processing of serum samples.

	None	Cutting (900 – 1800 cm^{-1})					
		Cut only	Rubberband Baseline correction				2nd Differential SG
			+ Normalisation method:				
			None	Amide I	Amide II	Vector	Vector
HC	0.0181	0.0212	0.0181	0.0157	0.0176	0.0023	0.0054
CVID	0.0176	0.0213	0.0180	0.0161	0.0184	0.0024	0.0050

Figure 2.3.2, and Table 2.3.1, illustrates the mean spectra and SDs following rubberband and second-differential (SG smoothed) baseline correction and subsequent normalisation using three methods (Amide I, Amide II and Vector). Once again, visual interpretation of the spectra revealed high similarities between the pre-processed spectra for the two classes, however the SDs, represented by the shaded regions, are greatly reduced. This indicates that the reproducibility of the data is greater following baseline correction and normalisation due to the reduction of noise and correction of interference from sample or environmental factors.

Table 2.3.1 shows the mean SD calculated across each wavenumber in the spectral dataset for each class. In comparison to the unprocessed, cut spectral data for HC and CVID samples (both at 0.021), the SDs are lower for all pre-processed datasets, improving reproducibility. Between each class (HC and CVID), the SDs associated with each pre-processing technique are comparable, indicating that a similar level of variance

(potentially attributed to environmental and sampling variations) is inherent to the collection of spectra from both classes.

For the rubberband baseline corrected spectra, the normalisation method which showed the lowest average SD for both CVID and HC spectral data was the vector normalisation method (at 0.0024 and 0.0023 respectively). SDs of the mean spectra were also calculated following second-differential (SG smoothed) baseline correction with vector normalisation of the CVID and HC spectral data (at 0.0050 and 0.0054, respectively), and included in Table 2.3.1.

For polynomial based methods of baseline correction, vector normalisation is the only suitable normalisation method that can be applied. Following this initial analysis, the two pre-processing methods chosen to take forward for subsequent multivariate analysis were rubberband baseline corrected spectra with vector normalisation and second-differential (SG smoothed) baseline correction with vector normalisation.

The variance and segregation of serum samples from each class (HC and CVID) were investigated by PCA and PCA-LDA following two different methods of pre-processing (rubberband, vector normalised and second differential, vector normalised) to determine the best possible separation techniques (illustrated in Figures 2.3.3 and 2.3.4). The number of PCs used was optimised using the Pareto function in the IRootLab toolbox to capture maximal variance whilst minimising inclusion of noise.

Figure 2.3.3A illustrates 98.9% of the variance to be captured within the first 10 PCs; we therefore limited subsequent analysis to a maximum of 10 PCs, dramatically reducing the size of the dataset prior to further exploratory analysis. During PCA cluster analysis, each spectrum can be plotted in either 2D or 3D within the linearly transformed space. The use of multiple plots enables improved visualisation and selection of the PCs responsible for separating the classes.

Figure 2.3.3B and 2.3.3C shows the scatterplots generated from plotting the first three PCs against each other, which together are responsible for 84% of the spectral variance. In figure 2.3.3B, the projected axis drawn using PC3 shows superior separation of the two classes compared to PC1 and PC2. Whilst PCA is useful to gain an initial overview of the spectral data and assess the presence or absence of any clear class separation, this mode of multivariate analysis has weak discriminatory power. Supervised PCA-LDA analysis

(using the first 10 PCs) was therefore performed to explore class segregation following rubberband vector normalised pre-processing.

The scores plot in Figure 2.3.3D demonstrates significant segregation of classes following PCA-LDA ($P<0.0001$). Second-order differential baseline correction is a widely accepted pre-processing technique for biological samples and was also considered within this study. The variance and segregation between classes was assessed using PCA and PCA-LDA, respectively, as previously described.

Figure 2.3.4A confirms that inclusion of the first 10 PCs is also appropriate for the second-differential, vector normalised pre-processed spectra. In Figure 2.3.4B and 2.3.4C, PCA analysis using the first three PCs (capturing 70% of the total variance) revealed PC2 has the greatest ability to separate classes. The scores plots derived from subsequent supervised analysis (PCA-LDA) demonstrated improved and significant ($P<0.0001$) class separation, as illustrated in figure 2.3.4D.

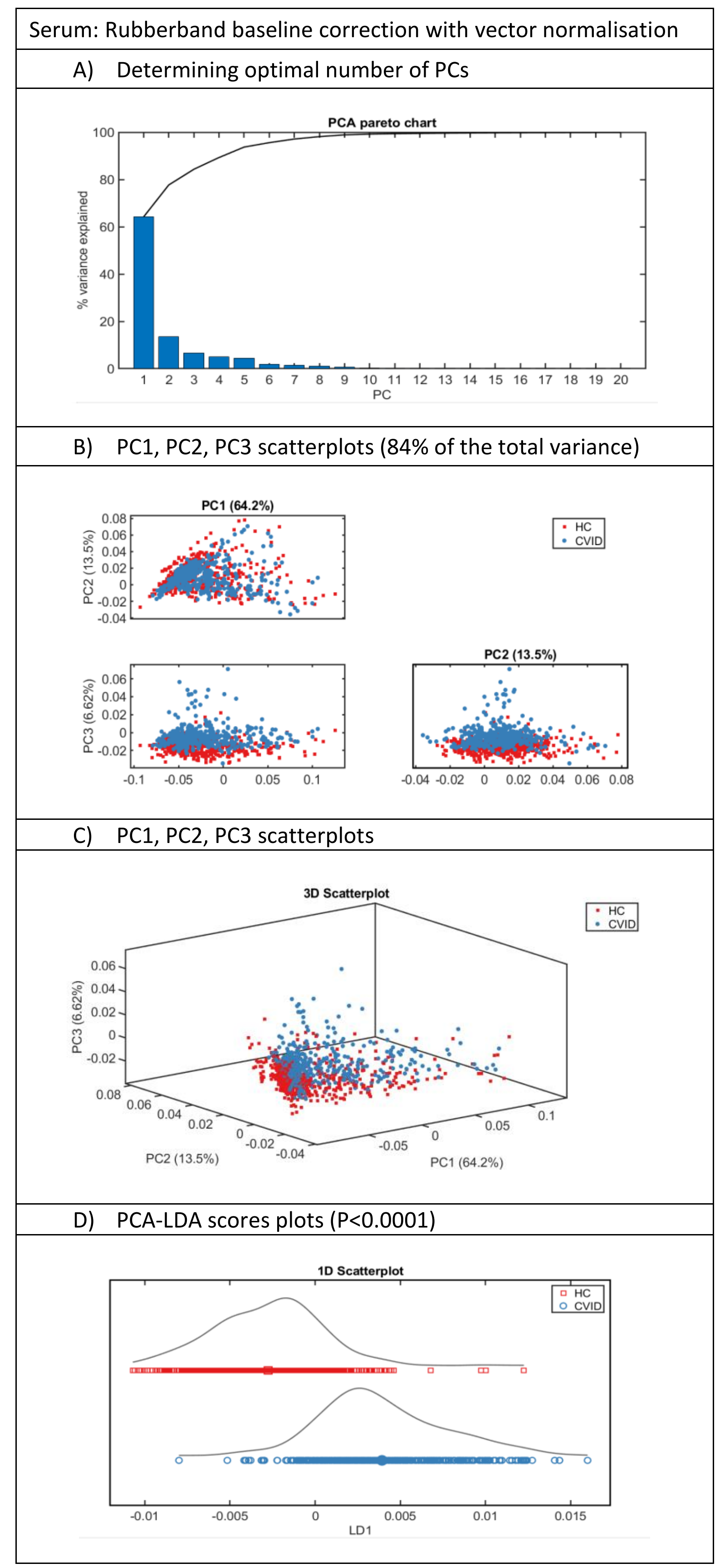


Figure 2.3.3. Examining the effect of pre-processing techniques on subsequent multivariate analysis techniques (unsupervised and supervised separation).

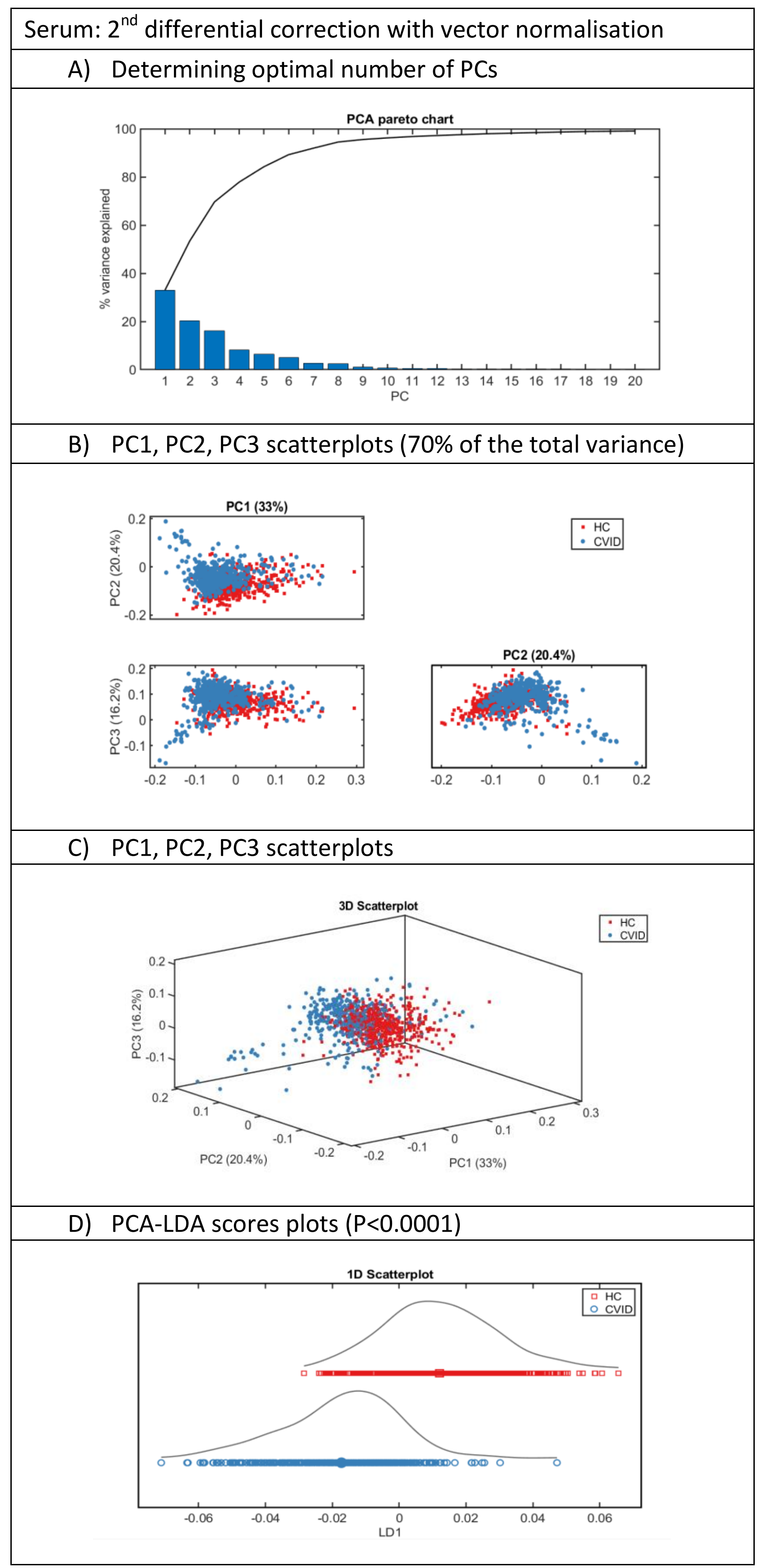


Figure 2.3.4. Examining the effect of pre-processing techniques on subsequent multivariate analysis techniques (unsupervised and supervised separation).

2.3.2 Verification of pre-processing techniques – Discussion

Pre-processing techniques are required prior to chemometric analysis in order to correct for unwanted signals (background noise) such as sloping baseline effects, fluorescence, scattering, variations in sample thickness, environmental conditions and instrumental variations[2,15]. These unwanted signals cause oscillations in the spectrum, producing false absorption bands that do not represent the sample composition.

The effects of the major pre-processing techniques (cutting, de-noising, baseline correction and normalisation) were explored within this study. Visualisation of the mean class spectra (with standard deviations), alongside the average calculated standard deviations, illustrated high levels of variance in the original unprocessed data which could be successfully reduced following application of these widely-used pre-processing techniques. As the most important spectral regions measured in biological samples are typically found within the fingerprint region, all spectra were firstly cut to extract this dataset (1800-900 cm^{-1}), which also includes the amide I and amide II regions (1500–1700 cm^{-1}).

The standard deviation calculated for both HC and CVID groups increased following cutting of the spectrum and extraction of the fingerprint region. This was likely a result of reducing the wavenumber range of the spectrum. In focussing analysis on a smaller specific wavenumber range, several wavenumber regions which would have exhibited very low absorbance signals ('dead' areas) would be eliminated. These dead areas would have had small levels of variation attributed with them thus the increased SD associated with cutting is likely a concentration effect once these regions were removed. In addition, it is known that the fingerprint region contains the spectral data for the majority of biologically relevant molecules so we would expect higher levels of variation, both within-class and between-class across this region.

Following application of rubberband baseline correction and normalisation to amide I, amide II or vector, the variation across the spectral data is reduced for both HC and CVID patient groups. These observations are similar to those reported in other studies using ATR-FTIR spectroscopy on human serum[19,20], with Hands et al. reporting reproducibility data (average SD of 0.014 and 0.0015, for unprocessed and rubberband vector-normalised spectra, respectively).

Amide I normalisation is an accepted method of normalisation for biological samples and reduces the intra-class variance compared to the cut, unprocessed data for both sample groups. However it is unknown at this stage in our investigations whether the contribution of Amide I absorbance to group-specific variance is significant, (thus a potential disease-specific biomarker). As normalisation to the amide I peak will eliminate variance between the classes at that wavenumber (1650 cm^{-1}), vector normalisation is concluded to be more suitable. Furthermore, vector normalisation is the only choice of normalisation for polynomial-based background correction techniques therefore was chosen for subsequent analysis for both rubberband and second-order differentiation pre-processing techniques prior to the application of unsupervised and supervised multivariate analysis.

The ability to separate the serum sample spectra into the two study groups (HC and CVID) was examined using unsupervised (PCA) and supervised (PCA-LDA) analysis techniques for both the rubberband- and second-order differentiated-vector normalised data. The pareto charts obtained for both sets of pre-processed data illustrates that 98% of the sample variance is captured within the first ten PCs, this is in keeping with previously reported studies using biological samples[21].

As an unsupervised exploratory data analysis, the group labels or class information (i.e. HC or CVID) is omitted from PCA to prevent over-fitting of the data and introduction of bias. The goal of PCA is to reduce the complexity of a dataset by producing a fewer number of independent variables (called loadings vectors), while retaining the maximum levels of variation present in the original data. This enables the data to be visualised in a reduced dimensional space and any relationship between classes can be identified with less difficulty[16].

Visualisation of the scatterplots from the rubberband- and second differential-baseline corrected spectra was done using PC1, PC2 and PC3 (Figures 2.3.3B and 2.3.3C; and 2.3.4B and 2.3.4C) which accounted for 84% and 70% of the total variance, respectively. These plots illustrate that the spectra from each class are clustered together to some degree, and that prior to any consideration of class within the analysis, there is evidence of within-class similarities and between-class differences. As an unsupervised multivariate method, PCA has weak discriminatory power, illustrated in the scatterplots by the high level of overlap between class cluster, which limits its use in classification and prediction

models. This can be overcome by combining PCA with more powerful supervised methods such as linear discriminant analysis (LDA). PCA will therefore be applied as a further pre-processing step for subsequent computational analysis and data mining algorithms used in the main body of the project.

In conclusion, both rubberband- and second-order differentiation methods for baseline correction are accepted in the field. As demonstrated in this method development section, both methods are vital pre-processing steps to improve the quality of the spectral data and increase the signal-to-noise ratio prior to further multivariate analysis. Encouragingly, in this initial piece of exploratory work, application of both methods demonstrated clear separation between classes when PCA-LDA multivariate analysis was applied to the pre-processed spectra from serum samples.

As the rubberband-vector normalised technique does not contain a smoothing step and that second-order differentiation correction is considered to be a more powerful technique (it increases the amplitude of discriminant spectral features between the classes), the second-differential baseline correction (with SG smoothing) followed by vector normalisation will be the pre-processing method applied to the subsequent classification work undertaken for this research study. These findings support the continuation of further analysis and development of classification models to potentially generate a novel diagnostic platform for CVID.

2.3.3 Investigating choice of biofluid - Results

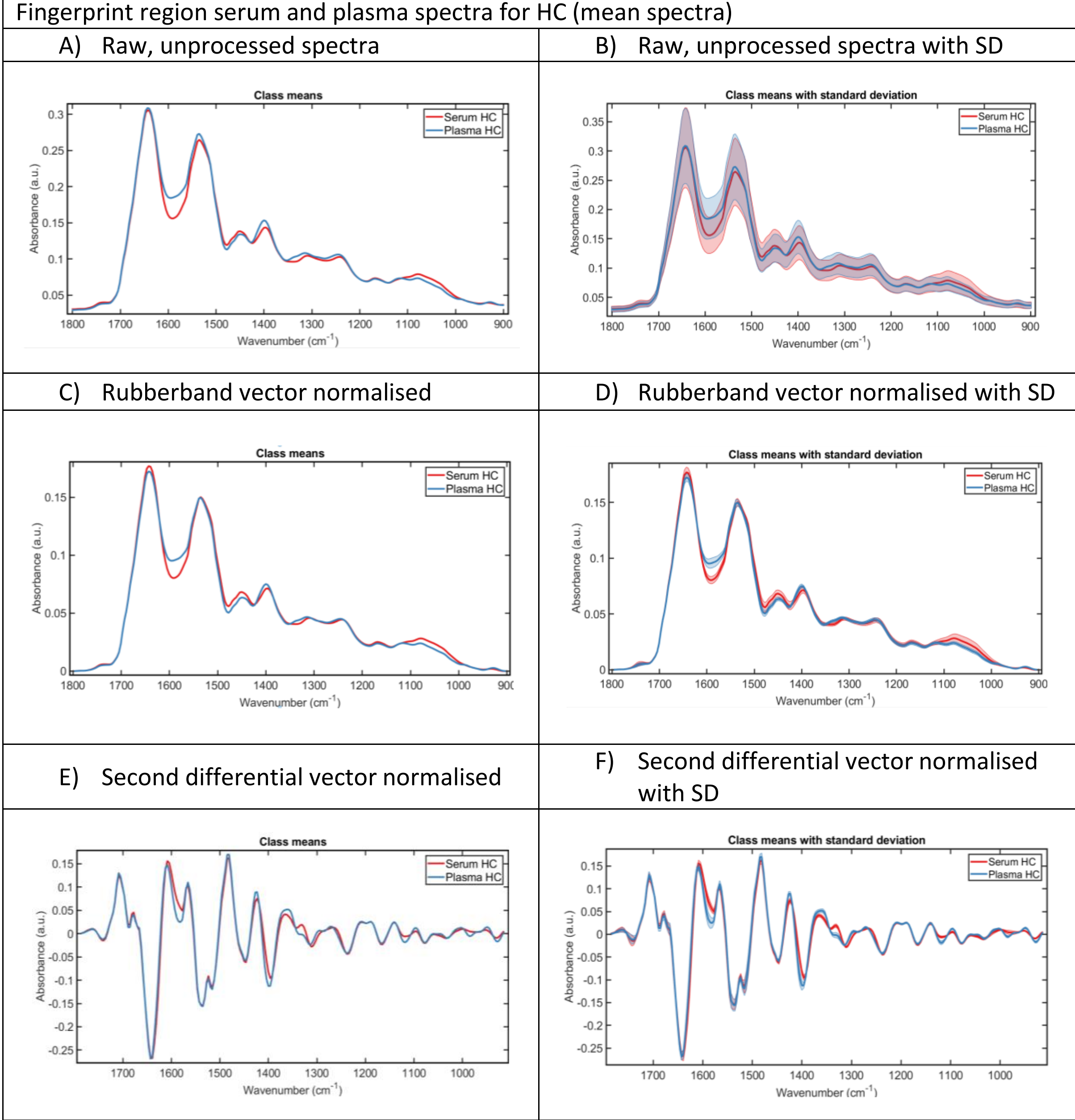


Figure 2.3.5. Visualisation of raw and pre-processed spectra for blood serum and plasma in healthy controls (mean spectra) with and without standard deviation (SD) displayed.

To further validate and confirm the observations from the previous pre-processing techniques evaluation, the mean absorbance spectra and associated standard deviations of the serum and plasma samples from healthy controls were visualised (Figure 2.3.5A-F). The results mirrored the earlier findings in that the standard deviation (shaded areas) of the spectra within each class is reduced following the application of well-recognised approaches to baseline correction and normalisation processes. The chosen methods to take forward within the previous section were rubberband- and second-order differential

followed by vector normalisation, both of which result in improvements in the overall spread of the spectral data for each class, as illustrated by the reduction in shaded areas associated with the mean class spectra in Figure 2.3.5 (B, D and F).

In this initial visualisation and comparison of the mean spectra from HC serum and plasma samples (Figure 2.3.5A-F) it is noteworthy that there are fewer overlapping areas and greater peak separation between the classes (serum vs plasma) when compared to the previous method section, which compared classes (HC vs CVID) within the same biofluid type (i.e. serum or plasma). This suggests that even prior to the application any further unsupervised or supervised multivariate analysis to identify subtle variances, there are significant spectral differences evident between the two biofluids.

To explore these differences further, an unsupervised comparison of the mean spectra (cut to the fingerprint region) from HC serum and plasma biofluids was performed on second differential vector normalised pre-processed spectra. Figure 2.3.6A represents the PCA scatterplots generated using the first three PCs (includes 82% total variance). The clustering observed within the same class, and separation between the two classes (serum = red; plasma = blue) is clearly illustrated within PC1, and to some degree within PC2. This indicates that there are fundamental differences between the spectra of each class and that the two sample types could not be used interchangeably if biospectroscopy methods were translated into the routine pathology laboratory.

The findings from the unsupervised analysis warrant further investigation to determine which wavenumbers (and associated biomolecular assignments) account for the major variances observed. A supervised multivariate method (PCA-LDA) was used to derive the scores plots and corresponding cluster vector plots for the two sample types. Figure 2.3.6B illustrates clear and statistically significant separation between the classes on the PCA-LDA scores plots ($p < 0.0001$).

The PCA-LDA loadings plot indicates which variables have the largest effect on the class separation and are responsible for clustering (as described in the methods section). As there are only two classes being compared, the cluster vector plots are generated from only one LDA loadings plot (Figure 2.3.6C). This results in the cluster vector plots (Figure 2.3.6D) being a mirror image of each other. The application of a peak detector to the loadings and cluster vector plots identifies the twelve most prominent peak variances

(positive or negative) between the classes, and thus can be used to identify the molecular groups which contribute to the major variances between the two classes.

For the healthy control group, the twelve most prominent wavenumber peaks that distinguish between the serum and plasma biofluids are labelled on the cluster vector plot in Figure 2.3.6D and described in Table 2. The wavenumbers (and associated molecular assignments) are 926 cm^{-1} and 984 cm^{-1} (protein phosphorylation vibrations), 1041 cm^{-1} (glycogen), 1119 cm^{-1} (C-O stretching vibrations), 1304 cm^{-1} (amide III), 1354 cm^{-1} and 1393 cm^{-1} (COO- symmetric stretching vibrations of fatty acids and amino acids), 1481 cm^{-1} (proteins), 1531 cm^{-1} (amide II), 1589 cm^{-1}, 1624 cm^{-1} and 1666 cm^{-1} (amide I).

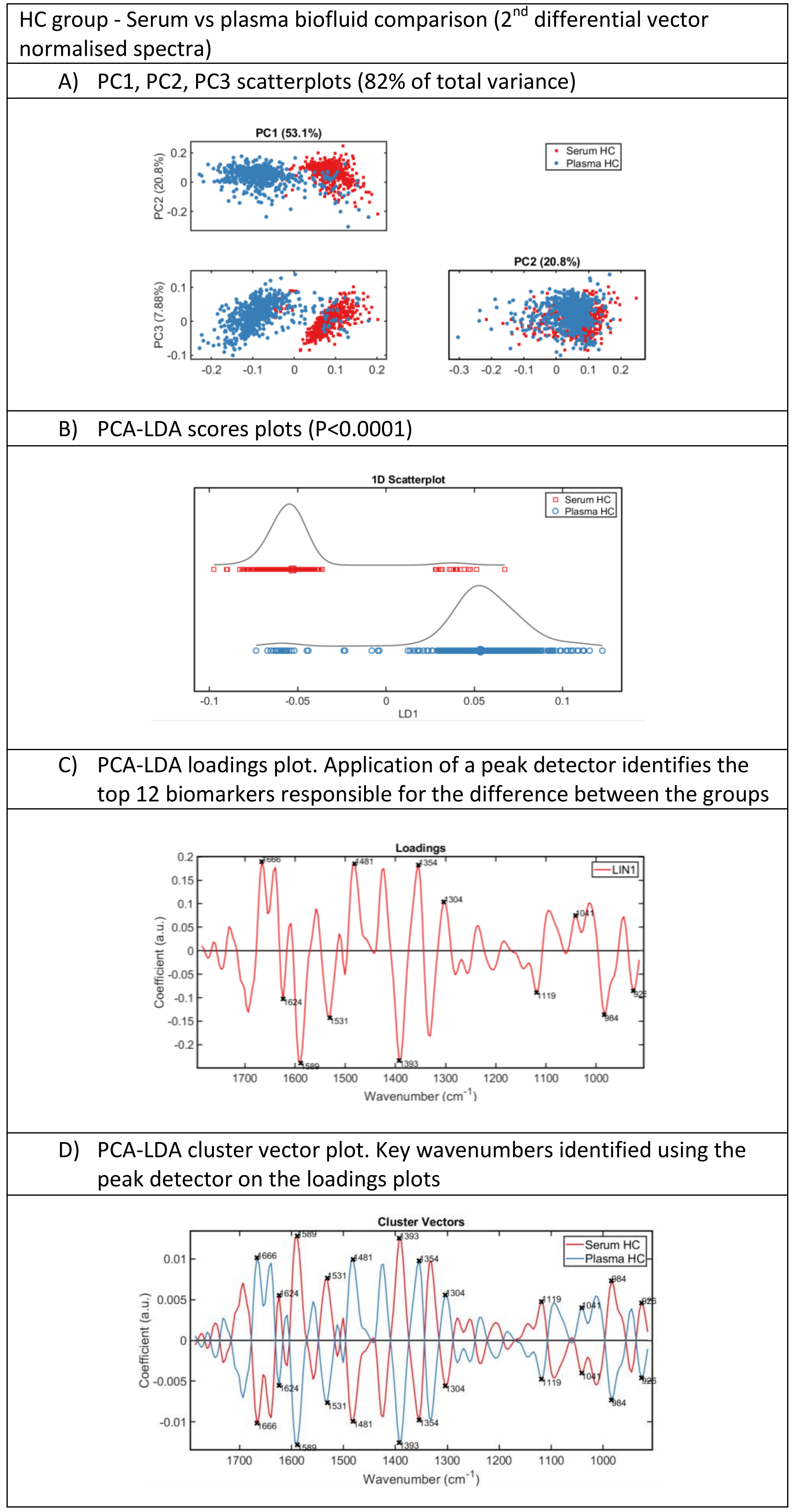


Figure 2.3.6. Unsupervised (PCA) and supervised (PCA-LDA) analysis of spectral variance between serum and plasma samples from healthy controls.

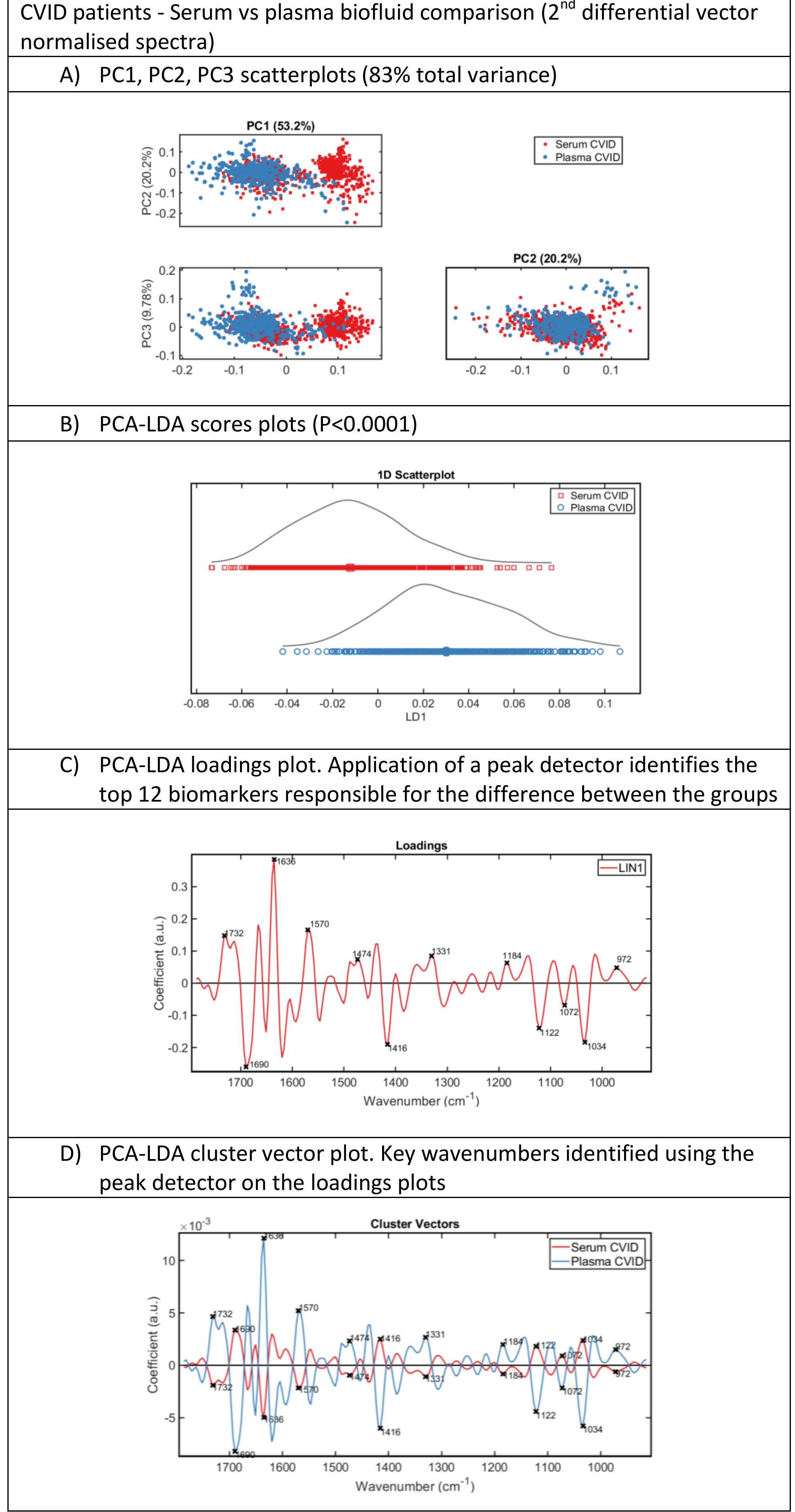


Figure 2.3.7. Unsupervised (PCA) and supervised (PCA-LDA) analysis of spectral variance between serum and plasma samples from CVID patients

In order to assess whether the spectral differences observed between the serum and plasma samples in the HC cohort were replicated in the disease group (CVID patients), the same analysis process was undertaken with the spectral data from the CVID patient samples. Figure 2.3.7A demonstrates the clustering of the generated PCA scores for each class (serum = red; plasma = blue) analysed using PC1, PC2 and PC3 (which covers 83% total variance).

As observed within the HC group scatterplots, two discrete clusters could be observed when using PC1, confirming the existence of significant spectral differences between the two biofluids. Although the unsupervised analysis was capable of separating the two clusters using PC1, the scatterplots of the CVID group spectra showed a greater degree of overlapping scores from each class, and the clusters appeared less clearly defined compared to the HC group.

Following the inclusion of class data into the PCA-LDA supervised analysis, the CVID patient spectra associated with each biofluid could be clearly separated, illustrated using PCA-LDA scores plots in Figure 2.3.7B. Once again, the separation observed between the biofluid classes of the CVID patients was less apparent on PCA-LDA scores plots when compared to the HC samples. To assess the significance of the class separation, a one-way ANOVA was used to calculate p-values for PCA-LDA scores, which demonstrated a statistical significance of $P < 0.0001$. Loadings plots (Figure 2.3.7C) and cluster vector plots (Figure 2.3.7D) were generated as previously described, with the twelve most distinguishing wavenumbers selected by application of the peak detector, as described in Table 2.3.2.

For the CVID patient group, the twelve most prominent wavenumber peaks (and associated molecular assignments) that distinguish between the serum and plasma biofluids were 972 cm^{-1} (protein phosphorylation vibration), 1034 cm^{-1} (glycogen), 1072 cm^{-1} (symmetric phosphate stretching), 1122 cm^{-1} (C-O stretching vibrations), 1184 cm^{-1} (carbohydrate), 1331 cm^{-1} (COO- symmetric stretching vibrations of fatty acids and amino acids), 1416 cm^{-1} and 1473 cm^{-1} (proteins), 1570 cm^{-1} (amide II), 1636 cm^{-1} and 1690 cm^{-1} (amide I), and 1732 cm^{-1} (C=O stretching band mode of the fatty acid ester).

The key wavenumbers extracted from the biofluid comparison for both HC and CVID patients are listed in adjacent columns in Table 2.3.2. The peaks occurring at closely associated wavenumbers or molecular assignments have been grouped together by

colour to illustrate any shared discriminating features observed within each group. In total there were seven overlapping molecular assignments associated with key spectral differences between serum and plasma samples. The discriminating features observed in both HC and CVID patients were; protein phosphorylation, glycogen, C-O stretch, COO- symmetric stretching, proteins, amide II and amide I.

Table 2.3.2. Key spectral wavenumbers associated with greatest spectral variance between serum and plasma samples identified using PCA-LDA cluster vector analysis. Closely associated wavenumbers observed within the HC and CVID group have been grouped together by colour across the two columns. These wavenumbers were identified following independent analysis of the serum and plasma samples from HC and CVID patients, respectively. Tentative molecular assignments generated from the IRootLab toolbox.

waves /cm^{-1}	HC	waves /cm^{-1}	CVID
926	Protein phosphorylation	972	Protein phosphorylation
984	Protein phosphorylation		
1041	Glycogen	1034	Glycogen
		1072	Symmetric phosphate
1119	C-O stretch (nu CO)	1122	C-O stretch (nu CO)
		1184	Carbohydrate
1304	Amide III		
1354	COO- symmetric stretching vibrations of fatty acids and amino acid	1331	COO- symmetric stretching vibrations of fatty acids and amino acid
1393	COO- symmetric stretching vibrations of fatty acids and amino acid		
		1416	Proteins
1481	Proteins	1473	Proteins
1531	Amide II	1570	Amide II
1589	Amide I		
1624	Amide I	1636	Amide I
1666	Amide I	1690	Amide I
		1732	Lipid

2.3.4 Investigating choice of biofluid - Discussion

The diagnosis of pathological disease can be greatly improved by undertaking appropriate, highly sensitive and specific laboratory testing of biological samples. The choice of sample type (cells, tissues, or biofluids) should be carefully considered to ensure that the appropriate compartment is being investigated, and that any disease-specific changes can be detected. The most widely used sample type for clinical investigation of human disease is blood, due to the fact that collection is relatively non-invasive, it is easily accessible, and that its molecular composition is reflective of structural and functional changes occurring within the human body.

An enormous number of discrete, measurable analytes can be detected in the blood. Whilst analytes shown to have disease-specific changes in concentration or function can be used as diagnostic biomarkers, there continues to be a lack of reliable biomarkers for the diagnosis and management of many pathological disorders and thus an unmet need for the development of innovative approaches to diagnostic medicine [22].

In this study we have chosen to analyse blood serum and blood plasma by ATR FTIR spectroscopy, assessing the complete composition of the sample as a whole rather than individual analytes. In this section of the method evaluation we have considered the molecular variances that are present between these two biofluids to avoid misclassification of samples in the main study due to sample-specific spectral variation rather than disease-specific variation.

In contrast to the subtle differences observed between the pre-processed mean spectra of serum samples from CVID patients and healthy controls in the previous results section, the differences between the mean spectra for serum and plasma samples (within the same class, healthy controls) were immediately apparent. This was anticipated given that the two sample types have fundamental differences in biochemical composition (removal of clotting proteins from serum), and based on the metabolic differences observed in previous studies using mass spectrometry [11,12].

The initial visualisation of the mean pre-processed spectra and SDs from serum and plasma samples mirrored the previous findings and confirmed the suitability of the pre-processing methods selected in the previous section (rubberband or second differential baseline correction followed by vector normalisation).

Subsequent multivariate analysis using PCA and PCA-LDA demonstrated significant variances between the two sample types. In both the HC and CVID patient group the blood sample types are not interchangeable for this methodology platform, unlike the majority of assays currently performed in the immunology laboratory (i.e. ELISA, immunoblot, turbidimetry, nephelometry, radioimmunoassay, electrophoresis, and indirect immunofluorescence). Notably, although the PCA-LDA score plots for both classes (HC and CVID) demonstrated a significant difference between the classes ($p<0.0001$), in the HC group the separation between the classes on the supervised (PCA-LDA) scores plots was greater, with less overlap compared to the CVID patient group.

We hypothesise that this may be due to the presence of additional disease-specific variables within the CVID group, which may relate to co-morbidities such as infection, inflammation, autoimmunity, or a result of treatment such as replacement immunoglobulin or antibiotics. If these disease-specific variances are present in both the serum and plasma compartments, the variances separating the two sample types will be less apparent when applying unsupervised methods to separate the classes, as illustrated in PC2 separation within Figure 2.3.6A. This reiterates the requirement for supervised methods of multivariate analysis (such as LDA), which consider the class data within the analysis to aid class separation and enable the detection of subtle class-specific variances within the spectral data.

To investigate the biochemical differences between the serum and plasma samples further and assess which spectral features account for the major variance between the two sample types, a peak detector was applied to the PCA-LDA loadings plots. In the HC group, the twelve most prominent wavenumbers were 926 cm^{-1} and 984 cm^{-1} (protein phosphorylation vibrations), 1041 cm^{-1} (glycogen), 1119 cm^{-1} (C-O stretching vibrations), 1304 cm^{-1} (amide III), 1354 cm^{-1} and 1393 cm^{-1} (COO- symmetric stretching vibrations of fatty acids and amino acids), 1481 cm^{-1} (proteins), 1531 cm^{-1} (amide II), 1589 cm^{-1}, 1624 cm^{-1} and 1666 cm^{-1} (amide I); and in the CVID group, 972 cm^{-1} (protein phosphorylation vibration), 1034 cm^{-1} (glycogen), 1072 cm^{-1} (symmetric phosphate stretching), 1122 cm^{-1} (C-O stretching vibrations), 1184 cm^{-1} (carbohydrate), 1331 cm^{-1} (COO- symmetric stretching vibrations of fatty acids and amino acids), 1416 cm^{-1} and 1473 cm^{-1} (proteins), 1570 cm^{-1} (amide II), 1636 cm^{-1} and 1690 cm^{-1} (amide I), and 1732 cm^{-1} (C=O stretching band mode of the fatty acid ester).

As the most distinguishing peaks separating the serum and plasma biofluids occurred at closely associated wavenumbers for both the HC and CVID patient groups, this suggests that the differences are due to the biochemical differences between the sample types, and not due to within-group (disease-related) variances.

Further work to assess the differences in these spectral biomarkers between sample types, i.e. whether they increase or decrease in concentration would be informative. Both serum and plasma have approximately 60-80 mg/mL of total protein, largely made up of albumin and globulins, with plasma having the higher total protein content of the two biofluids, due to the removal of clotting factors and fibrinogen in serum [9,23]. As several of the tentative molecular assignments from the key wavenumbers identified are associated with proteins; 1304 cm^{-1} (amide III), 481 cm^{-1} (proteins), 1531 cm^{-1} (amide II), 1589 cm^{-1}, 1624 cm^{-1} and 1666 cm^{-1} (amide I), 1416 cm^{-1} and 1473 cm^{-1} (proteins), 1570 cm^{-1} (amide II), 1636 cm^{-1} and 1690 cm^{-1} (amide I), we hypothesise that an increased contribution from these molecular vibrations will be observed in the plasma samples.

Prevention of the clotting cascade in plasma samples occurs through addition of Ethylenediaminetetraacetic acid (EDTA) and chelation of calcium, magnesium, zinc and potassium ions [10]; the concentration of these metallic ions will therefore be increased in serum samples compared to plasma. However, as FTIR spectroscopy can only be used to characterise molecules with a dipole moment, it would not be possible to detect these monoatomic ions and thus any concentration differences between the biofluids using ATR-FTIR.

This piece of work validates the requirement to include both biofluids as separate experiments within the main study, as at this point we are unsure which will identity the most significant spectral changes which could be used as CVID-specific biomarkers in the future. This investigation also confirms that the two biofluids are fundamentally different and have discrete FTIR molecular fingerprints. These two sample-types therefore cannot be used interchangeably within this testing platform.

In the main study, we propose that disease classification models (HC versus CVID spectra) will be built for each sample type (serum and plasma). In order to assess which biofluid has superior performance for the potential use as a diagnostic test, we will compare the calculated classification rates (ability to correctly classify spectra as either a HC or CVID sample), and the sensitivity and specificity values associated with each biofluid

classification model. We also aim to extract major wavenumber variances between HC and CVID patients and evaluate the value of using key spectral peaks as individual disease biomarkers. The spectra of both biofluids will be explored; if successful, a wavenumber library of potential CVID-specific wavenumbers will be built for both serum and plasma samples.

2.3.5 Examining environmental variance - Results

In the final aspect of the method evaluation section, the contributing variances from environmental factors was considered by investigating whether significant variance could be observed between serum samples analysed on different days (inter-assay variation). Figure 2.3.8A illustrates the mean unprocessed spectra of samples analysed on different dates for the HC group, and Figure 2.3.8B for the CVID patient group.

In total, 10 individual data acquisition dates were compared within this evaluation (08/06/2017, 14/06/2017, 05/07/2017, 26/07/2017, 03/08/2017, 18/08/2017, 30/08/2017, 06/09/2017, 28/09/2017, and 03/10/2017). As expected and observed throughout this method evaluation study, any obvious initial visual differences of the mean spectra between the classes were no longer observed following pre-processing (cutting, second differential baseline correction followed by vector normalisation), as shown in Figure 2.3.8C and 2.3.8D, for HC and CVID populations, respectively.

The exploratory analysis was undertaken using PCA and MANOVA statistical testing. Examination of the scores within PC1, PC2 and PC3 demonstrated there to be no significant difference between the spectra acquired on the different days for the CVID patient population ($p = 0.351$), illustrated in Figure 2.3.8F. Of note, there was a significant difference observed within spectra collected over different days within the HC group ($p < 0.0005$), illustrated in Figure 2.3.8E. Due to the complexity of the PCA results output and number of classes being compared, it is not possible to clearly discern any discrete clustering of scores within the 3D scatterplots with visual examination alone.

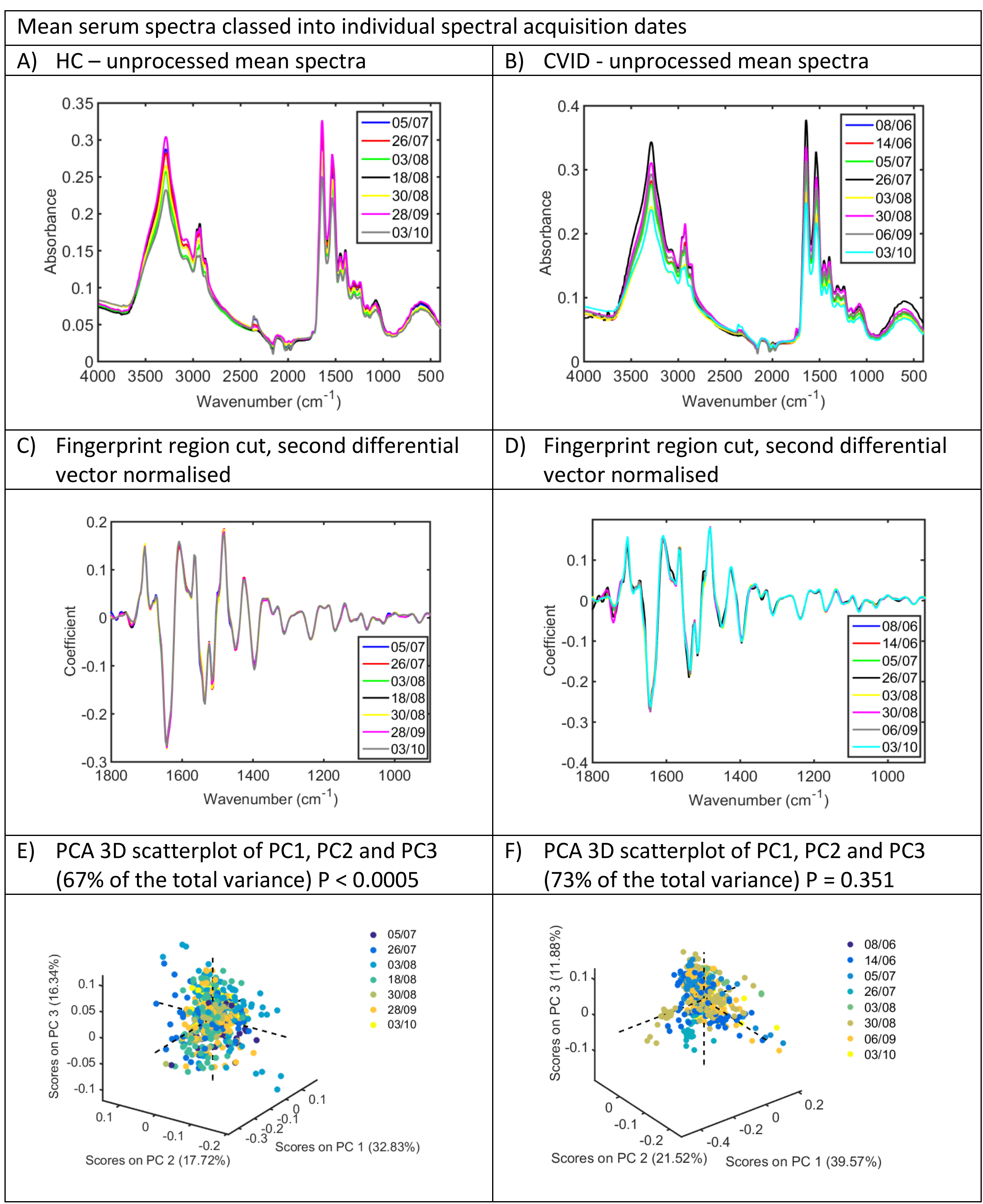


Figure 2.3.8. Examining the spectral differences between serum samples analysed across different dates (inter-assay variation) in HC and CVID patient populations using unsupervised PCA analysis.

2.3.6 Examining environmental variance - Discussion

In the environmental stability data, there is a significance ($p < 0.0005$) observed between the spectra collected on different days in the HC samples, which was not observed within the CVID patient study population. As highlighted in the introduction, environmental variances (including temperature, humidity, instrumentation or user input) can have an impact on the quality of spectra acquired and potentially lead to the incorrect identification of spectral variance between classes that is not disease or class-related. Several steps have been undertaken to account for these factors and eliminate bias within this research project, these include: using standardised processes for sample collection, sample storage, slide preparation and spectral acquisition, having a single researcher performing all of the aforementioned processes and by applying appropriate pre-processing steps to the data prior to multivariate analysis.

This aspect of the method evaluation has highlighted the fact that significant spectral variances can be observed between samples (or groups of samples) within the sample study population when the analysis is conducted over multiple different days. At this point it would not be possible to attribute this solely to environmental differences, as the findings were not mirrored in the CVID patient group, with no significant variance observed between samples processed on different days.

An alternative explanation would be that the observed variances detected between HC samples acquired on different days are attributed to inter-donor variability, which has become more apparent when the samples are divided into smaller groups of individual participants for comparison. This would not be surprising given the wide variety of demographic and participant characteristics that could be included in a 'Healthy Control' cohort. Further evaluation to elucidate the significance of wavenumber variances between individual HCs would be informative; however it was outside the scope and timescale of this piece of work.

Exclusion criteria were applied to the HC cohort to control the age range (participants were all 18-50 years old) and to prevent the inclusion of pregnant women or immune deficient individuals. Further steps were taken to collect information on participant infection history, medication and the presence of any immune-related disorders through the use of a brief questionnaire. Nevertheless, detailed demographics were not collected

for the HC group thus it would not be possible to fully explore within-class spectral variance within this study.

To summarise, the findings from the evaluation into the impact of environmental influence are not conclusive however this piece of work adds value towards the determination of analytical methods chosen in the main study. Whilst there were no significant differences observed between CVID patient samples tested on different days, the HC group demonstrated that environmental variability could potentially play a role as a source variance between samples.

This can be accounted for in the building of the classification modes in the main study. The construction of a classification model requires the data to be split into a 'training' dataset and a 'test' dataset. The splitting of the data into two discrete subsets provides the opportunity to use the first set in the calibration stage of the classification model construction. The model can subsequently be 'tested' for prediction capacity using a second subset of study data. It is important that most sources of variance are captured within the training dataset to ensure that the final classification model is not skewed by inter-assay variation and environmental factors. In light of the findings here, it is vital that the training dataset includes samples tested across multiple different days.

The two most commonly used computational-based splitting methods applied to spectroscopy data are Random selection (RS) and the Kennard-Stone (KS) algorithm[24]. Whist the RS method is favoured for its simplicity; KS is often the preferred choice for biochemical analysis. By including most of the variability in the training dataset, it provides a more uniform and representative training model for classification purposes[25]. In the construction of classification models within the main study, the KS algorithm will be used to ensure that all potential sources of environmental and inter-donor variability are contemplated in the training process.

Further to this, this splitting process will be performed on an individual patient basis rather than an individual spectrum basis (as 20 spectra were acquired per patient), to ensure that the training and test groups do not contain spectra from the same patient.

2.4 Summary of all Method Evaluation Results

- The pre-processing techniques chosen to be applied to spectral data prior to constructing the classification models within the main body of work will be second differential (SG) baseline correction followed by vector normalisation.
- Serum and plasma biofluids both show suitability to use as sample types within the main study.
- Significant spectral variances identified between the two biofluids will require that each sample type is analysed independently to assess which provides the superior classification model in the main study.
- The potential impact from environmental factors will be considered during the analysis of the spectral data. The construction of a classification model in the main study will be done using a training dataset, generated using recommended computational methods. To prevent skewing of the final model, the training dataset will include sample spectra collected from a variety of different dates, ensuring that all sources of environmental variability are considered.

References

1. Rinnan, Å., Berg, F. van den & Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* **28**, 1201–1222 (2009).

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

3. Baker, M. J. *et al.* Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protocols* **9**, 1771–1791 (2014).

4. Martin, F. L. *et al.* Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nat. Protocols* **5**, 1748–1760 (2010).

5. Luna, A. S. & Gois, J. S. Application of Chemometric Methods Coupled With Vibrational Spectroscopy for the Discrimination of Plant Cultivars and to Predict Physicochemical Properties Using R. in *Comprehensive Analytical Chemistry* vol. 80 165–194 (Elsevier, 2018).

6. Brereton, R. G. Pattern recognition in chemometrics. *Chemometrics and Intelligent Laboratory Systems* **149**, 90–96 (2015).

7. HéBerger, K. Chemoinformatics—multivariate mathematical–statistical methods for data evaluation. in *Medical Applications of Mass Spectrometry* 141–169 (Elsevier, 2008).

8. Section VI Cardiovascular Physiology. in *Ganong's review of medical physiology* (eds. Barrett, K. E. & Ganong, W. F.) (McGraw-Hill Medical, 2010).

9. Gonzalez-Quintela, A. *et al.* Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities: Serum immunoglobulin levels in adults. *Clinical & Experimental Immunology* **151**, 42–50 (2007).

10. Bowen, R. A. R. & Remaley, A. T. Interferences from blood collection tube components on clinical chemistry assays. *Biochemia Medica* 31–44 (2014) doi:10.11613/BM.2014.006.

11. Liu, X. *et al.* Serum or plasma, what is the difference? Investigations to facilitate the sample material selection decision making process for metabolomics studies and beyond. *Analytica Chimica Acta* **1037**, 293–300 (2018).

12. Yu, Z. *et al.* Differences between Human Plasma and Serum Metabolite Profiles. *PLoS ONE* **6**, e21230 (2011).

13. Conley, M. E., Notarangelo, L. D. & Etzioni, A. Diagnostic Criteria for Primary Immunodeficiencies. *Clinical Immunology* **93**, 190–197 (1999).

14. Savitzky, A. & Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **36**, 1627–1639 (1964).

15. Kelly, J. G. *et al.* Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. *J. Proteome Res.* **10**, 1437–1448 (2011).

16. Bro, R. & Smilde, A. K. Principal component analysis. *Anal. Methods* **6**, 2812–2831 (2014).

17. Morais, C. L. M. & Lima, K. M. G. Principal Component Analysis with Linear and Quadratic Discriminant Analysis for Identification of Cancer Samples Based on Mass Spectrometry. *Journal of the Brazilian Chemical Society* **29**, 472–481 (2018).

18. Martin, F. L. *et al.* Identifying Variables Responsible for Clustering in Discriminant Analysis of Data from Infrared Microspectroscopy of a Biological Sample. *Journal of Computational Biology* **14**, 1176–1184 (2007).

19. Hands, J. R. *et al.* Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples. *Journal of Biophotonics* **7**, 189–199 (2014).

20. Lovergne, L. *et al.* Investigating optimum sample preparation for infrared spectroscopic serum diagnostics. *Anal. Methods* **7**, 7140–7149 (2015).

21. Fearn, T. Discriminant Analysis. in *Handbook of Vibrational Spectroscopy* (eds. Chalmers, J. M. & Griffiths, P. R.) s4302 (John Wiley & Sons, Ltd, 2006). doi:10.1002/0470027320.s4302.

22. Naumann, D. Vibrational Spectroscopy in Microbiology and Medical Diagnostics. in *Biomedical Vibrational Spectroscopy* (eds. Lasch, P. & Kneipp, J.) 1–8 (John Wiley & Sons, Inc., 2008).

23. *Ganong's review of medical physiology*. (McGraw-Hill Medical, 2010).

24. Kennard, R. W. & Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **11**, 137–148 (1969).

25. Morais, C. L. M., Santos, M. C. D., Lima, K. M. G. & Martin, F. L. Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard-Stone algorithm approach. *Bioinformatics* **35**, 5257–5263 (2019).

Chapter 3: Research Study

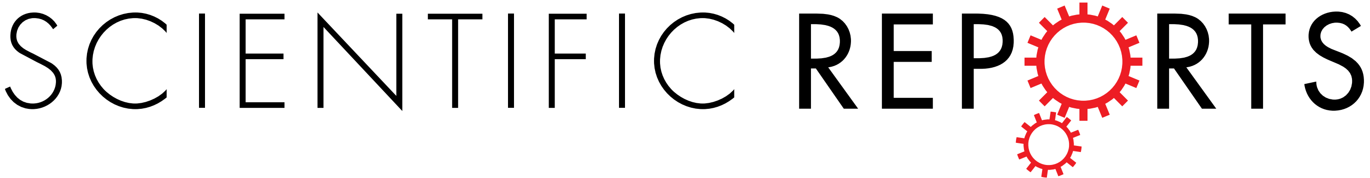
Published Research Article: New approach to investigate Common Variable Immunodeficiency patients using spectrochemical analysis of blood

Authors: Emma L. Callery, Camilo L. M. Morais, Maria Paraskevaidi, Vladimir Brusic, Pavaladurai Vijayadurai, Ariharan Anantharachagan, Francis L. Martin, & Anthony W. Rowbottom

Key words: Diagnostic markers, Primary immunodeficiency disorders, Infrared FTIR, Computational Analysis, Diagnosis, Common Variable Immune deficiency, Spectroscopy, Serum levels

Paper published in Nature Scientific Reports (Sci.Rep, 2019)

DOI: https://doi.org/10.1038/s41598-019-43196-5



OPEN

Received: 14 January 2019
Accepted: 4 April 2019
Published online: 10 May 2019

New approach to investigate Common Variable Immunodeficiency patients using spectrochemical analysis of blood

Emma L. Callery[1], Camilo L. M. Morais[2], Maria Paraskevaidi[2,5], Vladimir Brusic[3], Pavaladurai Vijayadurai[1], Ariharan Anantharachagan[1], Francis L. Martin[2] & Anthony W. Rowbottom[1,4,6]

Common variable immune deficiency (CVID) is a primary immunodeficiency disease, characterized by hypogammaglobulinemia, recurrent infections and various complications. The clinical heterogeneity of CVID has hindered identification of an underlying immune defect; diagnosis relies on clinical judgement, alongside evidence-based criteria. The lack of pathognomonic clinical or laboratory features leads to average diagnostic delays of 5 years or more from the onset. Vibrational spectroscopic techniques such as Fourier-transform infrared (FTIR) spectroscopy have recently gained increasing clinical importance, being rapid-, non-invasive and inexpensive methods to obtain information on the content of biological samples. This has led us to apply FTIR spectroscopy to the investigation of blood samples from a cohort of CVID patients; revealing spectral features capable of stratifying CVID patients from healthy controls with sensitivities and specificities of 97% and 93%, respectively for serum, and 94% and 95%, respectively for plasma. Furthermore we identified several discriminating spectral biomarkers; wavenumbers in regions indicative of nucleic acids (984 cm^{-1}, 1053 cm^{-1}, 1084 cm^{-1}, 1115 cm^{-1}, 1528 cm^{-1}, 1639 cm^{-1}), and a collagen-associated biomarker (1528 cm^{-1}), which may represent future candidate biomarkers and provide new knowledge on the aetiology of CVID. This proof-of-concept study provides a basis for developing a novel diagnostic tool for CVID.

Common variable immunodeficiency (CVID) is the most frequent life-threatening and symptomatic primary immune deficiency (PID)[1]. The estimated prevalence is between 1:10,000 and 1:100,000 of the population, with two peak ages of onset, one before the age of ten and another between 30–40 years of age[2,3]. The majority (>80%) of CVID cases are sporadic and the main diagnostic method is exclusion; often with a delay of approximately 5 years[4]. Failure to produce sufficient immunoglobulins results in recurrent infections in 90% of CVID patients; an increased risk of autoimmune disorders (22% of patients) and malignancy (16% of patients)[5–8]. CVID is a heterogeneous group of polygenic disorders for which the exact pathogenesis remains poorly understood[9–11]. Genetic mutations are implicated in CVID in 10–20% of patients; with defects found in more than 30 genes[4,8,12].

There are no clinical or laboratory features that are pathognomonic for CVID. Diagnostic criteria have therefore been developed which require sequential application of both clinical and laboratory findings in order to increase the specificity of the diagnosis. Current diagnostic criteria[5,13,14] define hypogammaglobulinemia as a major requirement for the diagnosis of CVID, but only when used in conjunction with further clinical or laboratory findings. This is due to the low diagnostic specificity of hypogammaglobulinemia for CVID; as reduced serological levels of immunoglobulin are also associated with a vast array of other primary and secondary immune disorders. The incidence and prevalence rates of hypogammaglobulinemia are not clearly defined, however

[1]Department of Immunology, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, PR2 9HT, UK. [2]School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston, PR1 2HE, UK. [3]University of Nottingham Ningbo China, Ningbo, 315100, China. [4]School of Medicine, University of Central Lancashire, Preston, PR1 2HE, UK. [5]Department of Surgery and Cancer, Imperial College London, London, W12 0HS, UK. [6]NIHR Lancashire Clinical Research Facility, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, PR2 9HT, UK. Correspondence and requests for materials should be addressed to E.L.C. (email: emma.callery@lthtr.nhs.uk) or A.W.R. (email: anthony.rowbottom@lthtr.nhs.uk)

secondary causes are more common[15]. Although the requirement to exclude secondary causes of hypogammaglobulinemia limit the use of this finding as a diagnostic marker for CVID, it remains a fundamental test, given that the hallmark of the disease is a reduced serum level of IgG, IgA and/or IgM. The definition of local or regional population reference ranges will impact the diagnostic utility of immunoglobulin results, as will the laboratory choice of analytical method (nephelometric and tubidmetric methods are most widely used). An absolute lower limit value of IgG at 4.5 g/L for adults has been proposed, as despite the wide range of IgG levels observed in CVID patients, Chapel and Cunningham-Rundles[2] described the majority of their 334 patients (94.2%) as having initial IgG levels <4.5 g/l at diagnosis.

Further laboratory testing, such as measuring specific antibodies to vaccine responses, enumeration of lymphocyte subsets (B, T and NK cells) and class-switched memory B cells by flow cytometry, can provide additional evidence to suggest defective antibody production. Findings are variable across the disease group and also within individual patients. Repeat testing has therefore been suggested to confirm any sub-normal findings; these limitations have been discussed in recent diagnostic criteria[5,13].

Efforts to categorise clinical subgroups within cohorts of established CVID patients using flow cytometry have further demonstrated the complex aetiology of this disorder, emphasising the variety of B cell differentiation defects that can contribute to the disease. Phenotypic analysis of B cells using population sizes of class-switched memory- and transitional B cells, in correlation with clinical aspects has generated three classification protocols for patients with CVID[16–18]. Whilst these studies demonstrate that B cell homeostasis is a pathogenic and clinically meaningful parameter for classification, reduced numbers of class-switched or memory B cells are not specific to CVID hence the diagnostic utility of these tests in isolation is limited[2].

Further challenges with current diagnostics relate to a lack of understanding as to which physiological compartment should be investigated for CVID-associated abnormalities, *i.e*., tissues, biological fluids, or cells. A potential novel diagnostic methodology for CVID is vibrational biospectroscopy. High resolution spectroscopy methods such as Fourier-transform infrared (FTIR) spectroscopy can provide unique spectral patterns that reflect the chemical and molecular composition of biological samples. We hypothesised that pathological changes in CVID patients produce characteristic FTIR spectra that distinguish them from healthy controls. The interaction of infrared (IR) light with biological matter produces an absorption plot, or 'spectral fingerprint', for each biological sample. The principles and biological applications of FTIR spectroscopy have been reviewed in detail in Baker *et al.*[19]. Vibrational spectroscopy is gaining recognition in the field of diagnostic medicine for a range of complex pathologies, mostly for malignancies[19–21]. A key requirement for diagnostic investigations is the highly accurate discrimination of pathological features from healthy neighbouring tissue or cells. These measurements are often performed on samples characterised by high background signals (or 'noise') relating to biological activity (*e.g*., increased cell turnover or inflammatory states). Due to the high complexity of vibrational spectroscopy data, computational-based methods (chemometrics) are needed to explore and extract relevant information from the experimentally acquired spectra. For this, multivariate classification techniques can be employed for feature extraction and classification, allowing biochemically-relevant information to be extracted and the automatic grouping of samples into pre-defined categories. This can be achieved using a combination of chemometric algorithms, such as forward feature selection (FFS), principal component analysis linear discriminant analysis (PCA-LDA), and principal component analysis support vector machines (PCA-SVM). All of these algorithms are based on a principal component analysis (PCA) decomposition, which significantly reduces the original data complexity to a fewer number of relevant factors, named principal components (PCs). PCA-LDA performs a linear discriminant analysis (LDA) of the PCA scores to assign the samples to their predicted groups; whereas PCA-SVM does the same procedure but in a non-linear classification fashion through a support vector machines (SVM) algorithm. FFS allows identification of main biomarkers responsible for class differentiation by calculating p-values for the spectral wavenumbers with larger PCA loadings. Vibrational spectroscopy has been successfully applied across a wide area of clinical medicine; providing a new approach to detect molecular and structural changes caused by complex disorders such as Alzheimer's disease[22,23], multiple sclerosis[24], mental disorders[25,26], HIV/AIDS[27], diabetes[28] and carcinogenesis[25,29–32]. High diagnostic accuracy was demonstrated for classification of numerous cancer types and other biological applications[33–38]. We hypothesised that vibrational spectroscopy will demonstrate similar analytical advantages within our cohort, allowing for the sensitive detection of characteristic spectral fingerprints that represent underlying pathological processes in CVID patients.

To our knowledge, this study reports for the first time, the application of FTIR methods, specifically, attenuated total reflection-FTIR (ATR-FTIR) spectroscopy, for detection of CVID. In this first phase we have explored the application of this technique to stratify CVID patients from healthy controls (HC) using serum and plasma. We performed stratified spectroscopic classification at multiple levels, to differentiate subgroups of CVID patients with- and without- further clinical complications. Finally, we have identified a number of meaningful and discriminating spectral biomarkers, tentatively assigned to specific molecular entities. These promising initial findings encourage further development of FTIR spectroscopy as a diagnostic technique for immune deficiency.

Results

The major aim of this study was the discrimination of CVID patients from HC in two biofluids; serum and plasma using FTIR-spectroscopy and multivariate analysis techniques. The ATR-FTIR spectra from 51 subjects (1020 spectra) were obtained and analysed using multiple chemometric methods. An exploratory (unsupervised) analysis using PCA model was undertaken, followed by classification using supervised methods (PCA-LDA, FFS, SVM) to enable successful segregation of subjects into their respective groups, CVID patients and healthy controls.

Discrimination of CVID patients from Healthy controls

Analysis of prominent IR spectral regions reveals significant variance between CVID and HC groups. Rubber-band baseline correction and vector normalisation produces spectra for crude visualisation of differences between the two groups, and corrects for experimental variation; this recognised technique improves accuracy and interpretability of the data whilst maintaining spectral integrity. The generated figures (Supplemental Fig. 1a,b (fingerprint region); Supplemental Fig. 1e,f (high region)) demonstrate visual spectral similarities for each class (CVID, HC), prior to the application of multivariate analysis tools. As expected for biological samples, the Amide I band was most prominent in the IR spectrum of the fingerprint region, dominated by C=O stretching, and N-H bending vibrations of proteins[39,40]. To enhance spectral variability between groups, second-order differentiation was applied (Supplemental Fig. 1c,d (Fingerprint region); Supplemental Fig. 1g,h (High region)) prior to implementation of multivariate approaches.

Serum. Key differences observed in the fingerprint region of serum (Supplemental Fig. 1a) at the pre-processing stage were lower absorbance intensities in the CVID group compared to HC at the nucleic acid-associated asymmetric stretching (ν_{as}) of PO_2^- (DNA/RNA) [1242 cm^{-1} ($p = 0.0003$). At the high region (Supplemental Fig. 1e) the CVID group revealed increased absorbance peaks within the lipid and protein associated (CH_3 and CH_2) stretching vibrations[21] (serum 3000–2800 cm^{-1} $p =< 0.0001$).

Plasma. The fingerprint region of the plasma spectra (Supplemental Fig. 1b) revealed greatest variance between the two groups at the Amide I [1643 cm^{-1} ($p < 10^{-6}$)] and Amide II [1535 cm^{-1} ($p < 10^{-6}$)] and Amide III peaks [1315 cm^{-1} plasma $p =< 0.0001$)], with lower absorbance found in the CVID group compared to HC. At the high region (Supplemental Fig. 1f), peak increases at 2928–2932 cm^{-1} ($p =< 0.0001$) were found in the CVID group compared to HC.

Segregation of CVID patients from HC can be demonstrated following multivariate analysis of the entire spectral dataset for both Fingerprint (1800–900 cm^{-1}) and High (3700–2800 cm^{-1}) regions. Additional examination of the spectra was performed using cross-validated PCA-LDA. The 1D PCA-LDA scores plots (Fig. 1A–D) were generated, and utilised to illustrate the significant differences between the CVID group (red) and the HCs (blue) ($p < 0.0005$); the "scores" here represent individual spectra, (fingerprint region serum $p < 10^{-6}$ and plasma $p < 10^{-6}$; high region serum $p < 10^{-6}$ and plasma $p \approx 10^{-4}$). To further explore whether the classes could be significantly separated on a study subject-level basis, mean values of the 20 (second-order differentiated) spectral replicates per sample were calculated prior to performing cross-validated PCA-LDA (illustrated in Supplemental Fig. 1a–d), in which each score represents an individual study subject. We again demonstrated significant differences between HC and CVID groups, this time on a patient level, for both serum and plasma at the fingerprint region ($p < 10^{-6}$ for both), and at the high region ($p < 10^{-6}$ for both).

Discrimination of CVID patients from HC using classification models

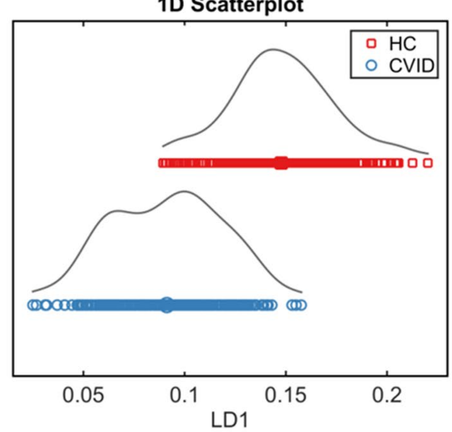
SVM classification prior to incorporation of additional clinical information. Following successful segregation of classes using PCA-LDA scores plots, the ability and performance of FTIR as a tool to discriminate CVID patients from HC was assessed through creation of classification models. Classification of CVID and HC was performed on the Fingerprint and High regions of the IR spectrum using support vector machine (SVM) learning algorithms as described in the methods. The SVM models were generated using 2/3 of the spectral data for the four distanced groups (Serum Fingerprint, Serum High, Plasma Fingerprint and Plasma High) prior to being tested with the remaining 1/3. The confusion matrices generated following the input of the test data into each classification model can be illustrated graphically in the form of confusion balls (Fig. 2a–d). Supplemental Table 2 specifies the c and γ values outputted by the four grid searches. Within the serum, correct classification was achieved for 99% of HC and 92% of CVID patients using the fingerprint region (Fig. 2a); and 71% of HC and 44% of CVID patients using the high region (Fig. 2c). Within the plasma, correct classification was achieved for 96% of HC and 92% of CVID patients using the fingerprint region (Fig. 2b); and 72% of HC and 51% of CVID patients using the high region (Fig. 2d). The highest sensitivities and specificities were obtained for the fingerprint region, achieving 97% and 93% respectively for serum; 94% and 95% respectively for plasma. In the high region, sensitivities and specificities were lower, at 66% and 91% respectively for serum; 55% and 69% for plasma (Fig. 2d).

SVM classification subsequent to incorporation of clinical subgrouping. CVID patients were further divided based on clinical manifestations (see methods) prior to application of SVM learning algorithms for classification (Fig. 3); the parameters identified during grid searches of the training data are included in Supplemental Table 3.

Serum. In the HC group, 10% of subjects were classified correctly; within the two CVID sub-groups, the classification model correctly assigned 56% of the CVID patients without further complications and 56% of patients with further complications into their respective groups (Fig. 3a). Classification using the high region achieved correct group assignment for 79% of HC subjects, 16% of CVID patients without further complications and 51% of CVID patients with further complication (Fig. 3d). The sensitivities and specificities of the fingerprint SVM model after incorporation of clinical data are documented in Fig. 3c; with greatest specificity, 91%, achieved when classifying HC subjects using the fingerprint region. For the classification of CVID subgroups using the high region, specificities of 92% for CVID-non complication patients and 82% for CVID-complication patients were achieved (Fig. 3f). In comparison to the classification model based on HC vs all CVID patients, sensitivities achieved

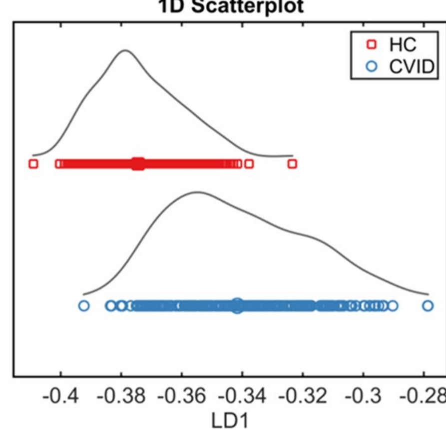
Serum Fingerprint Region

A) Serum fingerprint region cross-validated PCA-LDA scores plots (p < 0.0005)



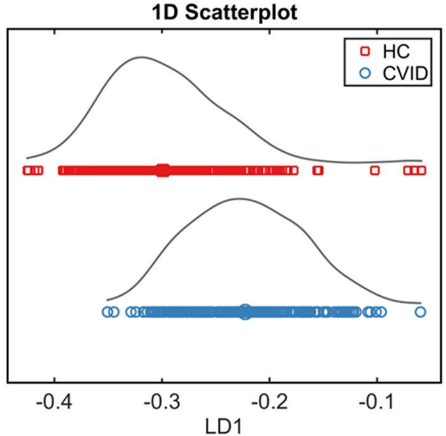
Plasma Fingerprint Region

B) Plasma fingerprint region cross-validated PCA-LDA scores plots (p < 0.0005)



Serum High Region

C) Serum high region cross-validated PCA-LDA scores plots (p < 0.0005)



Plasma High Region

D) Plasma high region cross-validated PCA-LDA scores plots (p < 0.0005)

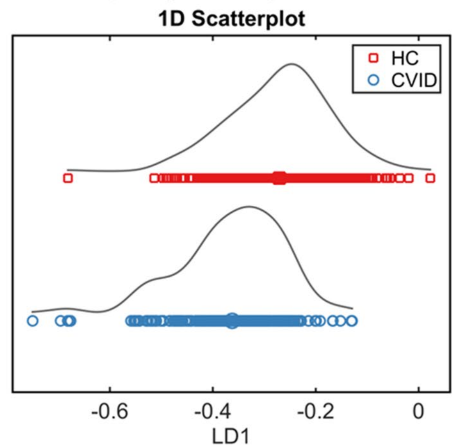


Figure 1. Supervised multivariate analysis techniques (PCA-LDA) successfully segregate classes (CVID vs HC). (**A**,**B**) Fingerprint region (900–1800 cm^{-1}); 1D scores plots (LD1) after cross-validated PCA-LDA of the training dataset (CVID $n = 13$; HC $n = 18$) for serum and plasma respectively. (**C**,**D**) High region (2800–3700 cm^{-1}); 1D scores plots (LD1) after cross-validated PCA-LDA of the training dataset (CVID $n = 13$; HC $n = 18$) for serum and plasma respectively.

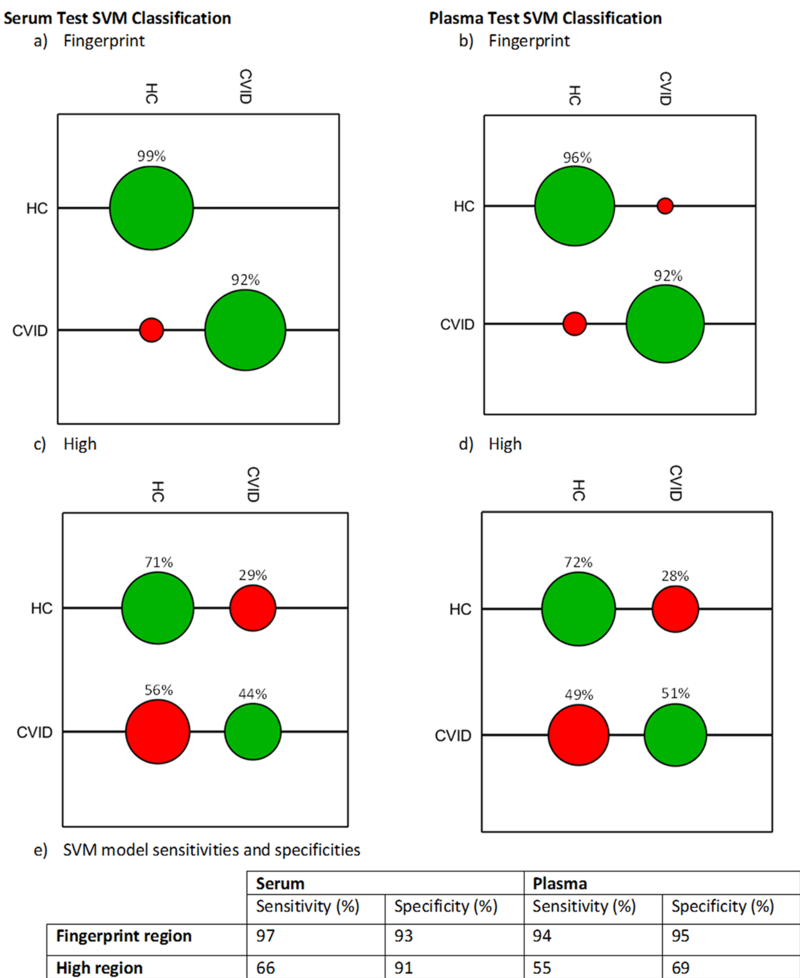
following clinical subgrouping were lower, at 41% for CVID-non complication and 71% for CVID-complication patients using the fingerprint region.

Plasma. SVM models generated for the plasma data demonstrated increased classification ability compared to the serum, correctly assigning 93% of HC, 77% of CVID-non complication patients and 76% of CVID-complication patients using the fingerprint region (Fig. 3b); classification rates in the high region were 75%, 41% and 52% respectively (Fig. 3e). Sensitivities and specificities for the three groups are documented in (Fig. 3c,f). Classification using the fingerprint region achieved highest sensitivities and specificities; HCs were classified with a sensitivity and specificity of 93% and 87% respectively, CVID-non complication patients with 73% and 93% respectively and CVID-complication patients with 73% and 95% respectively.

Biomarker analysis

Feature extraction was performed as described in the methods; the key biomarkers extracted from each technique are illustrated in Fig. 4 (fingerprint) and Supplemental Fig. 3 (high region) and documented in Supplemental Tables 4 and 5, along with the tentative molecular assignments previously described for individual wavenumbers. Relative increases or decreases in the absorbance intensity of CVID spectra are indicated where further subject-level analysis was performed.

Figure 4 and Supplemental Fig. 3(a,b) illustrates the 6 most variant peaks elucidated using the Student's T-Test method; Fig. 4 and Supplemental Fig. 3(c,d) shows the biomarkers selected from the cluster vectors using



	Serum		Plasma	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Fingerprint region	97	93	94	95
High region	66	91	55	69

Figure 2. SVM classification model for CVID vs HC using each biofluid at the fingerprint (900–1800 cm^{-1}) and high region (2800–3700 cm^{-1}) of the spectrum. (**a–d**) SVM confusion matrices for (**a**) serum fingerprint, (**b**) plasma fingerprint, (**c**) serum high and (**d**) plasma high regions. The tuning parameters (c, γ) extracted from a grid search of the training dataset were used to subsequently generate confusion matrices (coloured balls) and associated classification rates for the test dataset (CVID $n = 8$; HC $n = 12$). (**e**) Sensitivity and specificity of SVM models calculated using the corresponding 'accumulated hits' data (individual spectra).

PCA-LDA; and Fig. 4 and Supplemental Fig. 3(e,f) are FFS histograms illustrating the number of times each wavenumber was selected as a key feature for differentiating between the two classes. Figure 4 and Supplemental Fig. 3(g,h) visually represent the spectral wavenumber location where the 18 biomarkers were found using the three techniques. The points showing close proximity or overlap indicate the close agreement of the selected wavenumbers from the three methods of biomarker extraction.

Fingerprint region biomarkers. *Serum.* Within the serum a total of 10 spectral wavenumbers ($p = < 0.05$) were extracted. Wavenumber 1034 cm^{-1} was extracted by two different methods, which strengthens its utility as a serum biomarker.

Individual subject-level wavenumber intensity analysis was performed on each of the biomarkers (pre-processed (rubber-band, vector normalised) data). Four key peaks were shown to demonstrate significant differences ($p < 0.05$) between mean absorbance intensity for CVID and HC subjects (Fig. 5a–d); 1115 cm^{-1} (symmetric stretching P–O–C), 1034 cm^{-1} (collagen), 1528 cm^{-1} (C=N guanine, adenine, cytosine) and 1759 cm^{-1} (C=O ester group vibration of triglycerides). Intensity differences between the clinical sub-groups (patients with- and without further complications), were also explored. The increased absorbance intensity of wavenumbers 1115 cm^{-1}, 1034 cm^{-1} and 1759 cm^{-1} were statistically significant in CVID compared to the HC group; whereas the intensity of wavenumber 1528 cm^{-1} was lower. This finding was mirrored in the CVID sub-groups; patients with further complications demonstrated higher absorbance intensity for 1115 cm^{-1}, 1034 cm^{-1} and 1759 cm^{-1} compared to the patients without complications; with lower intensities observed for wavenumber 1528 cm^{-1}.

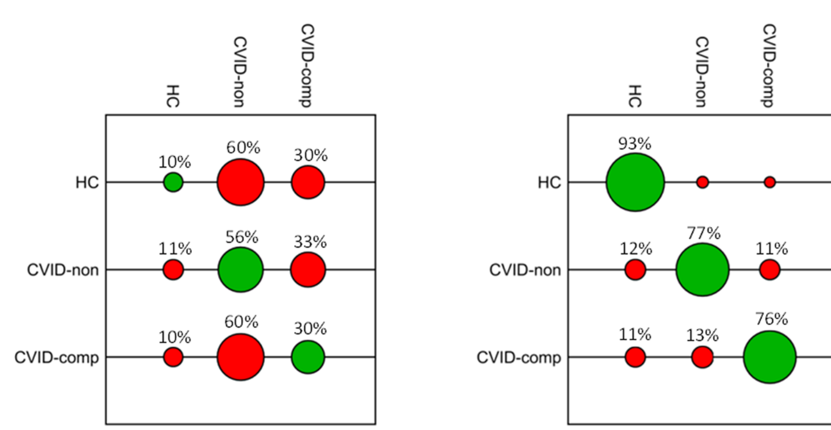
Plasma. 12 unique wavenumbers demonstrated significant absorbance intensity differences ($p = < 0.05$) within the plasma when comparing CVID patients and HC (Supplemental Table 4). Of these, 9 wavenumbers revealed

Serum Test SVM Classification

a) Fingerprint

Plasma Test SVM Classification

b) Fingerprint



c) SVM model sensitivities and specificities – Fingerprint region

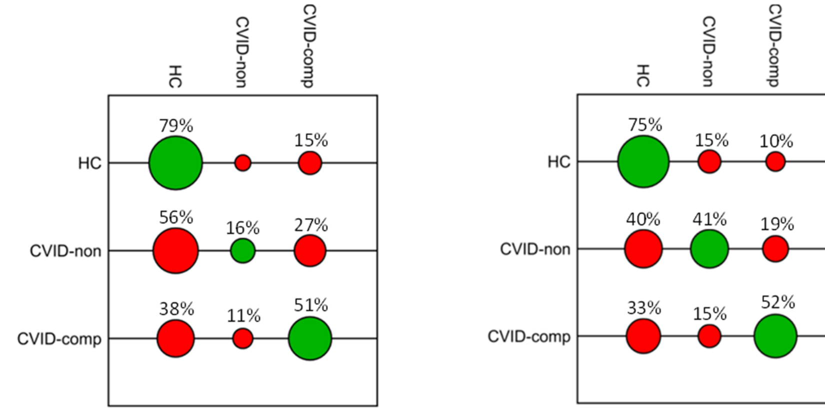
Serum Fingerprint	Sensitivity (%)	Specificity (%)
HC	11	91
CVID-non	65	41
CVID-comp	33	71
Plasma Fingerprint	**Sensitivity (%)**	**Specificity (%)**
HC	93	87
CVID-non	73	93
CVID-comp	73	95

Serum Test SVM Classification

d) High

Plasma Test SVM Classification

e) High



f) SVM model sensitivities and specificities – High region

Serum High	Sensitivity (%)	Specificity (%)
HC	75	61
CVID-non	39	86
CVID-comp	51	88
Plasma High	**Sensitivity (%)**	**Specificity (%)**
HC	78	54
CVID-non	15	92
CVID-comp	48	82

Figure 3. SVM classification model 'HC vs CVID-non complications vs CVID-complications' for each biofluid at the fingerprint (900–1800 cm^{-1}) and high region (2800–3700 cm^{-1}) of the spectrum. (**a**,**b**) SVM confusion matrices for a, serum fingerprint, (**b**) plasma fingerprint. The tuning parameters (**c**, γ) extracted from a grid search of the training dataset were used to subsequently generate confusion matrices (coloured balls) and associated classification rates for the test dataset (CVID $n = 8$; HC $n = 12$). (**c**) Sensitivity and specificity of SVM models using Fingerprint region, calculated using the corresponding 'accumulated hits' data (individual spectra). (**d**,**e**) SVM confusion matrices for (**c**) serum high and (**d**) plasma high regions. (**e**) Sensitivity and specificity of SVM models using High region, calculated using the corresponding 'accumulated hits' data (individual spectra).

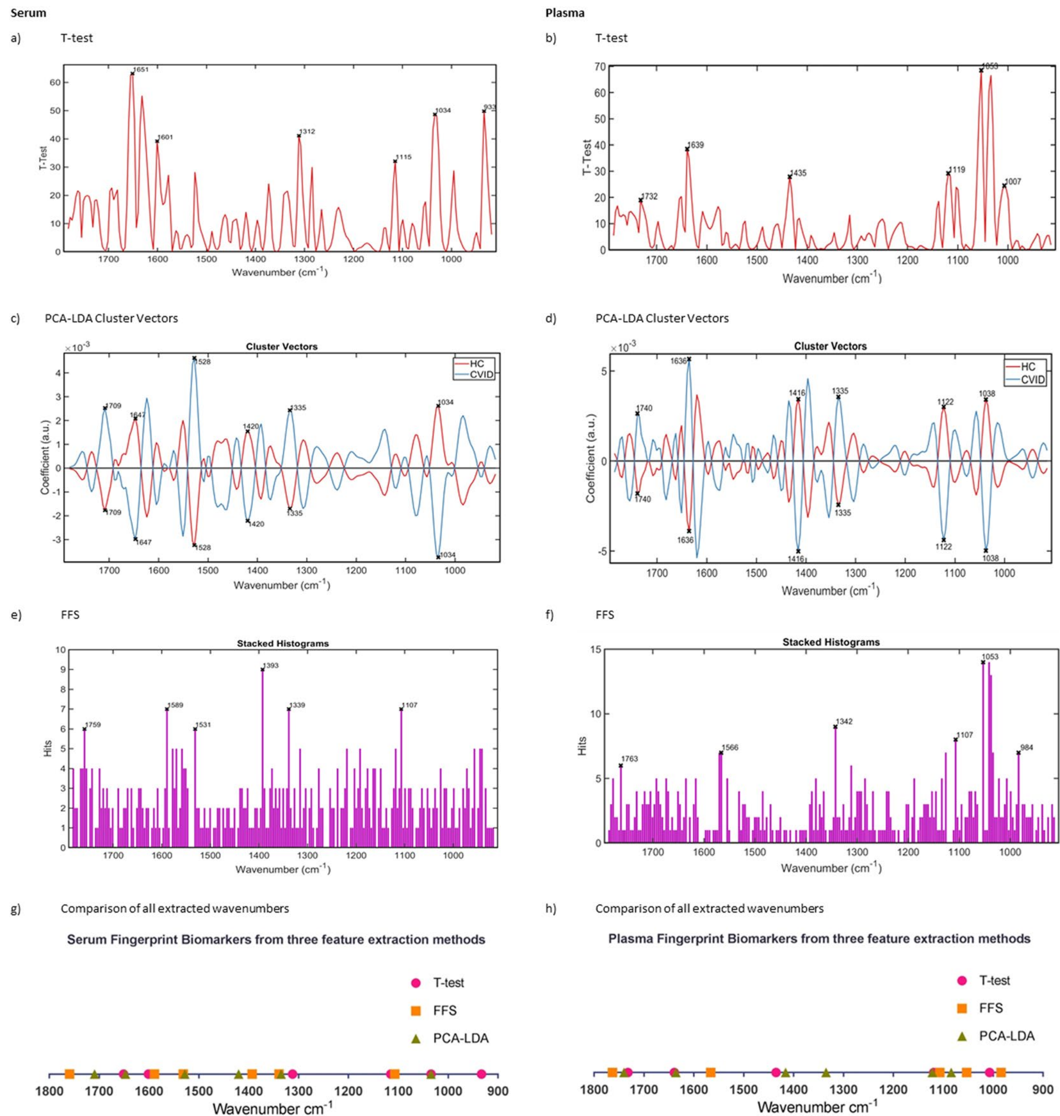


Figure 4. Serum and plasma Fingerprint region biomarkers identified using three feature extraction methods on the training dataset (CVID $n = 13$; HC $n = 18$). (**a**) Serum T-test. (**b**) Plasma T-test. (**c**) Serum FFS. (**d**) Plasma FFS. (**e**) Serum cross-validated PCA-LDA. (**f**) Plasma cross-validated PCA-LDA. (**i,j**) Visual representation of wavenumber location for extracted biomarkers from each method for serum and plasma respectively. FFS Forward Feature Selection.

significant differences on individual subject-level analysis (Fig. 6a–i); 984 cm^{-1}, 1007 cm^{-1}, 1053 cm^{-1}, 1084 cm^{-1}, 1107 cm^{-1} and 1119 cm^{-1} (within phosphodiester region 900^{-1} 300 cm^{-1}); 1416 cm^{-1}, 1566 cm^{-1} and 1639 cm^{-1} (within protein region; Amide I, Amide II). Wavenumber 1053 cm^{-1} was extracted by two different methods, strengthening its utility as a serum biomarker.

Absorbance intensities of 8 of the 9 wavenumbers increased in the CVID group compared to HC, with only wavenumber 1639 cm^{-1} (Amide I; thymine, adenine, guanine) demonstrating a lower absorbance intensity. Consistent with the serum data, the findings within the CVID subgroups reflected the intensity differences observed between the CVID patients and HC, with absorbance intensity for 8 of the 9 biomarkers found to be increased in the CVID patients with further complications compared to those patients without, and only wavenumber 1639 cm^{-1} demonstrating a lower intensity respectively.

Differences between HC subjects and CVID patients with further complications were calculated as significant for all 9 biomarkers. In contrast to the serum data, exploration of the plasma biomarkers revealed significant differences between HC and CVID patients without further complication. Wavenumbers 1107 cm^{-1}, 1119 cm^{-1} and 1416 cm^{-1} all demonstrated higher absorbance intensities in the CVID-non group compared to HC, whereas wavenumber 1639 cm^{-1} showed lower intensity. None of the 9 plasma wavenumbers demonstrated significant absorbance differences to segregate the two CVID subgroups.

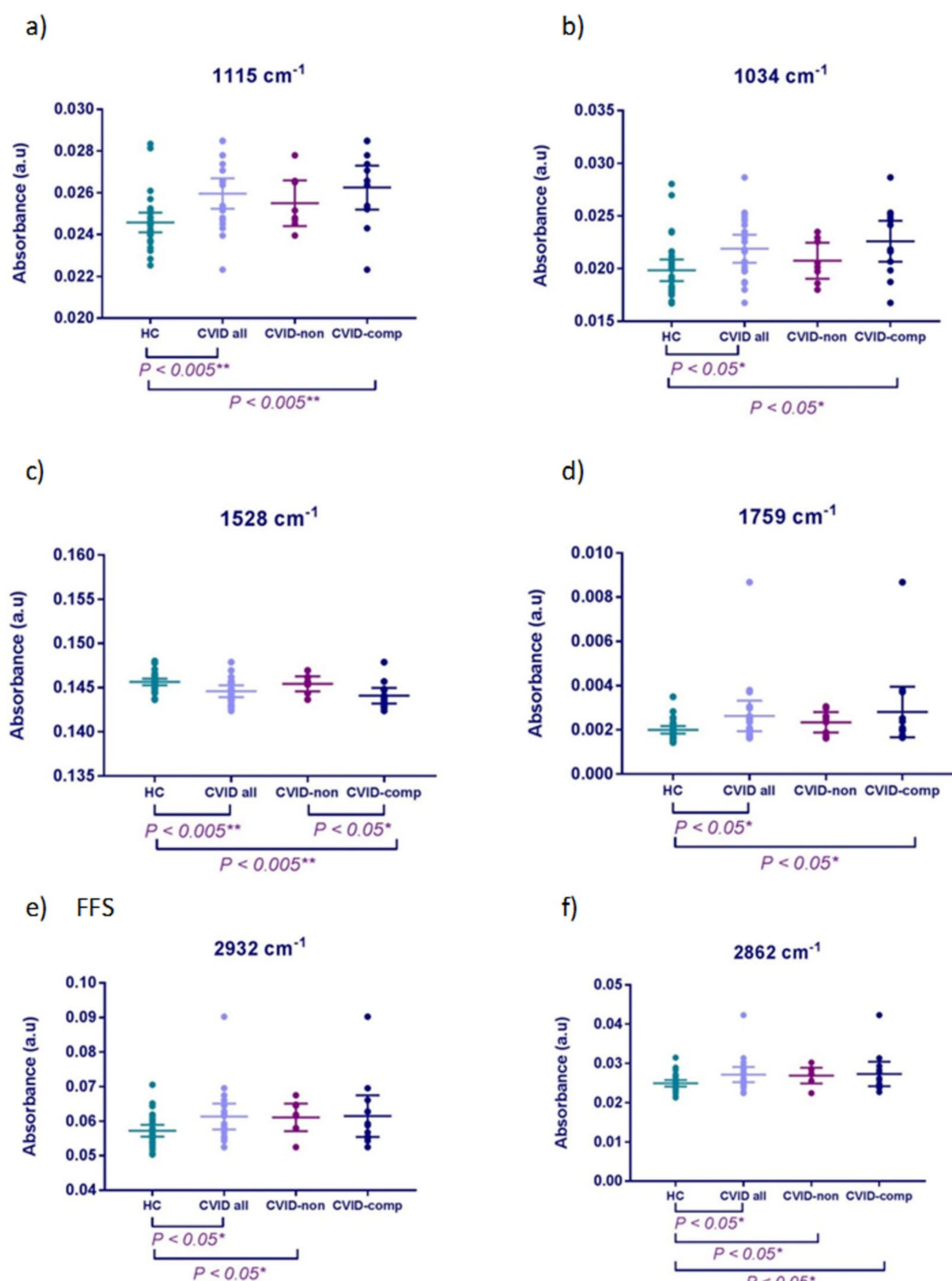


Figure 5. Serum Biomarkers. Between group absorbance intensity analysis for discriminating spectral wavenumbers. Mean absorbance intensity plotted for each study subject (20 replicates) per wavenumber. CVID $n = 21$ (CVID-non $n = 8$; CVID-comp $n = 13$); HC n = 30. Data are expressed as mean (±95% CI). *$P < 0.05$; **$P < 0.005$.

High region biomarkers. *Serum.* In the high region 11 biomarkers were calculated to be statistically significant when comparing intensities between the two classes (HC vs CVID) (Supplemental Table 5).

On individual subject-level analysis, 2 significant biomarkers ($p < 0.05$) were identified; 2932 cm^{-1} and 2862 cm^{-1}, corresponding to important CH, CH_2 and CH_3 molecular vibrations found in lipids and fatty acids (Fig. 5e,f). The absorbance intensity of both wavenumbers were increased within the CVID patients compared to HC; similarly, intensity increases were observed in CVID patients with further complications compared to those without. The increased absorbance intensities were calculated to be significant for HC vs CVID patients-with further complications, HC vs CVID patients-without complications, but not between the two CVID subgroups.

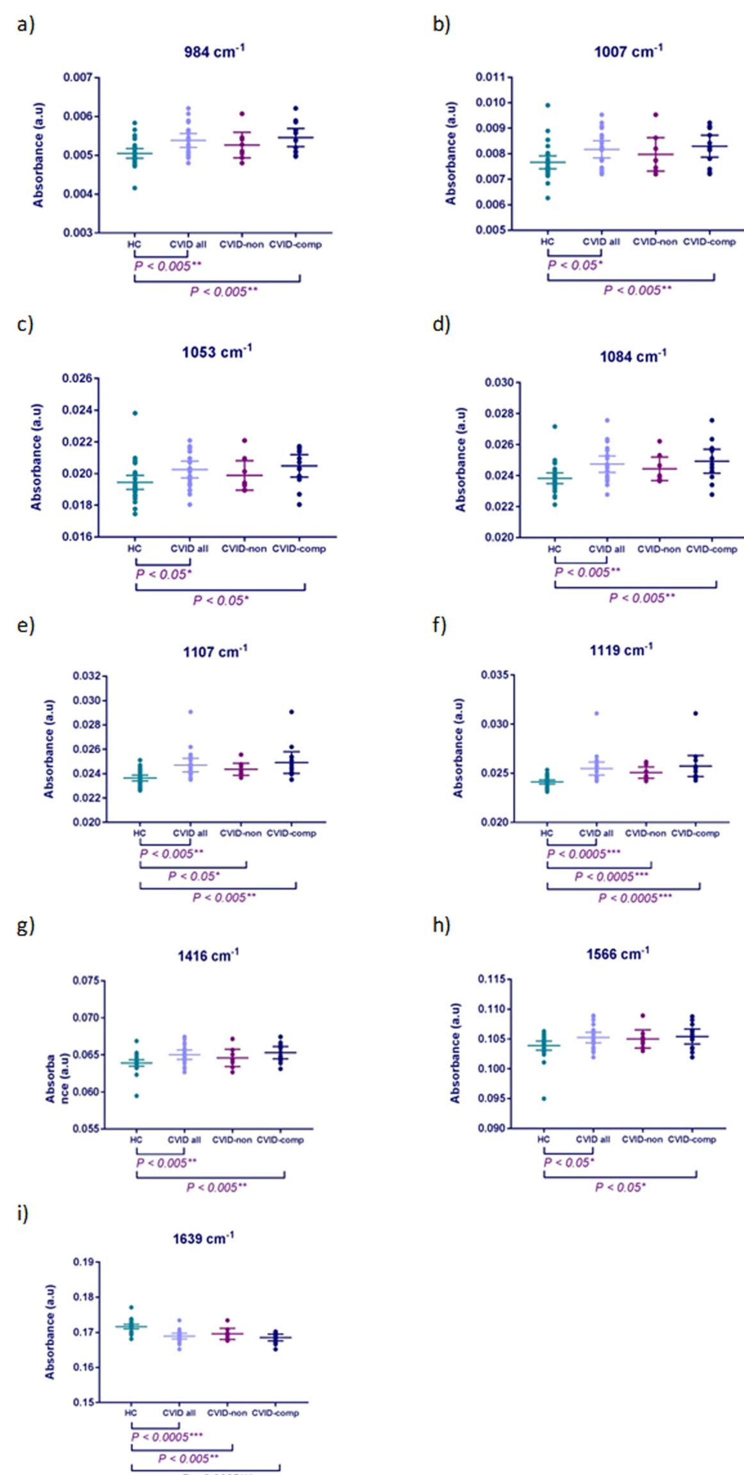


Figure 6. Plasma Biomarkers. Between group absorbance intensity analysis for discriminating spectral wavenumbers. Mean absorbance intensity plotted for each study subject (20 replicates) per wavenumber. CVID $n = 21$ (CVID-non $n = 8$; CVID-comp $n = 13$); HC $n = 30$. Data are expressed as mean (±95% CI). *$P < 0.05$; **$P < 0.005$; ***$P < 0.0005$.

Plasma. 10 biomarkers were found to demonstrate significant differences between CVID and HC groups (Supplemental Table 5). Only one wavenumber demonstrated significant intensity differences between HC and CVID patients on an individual subject level (3302 cm^{-1}) however the tentative assignment of this wavenumber to water renders it unsuitable for use as a biomarker.

Discriminating biomarkers between the CVID subgroups. Of all identified serum and plasma biomarkers (total 17) only one wavenumber, 1528 cm^{-1} (Fig. 5c) detected in the fingerprint region of the serum, was shown to be a discriminating biomarker between the CVID patients presenting with further complications (CVID-comp) and those remaining complication-free (CVID-non) ($p = 0.037$). The mean intensity of this wavenumber was reduced in both of the CVID groups compared to HC (HC = 0.1456, CVID-non = 0.1454, CVID-complications = 0.1455).

Discussion

Although great progress has been made in developing diagnostic and classification criteria for CVID[6,16–18,41], there is still no robust method to achieve this. This study has demonstrated the effectiveness of FTIR spectroscopic methods towards the diagnosis of CVID, correctly segregating CVID patients and HC into their respective groups following analysis of biofluids. This has been performed using a classification model on the fingerprint region of the ATR-FTIR spectrum with a sensitivity and specificity of 94% and 95% respectively for plasma, and 97% and 93% respectively for serum. The high region of the spectrum was similarly analysed, providing a classification model with a sensitivity and specificity of 55% and 69% respectively for plasma, and 66% and 91% respectively for serum; suggesting analysis of the fingerprint region would be more appropriate for classification of CVID. The use of blood-based vibrational spectroscopy to detect differences between clinical sub-groups of CVID patients has also been achieved; demonstrated by the successful assignment of individual study subjects into their respective groups; greatest sensitivities and specificities were achieved within the plasma fingerprint region, at 93% and 87% respectively for HC subjects, 73% and 94% for CVID patients without complications and 73% and 95% for CVID patients with further complications.

A further aim of the current study was to extract spectral biomarkers responsible for the differentiation between CVID patients and HC. As each wavenumber corresponds to molecular bonds within biochemical sample components, we could tentatively assign the most discriminating peaks for use as potential disease biomarkers. In the serum we found evidence of detectable alterations in band intensities at four wavenumbers 1034 cm^{-1} (collagen[42]), 1115 cm^{-1} (symmetric stretching P–O–C[43]), 1528 cm^{-1} (C=N guanine, adenine, cytosine[43,44]) and 1759 cm^{-1} (C=O ester group vibration of triglycerides[45,46]). We hypothesise that the significant increase in the collagen-associated peak (1034 cm^{-1}) observed in CVID patients may be associated with increased collagen turnover and production of degradation fragments following recurrent respiratory tract infection-driven lung damage. Abnormal distributions of extracellular matrix components, such as collagen type I and III, have been demonstrated in fibrotic lung conditions such as alveolitis, respiratory distress syndrome and chronic obstructive pulmonary disease (COPD)[47–50]; with degradation fragments detected and used as diagnostic makers in COPD and idiopathic pulmonary fibrosis[51]. We postulate 1034 cm^{-1} wavenumber analysis could have similar clinical utility, thus warranting a more detailed evaluation as a disease marker. Interestingly, in patients with further CVID-associated complications, a further increase in 1034 cm^{-1} peak intensity was observed, significantly segregating this group from the HCs ($p < 0.05$). This increase may signal the progression of early lung damage to bronchiectasis; if so, this marker could be used to identify high risk patients and avoid progression of this irreversible complication reported in over 2/3 of CVID patients[3]. Given the ubiquitous expression of collagen throughout the body, biomarker 1034 cm^{-1} intensity could reflect systemic serological levels, thus, further exploration in patients with other fibrotic co-morbidities or respiratory disorders are required to determine disease specificity.

Examination of nucleic acid-associated wavenumbers revealed a number of interesting observations. Three significant wavenumbers observed in plasma showed an increased intensity in CVID patients compared to HCs (984 cm^{-1}, 1053 cm^{-1}, 1084 cm^{-1}) as did wavenumber 1115 cm^{-1} in serum. Each of these wavenumbers have been assigned to bond vibrations found in molecules containing phosphodiester regions, PO_2 and P-O-C bonds, such as moieties found in the DNA/RNA sugar phosphate backbone[42,44,45,52–58]. Wavenumber intensity increases were accentuated further in the CVID-complications group compared to those patients without complications; albeit not to a statistically significant level ($p => 0.05$). We postulate that the increased trend in DNA/RNA-associated wavenumbers observed in the CVID group may be associated with increased levels of transcription for a number of genes involved in immune signalling pathways. Altered cytokine and chemokine profiles have been observed in CVID patients by several groups[59–65], thought to be driven by a state of chronic immune activation. This has been attributed to microbial activation of monocyte-macrophage and granulocyte lineages[59] or alternatively a predominance of the Th2 pathway[60]. CVID-associated profiles include increased serum levels of IL-4 and IL-10[60], IL-2 and IL-10[61], IL-6, IL-8, IL-1RA and TNF-a[62] and increased IL-10, IL-RA, and TNF-α[63]. Treatment with IVIg has been shown to dampen down this immune activation in CVID patients, although the exact mechanism of action remains unclear[59]. Of note within our study, we found CVID patients with further complications demonstrated a further intensity increase in the four DNA/RNA-associated wavenumbers. One suggestion for this could relate to poor control of immune activation due to higher requirements of replacement Ig in certain individuals. Further investigation into elevated serological levels of chemokines and cytokines in relation to key wavenumber intensity is warranted to determine the relationship between these CVID-associated biomarkers.

Surprisingly, wavenumbers 1528 cm^{-1} (in serum) and 1639 cm^{-1} (in plasma) were shown to have a decreased absorbance intensity in CVID patients compared to HC ($p < 0.005$ and $p < 0.0005$, respectively). These wavenumbers have been assigned more closely to C=N, C=C and N-H bonds found in nucleotide bases such as guanine, adenine, cytosine and thymine[37,44,66], as opposed to phosphate-associated moieties as in the previously described DNA/RNA-associated wavenumbers. We speculate this may reflect a general decrease in the nucleotide pool in CVID patients, and if so may elucidate some of the far-reaching biological and metabolic effects observed within this group. Since DNA replication is pertinent to chromosomal replication, amongst numerous other activities, these may all be directly or indirectly affected when nucleotide concentrations deviate from a physiologically normal range[67]. Unbalanced nucleotide pools have been linked to the activation of p53 and cell cycle arrest in actively dividing cells[68], therefore similar processes could relate to B cell maturational arrest in CVID.

The intensity in both wavenumbers 1528 cm^{-1} and 1639 cm^{-1} were further decreased in the CVID subgroup with further complications, compared to those without complications. Of heightened interest, wavenumber 1528 cm^{-1} was the only biomarker of the 18 identified to demonstrate significant differences in absorbance intensity between the two CVID subgroups ($p < 0.05$). This wavenumber could aid further elucidation of disease pathophysiology in addition to serving as a potential marker to determine severity of disease and development of further complications.

The findings from the first stage of this study are encouraging based on the impact that the translation of FTIR spectroscopy into a diagnostic platform for CVID could have on clinical practice. Whilst we have demonstrated the ability of this method to correctly classify CVID patients from HCs, diagnostic capabilities must be further established in subsequent multi-centre studies.

Within the setting of CVID, the diagnostic efficiency of current laboratory methods and FTIR spectroscopy will remain difficult to ascertain until disease-specific features and pathogenic disease mechanisms are further elucidated. The low diagnostic specificity of current tests such as serum immunoglobulins, vaccine responses and B cell immunophenotyping, and reliance on complex classification criteria for CVID, highlights the clinical requirement for an improved approach. In order to demonstrate the power of biospectroscopy as a novel diagnostic tool, the next phase of this study will work towards the validation and verification of the method. Test specificity will be addressed through inclusion of additional patient groups, across multiple centres, such as patients with other primary- or secondary immune deficiency disorders. Further work to fully determine how the identified FTIR biomarkers relate to the molecular and cellular composition of CVID patient samples will be a key milestone in determining whether any pathognomonic features can be identified. Until this is achieved, it is most likely that FTIR will be used alongside current diagnostic methods in order to add a further level of evidence-based criteria to the diagnosis of CVID.

Once achieved, a major advantage of using FTIR over current methods would be the capability for monitoring multiple biochemical changes in patient samples over a time-course analysis. The information collected would enable clinicians to adapt treatment options and undertake additional investigations in a timely manner. By detecting disease-associated complications earlier, before irreversible damage occurs, the life-expectancy of this patient group (in which secondary complications have the biggest impact), could potentialy be extended[8]. In contrast to some of the limitations facing FTIR spectroscopy analysis in other high-risk clinical areas such as malignancy, where it is often required as a one-shot investigation, CVID is a chronic, life-long condition and therefore would be a more suitable candidate for long-term monitoring.

In conclusion, our study has demonstrated that FTIR spectroscopy is a promising analytical tool for determining differences between healthy controls and CVID patients. A classification method based on the fingerprint region in serum was able to correctly discriminate up to 99% of the spectra representative of controls and 92% of spectra representative of disease, and for plasma, 96% of controls and 92% of disease. Several spectral wavenumbers have been identified as key biomarkers; each demonstrating significant statistical differences in band intensities when comparing subjects from the CVID and control groups. These biomarkers have been tentatively assigned to bond vibrations found in important biochemical moieties that should be further explored in relation to pathophysiological mechanisms causing CVID. This work therefore opens the way for the first application of FTIR spectroscopy in a clinical immunology laboratory, which could rapidly translate into a point-of-care device to enable 'while-you-wait' diagnostic testing in the immunodeficiency clinic.

Materials and Methods

Population. This study included 21 adult (>18 years old) CVID patients and 30 healthy age-matched controls recruited at Royal Preston Hospital. This study was approved by the ethics committee of the NHS Research Ethics Committee, Health Research Authority (HRA) (IRAS No. 212518). All samples were collected with informed written consent for study participation and all methods were carried out in accordance with relevant guidelines and regulations. Double-blinded unbiased acquisition of spectra was performed from all 51 samples following the allocation of a randomised unique study number to each subject at the point of recruitment. All patients clinically diagnosed with CVID fulfilled the European Society for Immunodeficiencies and the Pan American Group for Immunodeficiency (ESID/PAGID) (1999) diagnostic criteria[14]. The cohort characteristics are shown in Table 1 and more detailed patient demographics information is shown in Supplemental Table 1 [see Supplementary Information].

Immunoglobulin (IG) therapy. Management of CVID patients with immunoglobulin (IG) replacement was noted for the analysis the patient cohort results (see Supplemental Table 1). Serum immunoglobulin levels were measured for all patients and controls on the date of recruitment into the study, (IG-treated patient mean = 7.86 g/L, SD2.37; HC mean IgG = 10.48 g/L, SD = 1.72). Pre-treatment mean IgG levels for the patient cohort were 2.17 g/L, SD = 2.56.

Clinical sub-groups. The CVID patient cohort was sub-grouped based on the presence, or absence of additional clinical complications using data extracted from individual patient files, clinical notes, and laboratory test results. The CVID-complications group consisted of 13 patients in total. Each patient in this group had documented clinical history of one or more of the following complications: bronchiectasis ($n = 10$), autoimmunity ($n = 4$), splenomegaly ($n = 6$), malignancy ($n = 5$), or gastrointestinal complications ($n = 4$). The CVID complications-free group consisted of 8 patients; review of the available clinical history (clinical letters, patient discussion, electronic/paper notes) revealed no indication of any relevant clinical complications.

Sample Collection. Whole-blood samples were collected into EDTA-treated or serum gel-separator tubes and centrifuged at 110 × g for 5 min to separate the plasma or serum supernatant from the cells. Serum and

CVID patient demographics	
Clinical complications	n=
Bronchiectasis	10
Splenomegaly	6
AI	4
malignancy	5
ENT	4
GI	4
Passed away	2
Treatment	
IVIG	6
SC	13
NONE	2
IgG measurement (g/L)	
IgG levels < 6	6
IgG levels > 6	15
Age segregation (yrs)	
20–40	6
40–60	8
60+	7
Sex	
Female	9
Male	12

Table 1. Summary of CVID patient demographics.

plasma samples were then stored as 0.3 mL aliquots at −80 °C until required. Prior to spectroscopic analysis, individual aliquots were thawed; mixed and 50 µL from each aliquot was deposited onto IR-reflective glass slides (MirrIR Low-E slides; Kevley Technologies) in duplicate. Slides were left to air dry for up to 8 hours before being placed into a desiccator overnight. Once generated, dried blood spot slides were analysed the subsequent day. This process was undertaken for both serum and plasma samples.

ATR-FTIR spectral acquisition. The spectra were obtained using a Tensor 27 FTIR spectrometer with Helios ATR attachment (Bruker Optics Ltd) operated by OPUS 5.5 software. The sampling area, defined by the internal reflection element (a diamond crystal), was ≈250 × 250 µm. Spectral resolution was 8 cm^{-1} with two times zero-filling, giving a data spacing of 4 cm^{-1} over the range 4,000–400 cm^{-1}.

The acquisition of an FTIR spectrum involves collecting a 'single-beam' spectral measurement at one point within a sample. For each study subject, blood spots of both serum and plasma were produced in duplicate; 10 spectra were collected per 50 µL dried blood spot (total of 20 spectra per biofluid). In order to enlarge the area of acquisition and minimize bias associated with sample thickness and molecular heterogeneity, spectra were collected from 10 different point locations within each blood spot. In consideration of the well-described 'coffee ring effect'[69–71], point spectra from the peripheral edges of the dried blood spots were avoided. The diamond crystal was cleaned with distilled water and dried between samples and replicates. Pre-processing of spectra was performed according to recommended protocols[19,36,72].

Computational Analysis. The spectra files were pre-processed using the IRootLab toolbox (trevisanj.github.io/irootlab/), within MATLAB R2017a software (MathWorks). Initially, the 20 replica spectra per sample were averaged in order to work with a sample-based classification. Two pre-processing techniques were independently tested: (1) rubber-band baseline correction followed by vector normalisation[72] and (2) by Savitzky-Golay (SG) smoothing (second-order polynomial and nine filter coefficients)[73]. Once the spectra had been pre-processed, two regions of interest were extracted from the spectra; the Fingerprint region, which covers the area between wavenumbers 1800–900 cm^{-1}; and the High region, which covers wavenumbers 3700–2800 cm^{-1}.

Principal component analysis linear discriminant analysis (PCA-LDA) was used to observe inter-group differences by means of a linear discriminant function applied to the principal component analysis (PCA) scores[74]. PCA is an unsupervised classification technique of exploratory analysis that reduces the spectral dataset into a small number of principal components (PCs) responsible for the majority of the original data variance[75].

Support vector machine (SVM) is a supervised machine-learning was applied for classifying data. The data was pre-processed as above using SG-2nd differential baseline correction and de-noising and vector normalisation. We used an SVM algorithm performed in MATLAB, run with an n-fold leave one out, cross validation technique ($n = 5$) to select the best parameters for c and gamma (γ). The parameters (c, γ) for SVM are selected by using a grid search function in MATLAB[72]. To investigate the classification rate, sensitivities and specificities were calculated for each model tested[76]. The SVM was trained using 2/3 of the spectral data and tested using the remaining 1/3. The data set was split using the Kennard-Stone algorithm to achieve uniformity and representativeness within

the samples selected for the training set[77]. This splitting process was performed in a patient basis, where the spectral data assigned to the training and test sets were from different samples, so the training and test groups do not contain spectra from the same patient. The models were built using 10-fold cross-validation for optimization. The classification percentage calculated from the confusion balls (graphical representation of a confusion matrix) of each SVM model designates the rate of correct group assignation when applying the test dataset to the trained SVM model. Sensitivity and specificity of each SVM classification was calculated using the accumulative hits data (number of true positives, true negatives, false positives, and false negatives) generated from the confusion matrices.

Feature extraction was performed on the training dataset to extract potential biomarkers and identify the spectral wavenumbers that account for the largest differences between the CVID and HC groups. This was undertaken using three methods of biomarker extraction on the training dataset for serum and plasma: Student's T-Test, PCA-LDA and Feature Forward Selection (FFS), for both Fingerprint and High regions of the spectra. The six key biomarkers extracted from each method were subsequently investigated for relative increases or decreases in absorbance intensity between the classes (subject groups). Wavenumbers not demonstrating significant intensity variance between CVID and HC groups were not taken forward for individual subject level intensity analysis (using average intensities of 20 spectral replicates). Extracted wavenumbers within close proximity ($10\,cm^{-1}$) of an adjacent biomarker were omitted, as closely associated wavenumbers will be influenced from intensity increases or decreases in nearby peaks already identified as biomarkers.

The Student's T-Test method was performed on the training dataset for both fingerprint and high regions of the spectra. The −log10 of the P-value of the T-test for each wavenumber was then plotted to identify the potential biomarkers from the T-test. The biomarkers extracted following PCA-LDA were obtained from the cluster vector analysis. FFS was applied within IRootLab using the PCA loadings to identify the main biomarkers responsible for class segregation by calculating p-values for the variables with larger loadings coefficients[78]. A peak detection algorithm was applied to each method to identify the six most segregating peaks. Extracted wavenumbers within $10\,cm^{-1}$ proximity of an adjacent biomarker were also omitted ($n = 2$) resulting in a total of 10 spectral wavenumbers ($p =< 0.05$).

Statistical analysis. A student's t-test (two–tailed, 95% confidence interval (CI)) was performed to calculate statistical significance of spectral variance between groups, with a P-value of less than 0.05 being considered significant. A power test based on a two-tailed t-test (data input as mean and standard-deviation of the plasma pre-processed spectra in the fingerprint region for each class) indicated a minimum number of samples of 26 HC and 15 CVID patients for a power of 80%. The number of samples used herein (HC = 30, CVID = 21) is above this minimum. Statistical analysis was carried out on averaged spectra to account for differences between individuals and not spectra.

Data deposition. The data (raw spectra and pre-processed spectra) reported in this paper are available at the publicly accessible data repository Figshare (https://doi.org/10.6084/m9.figshare.7751309).

Data Availability

All data (raw and pre-processed spectra) along with appropriate code identifiers will be uploaded onto the publicly accessible data repository Figshare.

References

1. Boyle, J. M. & Buckley, R. H. Population prevalence of diagnosed primary immunodeficiency diseases in the United States. *J. Clin. Immunol.* **27**, 497–502 (2007).
2. Chapel, H. & Cunningham-Rundles, C. Update in understanding common variable immunodeficiency disorders (CVIDs) and the management of patients with these conditions. *Br. J. Haematol.* **145**, 709–727 (2009).
3. Gathmann, B. *et al*. Clinical picture and treatment of 2212 patients with common variable immunodeficiency. *J. Allergy Clin. Immunol.* **134**, 116–126.e11 (2014).
4. Picard, C. *et al*. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency 2015. *J. Clin. Immunol.* **35**, 696–726 (2015).
5. Bonilla, F. A. *et al*. International consensus document (ICON): common variable immunodeficiency disorders. *J. Allergy Clin. Immunol. Pract.* **4**, 38–59 (2016).
6. Chapel, H. *et al*. Common variable immunodeficiency disorders: division into distinct clinical phenotypes. *Blood* **112**, 277–286 (2008).
7. Cunningham-Rundles, C. The many faces of common variable immunodeficiency. *Hematol. Am. Soc. Hematol. Educ. Program* **2012**, 301–305 (2012).
8. Cunningham-Rundles, C. & Bodian, C. Common variable immunodeficiency: clinical and immunological features of 248 Patients. *Clin. Immunol.* **92**, 34–48 (1999).
9. Bacchelli, C., Buckridge, S., Thrasher, A. J. & Gaspar, H. B. Translational mini-review series on immunodeficiency: molecular defects in common variable immunodeficiency. *Clin. Exp. Immunol.* **149**, 401–409 (2007).
10. Salzer, U., Warnatz, K. & Peter, H. H. Common variable immunodeficiency - an update. *Arthritis Res. Ther.* **14**, 223 (2012).
11. Li, J. *et al*. Understanding the genetic and epigenetic basis of common variable immunodeficiency disorder through omics approaches. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1860**, 2656–2663 (2016).
12. Castigli, E. & Geha, R. Molecular basis of common variable immunodeficiency. *J. Allergy Clin. Immunol.* **117**, 740–746 (2006).
13. Ameratunga, R., Woon, S.-T., Gillis, D., Koopmans, W. & Steele, R. New diagnostic criteria for common variable immune deficiency (CVID), which may assist with decisions to treat with intravenous or subcutaneous immunoglobulin. *Clin. Exp. Immunol.* **174**, 203–211 (2013).
14. Conley, M. E., Notarangelo, L. D. & Etzioni, A. Diagnostic criteria for primary immunodeficiencies. *Clin. Immunol.* **93**, 190–197 (1999).
15. Perez, E. E. *et al*. Update on the use of immunoglobulin in human disease: A review of evidence. *J. Allergy Clin. Immunol.* **139**, S1–S46 (2017).
16. Piqueras, B. *et al*. Common variable immunodeficiency patient classification based on impaired B cell memory differentiation correlates with clinical aspects. *J. Clin. Immunol.* **23**, 385–400 (2003).

17. Wehr, C. *et al*. The EUROclass trial: defining subgroups in common variable immunodeficiency. *Blood* **111**, 77–85 (2008).
18. Warnatz, K. *et al*. Severe deficiency of switched memory B cells (CD27+IgM−IgD−) in subgroups of patients with common variable immunodeficiency: a new approach to classify a heterogeneous disease. *Blood* **99**, 1544–1551 (2002).
19. Baker, M. J. *et al*. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **9**, 1771–1791 (2014).
20. Butler, H. J. *et al*. Using raman spectroscopy to characterize biological materials. *Nat Protoc.* **11**, 664–687 (2016).
21. Movasaghi, Z., Rehman, S. & ur Rehman, D. I. Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **43**, 134–179 (2008).
22. Paraskevaidi, M. *et al*. Differential diagnosis of Alzheimer's disease using spectrochemical analysis of blood. *Proc. Natl. Acad. Sci.* **114**, E7929–E7938 (2017).
23. Peuchant, E. *et al*. Infrared spectroscopy: a reagent-free method to distinguish Alzheimer's disease patients from normal-aging subjects. *Transl. Res. J. Lab. Clin. Med.* **152**, 103–112 (2008).
24. Yonar, D., Ocek, L., Tiftikcioglu, B. I., Zorlu, Y. & Severcan, F. Relapsing-remitting multiple sclerosis diagnosis from cerebrospinal fluids via fourier transform infrared spectroscopy coupled with multivariate analysis. *Sci. Rep.* **8**, 1025 (2018).
25. Abbas, S. *et al*. Diagnosis of malignant pleural mesothelioma from pleural fluid by fourier transform-infrared spectroscopy coupled with chemometrics. *J. Biomed. Opt.* **23**, 1–14 (2018).
26. Depciuch, J. *et al*. Phospholipid-protein balance in affective disorders: analysis of human blood serum using raman and FTIR spectroscopy. A pilot study. *J. Pharm. Biomed. Anal.* **131**, 287–296 (2016).
27. Sitole, L., Steffens, F., Krüger, T. P. J. & Meyer, D. Mid-ATR-FTIR spectroscopic profiling of HIV/AIDS sera for novel systems diagnostics in global health. *OMICS J. Integr. Biol.* **18**, 513–523 (2014).
28. Firdous, S. *et al*. Measurement of diabetic sugar concentration in human blood using Raman spectroscopy. *Laser Phys.* **22**, 1090–1094 (2012).
29. Backhaus, J. *et al*. Diagnosis of breast cancer with infrared spectroscopy from serum samples. *Vib. Spectrosc.* **52**, 173–177 (2010).
30. Bunaciu, A. A., Hoang, V. D. & Aboul-Enein, H. Y. Applications of FT-IR spectrophotometry in cancer diagnostics. *Crit. Rev. Anal. Chem.* **45**, 156–165 (2015).
31. Paraskevaidi, M. *et al*. Potential of mid-infrared spectroscopy as a non-invasive diagnostic test in urine for endometrial or ovarian cancer. *The Analyst* **143**, 3156–3163 (2018).
32. Wang, X., Shen, X., Sheng, D., Chen, X. & Liu, X. FTIR spectroscopic comparison of serum from lung cancer patients and healthy persons. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **122**, 193–197 (2014).
33. Baker, M. J. *et al*. Developing and understanding biofluid vibrational spectroscopy: a critical review. *Chem. Soc. Rev.* **45**, 1803–1818 (2016).
34. Hands, J. R. *et al*. Brain tumour differentiation: rapid stratified serum diagnostics via attenuated total reflection fourier-transform infrared spectroscopy. *J. Neurooncol.* **127**, 463–472 (2016).
35. Khanmohammadi, M. & Garmarudi, A. B. Infrared spectroscopy provides a green analytical chemistry tool for direct diagnosis of cancer. *TrAC Trends Anal. Chem.* **30**, 864–874 (2011).
36. Martin, F. L. *et al*. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nat. Protoc.* **5**, 1748–1760 (2010).
37. Talari, A. C. S., Martinez, M. A. G., Movasaghi, Z., Rehman, S. & Rehman, I. U. Advances in Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **52**, 456–506 (2017).
38. Theophilou, G., G. Lima, K. M., L. Martin-Hirsch, P., F. Stringfellow, H. & L. Martin, F. ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer. *Analyst* **141**, 585–594 (2016).
39. Byler, D. M. & Susi, H. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **25**, 469–487 (1986).
40. Jackson, M. & Mantsch, H. H. The use and misuse of FTIR spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. Mol. Biol.* **30**, 95–120 (1995).
41. Driessen, G. J. *et al*. B-cell replication history and somatic hypermutation status identify distinct pathophysiologic backgrounds in common variable immunodeficiency. *Blood* **118**, 6814–6823 (2011).
42. Fabian, H. *et al*. A comparative infrared spectroscopic study of human breast tumors and breast tumor cell xenografts. *Biospectroscopy* **1**, 37–45 (1995).
43. Thumanu, K. *et al*. Diagnosis of liver cancer from blood sera using FTIR microspectroscopy: a preliminary study. *J. Biophotonics* **7**, 222–231 (2014).
44. Mossoba, M. M. *et al*. Printing microarrays of bacteria for identification by infrared microspectroscopy. *Vib. Spectrosc.* **38**, 229–235 (2005).
45. Caine, S., Heraud, P., Tobin, M. J., McNaughton, D. & Bernard, C. C. A. The application of fourier transform infrared microspectroscopy for the study of diseased central nervous system tissue. *NeuroImage* **59**, 3624–3640 (2012).
46. Stone, N., Kendall, C., Smith, J., Crow, P. & Barr, H. Raman spectroscopy for identification of epithelial cancers. *Faraday Discuss.* **126**, 141–157 (2004). discussion 169–183.
47. Bihlet, A. R. *et al*. Biomarkers of extracellular matrix turnover are associated with emphysema and eosinophilic-bronchitis in COPD. *Respir. Res.* **18**, 22 (2017).
48. Kirk, J. M., Bateman, E. D., Haslam, P. L., Laurent, G. J. & Turner-Warwick, M. Serum type III procollagen peptide concentration in cryptogenic fibrosing alveolitis and its clinical relevance. *Thorax* **39**, 726–732 (1984).
49. Last, J. A., Siefkin, A. D. & Reiser, K. M. Type I collagen content is increased in lungs of patients with adult respiratory distress syndrome. *Thorax* **38**, 364–368 (1983).
50. Seyer, J. M., Hutcheson, E. T. & Kang, A. H. Collagen polymorphism in idiopathic chronic pulmonary fibrosis. *J. Clin. Invest.* **57**, 1498–1507 (1976).
51. Leeming, D. J. *et al*. Serological investigation of the collagen degradation profile of patients with chronic obstructive pulmonary disease or idiopathic pulmonary fibrosis. *Biomark. Insights* **7**, BMI.S9415 (2012).
52. Wehbe, K. *et al*. Differentiation between normal and tumor vasculature of animal and human glioma by FTIR imaging. *The Analyst* **135**, 3052–3059 (2010).
53. Wood, B. R. *et al*. FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies. *Biospectroscopy* **4**, 75–91 (1998).
54. Yang, Y. *et al*. Study of tumor cell invasion by fourier transform infrared microspectroscopy. *Biopolymers* **78**, 311–317 (2005).
55. Yano, K. *et al*. Direct measurement of human lung cancerous and noncancerous tissues by fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool? *Anal. Biochem.* **287**, 218–225 (2000).
56. Zelig, U. *et al*. Pre-screening and follow-up of childhood acute leukemia using biochemical infrared analysis of peripheral blood mononuclear cells. *Biochim. Biophys. Acta* **1810**, 827–835 (2011).
57. Fung, M. F. K., Senterman, M. K., Mikhael, N. Z., Lacelle, S. & Wong, P. T. T. Pressure-tuning fourier transform infrared spectroscopic study of carcinogenesis in human endometrium. *Biospectroscopy* **2**, 155–165 (1998).
58. Fukuyama, Y., Yoshida, S., Yanagisawa, S. & Shimizu, M. A study on the differences between oral squamous cell carcinomas and normal oral mucosas measured by fourier transform infrared spectroscopy. *Biospectroscopy* **5**, 117–126 (1999).

59. Paquin-Proulx, D. & Sandberg, J. K. Persistent immune activation in CVID and the role of IVIg in its suppression. *Front. Immunol.* **5**, (2014).
60. Rezaei, N. *et al*. B-cell–T-cell activation and interaction in common variable immunodeficiency. *Hum. Immunol.* **71**, 355–362 (2010).
61. Kasztalska, K., Ciebiada, M., Cebula-Obrzut, B. & Górski, P. Intravenous immunoglobulin replacement therapy in the treatment of patients with common variable immunodeficiency disease: an open-label prospective study. *Clin. Drug Investig.* **31**, 299–307 (2011).
62. Ibáñez, C. *et al*. Modulating effects of intravenous immunoglobulins on serum cytokine levels in patients with primary hypogammaglobulinemia. *BioDrugs Clin. Immunother. Biopharm. Gene Ther.* **19**, 59–65 (2005).
63. Hel, Z. *et al*. Altered serum cytokine signature in common variable immunodeficiency. *J. Clin. Immunol.* **34**, 971–978 (2014).
64. Litzman, J. *et al*. Chronic immune activation in common variable immunodeficiency (CVID) is associated with elevated serum levels of soluble CD14 and CD25 but not endotoxaemia. *Clin. Exp. Immunol.* **170**, 321–332 (2012).
65. North, M. E., Spickett, G. P., Webster, A. D. & Farrant, J. Raised serum levels of CD8, CD25 and beta 2-microglobulin in common variable immunodeficiency. *Clin. Exp. Immunol.* **86**, 252–255 (1991).
66. Choo, L.-P. *et al*. Infrared spectra of human central nervous system tissue: Diagnosis of alzheimer's disease by multivariate analyses. *Biospectroscopy* **1**, 141–148 (1995).
67. Pai, C.-C. & Kearsey, S. E. A critical balance: dNTPs and the maintenance of genome stability. *Genes* **8** (2017).
68. Hastak, K. *et al*. DNA synthesis from unbalanced nucleotide pools causes limited DNA damage that triggers ATR-CHK1-dependent p53 activation. *Proc. Natl. Acad. Sci. USA* **105**, 6314–6319 (2008).
69. Choi, S. & Birarda, G. Protein mixture segregation at coffee-ring: real-time imaging of protein ring precipitation by FTIR spectromicroscopy. *J. Phys. Chem. B* **121**, 7359–7365 (2017).
70. Yunker, P. J., Still, T., Lohr, M. A. & Yodh, A. G. Suppression of the coffee-ring effect by shape-dependent capillary interactions. *Nature* **476**, 308–311 (2011).
71. Deegan, R. D. *et al*. Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389**, 827–829 (1997).
72. Trevisan, J., Angelov, P. P., Carmichael, P. L., Scott, A. D. & Martin, F. L. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst* **137**, 3202–3215 (2012).
73. Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
74. Morais, C. L. M., Lima, K. M. G., Morais, C. L. M. & Lima, K. M. G. Principal component analysis with linear and quadratic discriminant analysis for identification of cancer samples based on mass spectrometry. *J. Braz. Chem. Soc.* **29**, 472–481 (2018).
75. Bro, R. & Smilde, A. K. Principal component analysis. *Anal. Methods* **6**, 2812–2831 (2014).
76. Siqueira, L. F. S., Araújo Júnior, R. F., de Araújo, A. A., Morais, C. L. M. & Lima, K. M. G. LDA vs. QDA for FT-MIR prostate cancer tissue classification. *Chemom. Intell. Lab. Syst.* **162**, 123–129 (2017).
77. Kennard, R. W. & Stone, L. A. Computer aided design of experiments. *Technometrics* **11**, 137–148 (1969).
78. Trevisan, J., Angelov, P. P., Scott, A. D., Carmichael, P. L. & Martin, F. L. IRootLab: a free and open-source MATLAB toolbox for vibrational biospectroscopy data analysis. *Bioinformatics* **29**, 1095–1097 (2013).

Author Contributions

Emma L. Callery designed and performed the experiments, analysed and interpreted data, and wrote the manuscript. Anthony W. Rowbottom initiated, designed and led the research, helped write the manuscript, and provided discussions. Francis L. Martin initiated, designed and supervised the research. Vladimir Brusic designed the research, provided intellectual input and helped write the manuscript. Pavaladurai Vijayadurai and Ariharan Anantharachagan provided patient samples, clinical data and discussions. Camilo L. M. Morais provided chemometric support and analysed data. Maria Paraskevaidi provided help with the experiments, intellectual input and discussions.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-43196-5.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019

Chapter 4: Critical Appraisal of the Research Project

Chapter 4

Critical Appraisal of the Research Project

Abstract

Biospectroscopy has been gaining wider acceptance and application in the clinical setting over the past decade; however, it has yet to reach NHS laboratories and healthcare clinics as a routine platform for clinical assessment. In this research project we have used ATR-FTIR spectroscopy and multivariate analysis tools to examine its application to the primary immunodeficiency disorder, common variable immune deficiency (CVID).

To our knowledge, this is the first time that these methods have been applied to the investigation of CVID. In this final chapter we critically appraise the project, focussing on the choice of methods used, the disease population studied, and the contribution provided to theoretical and clinical practice.

The advantages offered by biospectroscopy include low-cost, whole-sample measurements without sample destruction, minimal sample preparation, and use of small sample volumes. Disadvantages include a lack of best practice guidance for analytical protocols, and the requirement for specialist staff for computational analysis. The classification power of this platform has been illustrated within our study, a strength that has been well-documented in the literature, with seven decades of cancer research studies demonstrating diagnostic accuracies in the order of 80-100%.

However, further progression towards clinical translation is being hindered by the unfamiliarity of the method, the limited repertoire of pathological conditions being investigated, and the fact that the majority of studies remain at the proof-of-concept stage, yet to be validated by larger-scale studies.

The scope for further work will be considered at the end of the chapter. We propose experiments to add value and strengthen the validity of our findings, and future work using Raman Spectroscopy, with the goal to embed vibrational spectroscopy into the immunology laboratory testing pathway for CVID.

4.1 Introduction

Biospectroscopy is an emerging field used to provide new insights into biological questions by coupling computational analysis with physical sciences. Spectroscopy

measurements and spectral patterns are derived following the absorption, emission or scattering of electromagnetic radiation by atoms or molecules. The application of spectroscopy methods to investigate biological material is termed biospectroscopy. The features of a spectrum generated from interrogating biological material are directly related to the molecular structure of the biological compounds within the sample. This includes specific absorption or emission wavenumber bands, their intensity and their shape; collectively known as the 'biochemical fingerprint'[1,2].

The detection of conventional single-analyte disease markers, commonly measured in biofluids and tissues are unlikely to be appropriate for heterogeneous, multidimensional disorders. Instead, there is an increasing look to apply a systems biology approach, such as genome-wide screening and mathematical modelling to add new knowledge and fulfil the void in appropriate diagnostic biomarkers. However, for the majority of immunological disorders it has not been possible to find the monogenic link to the disease[3–8]; this therefore limits the clinical utility of a genomic approach in routine diagnostics and an alternative strategy is required.

Other 'omics-based methods (proteomics, transcriptomics, metabolomics and epigenetics) have been applied to the study of immune-mediated disorders; a recent review of this topic has been published by Chu et al.[9]. The systems biology approach has been widely applied to complex diseases such as cancer, and more recently to immune-mediated pathologies[10–14]. The ultimate goal is to stratify heterogeneous patient groups and enable the progression of personalised medicine. The work presented in this thesis aligns with this new approach, as we have applied an analytical technology (biospectroscopy), capable of whole sample profiling, to the investigation of a complex and clinically heterogeneous immunological disorder.

The use of biospectroscopy as a screening, diagnostic or prognostic tool in clinical practice is gaining more widespread acceptance following an expansion of studies over the past decade; reports of using biospectroscopy to successfully diagnose cancers of the bladder, ovary, brain, breast, cervix, and colon in addition to applications in transplant rejection, cardiology and biomarker detection have ignited interest in applying this platform within clinical immunology (as described in Chapter 1, Section 1.1). As evidenced by the small number of studies in autoimmunity, allergy and immune deficiency, biospectroscopy is still in its infancy within clinical immunology[15].

If the full potential of biospectroscopy is to be harnessed, both as a diagnostic tool and to advance our theoretical knowledge of disease mechanisms, the frequency and design of spectroscopy studies must develop past the stage of 'proof-of-concept'. We envisage this will require strong clinical and academic partnerships, and a 'multi-omics' approach to gain further insight into the molecular assignments of key disease-specific spectral patterns.

In Chapters 2 and 3, the clinically heterogeneous primary immune deficiency disorder, Common Variable Immune Deficiency (CVID) was selected as the disease population to investigate with biospectroscopy. Within this group, we assess the suitability of applying ATR-FTIR spectroscopy and machine learning approaches for diagnostic classification and biomarker identification. As, to our knowledge, this is the first time such methods have been applied, we cannot compare our findings to those reported by other researchers.

Chapter 2 describes and evaluates the early data processing techniques that can be applied to big data such as pre-processing, principal component analysis (PCA) and PCA-linear discriminant analysis (PCA-LDA). These techniques not only apply to biospectroscopy data, but are used during the analysis of data produced using the other 'omics-based method. We consider vibrational biospectroscopy aligns well with 'omics-based approaches as a method group, in that it generates large complex datasets which require the application of machine learning and artificial intelligence to extract meaningful results from an abundance of background data[16].

In this study we successfully applied data reduction steps (PCA) to identify discriminating features within the data that influence separation, and subsequently enable classification of CVID patient cohorts from healthy controls (HCs); this technique has also been demonstrated within similar sized CVID studies using proteomics[17,18]. In addition, we demonstrate the importance of controlling pre-analytical factors in the production of unbiased and robust data. The selection of pre-processing techniques, evaluation of sample types, and use of a single researcher to perform sample processing have helped to control unwanted sources of variation which could skew the final classification models.

As a novel method within immune deficiency research, the selection of spectroscopy technique and computational methods used throughout the study have been influenced by the work of other researchers within the group, alongside guidance from published

protocols[1,2,19–25]. The use of complex mathematical modelling brings inherent variation; therefore prior to the development of a diagnostic platform for CVID, the classification models generated within this study would require further validation in this setting. This chapter provides a platform to critically reflect on the methods used within this project; to discuss the advantages and limitations, the theoretical benefits to clinical practice and to put forward suggestions for future work.

4.2 Critical appraisal of the study protocol and methodological approach

4.2.1 Common variable immune deficiency as the study population

CVID is the most common, symptomatic primary immunodeficiency (PID) worldwide[26–28]. It is now understood that rather than a discrete disease, CVID encompasses a wide spectrum of disorders which present with a heterogeneous (hence the name ‘variable’), clinical picture but share similar laboratory findings of hypogammaglobulinemia and poor vaccine responses[27]. For these reasons, as highlighted throughout this thesis, diagnosis remains one of exclusion, and a single pathognomonic clinical or laboratory feature has yet to be discovered, or more likely, does not exist.

Identification of CVID patients therefore relies on diagnostic criteria, which have been developed based on both clinical and laboratory findings[28–30]. Diagnostic delay remains a major issue for these patients, and whilst infectious complications are the most frequent clinical feature, non-infectious manifestations (autoimmunity, lung disease, gastrointestinal inflammatory disease, malignancy) related to immune dysregulation are associated with increased morbidity and mortality within the CVID patient group[31,32].

The rationale for choosing a cohort of CVID patients as a study population is based on a number of factors, some of which have been highlighted in the summary of the disorder above. Firstly, to our knowledge, vibrational spectroscopy has not been applied to the study of any primary or secondary immunodeficiency disorder; therefore immune deficiency provided a novel area to investigate within this research project. Given the clinical variability of CVID, it is likely that multiple immune regulatory pathways are implicated in causing the shared laboratory finding of hypogammaglobulinemia. Our understanding of these pathways has increased over the last four decades, however the underlying cause, and the factors influencing whether CVID patients present with- or without further inflammatory complications has yet to be defined[33].

As described in Chapter 1, Section 1.2, this is beginning to be elucidated as recent proof-of-concept studies using proteomics and machine-learning methods have identified signature panels of immune-related markers specific to CVID patients with immune dysregulation, not found in patients with infectious complications only[17,18]. These studies were published in 2020 and 2021, respectively, at a later date to our published research manuscript (2019). Although different technologies have been used for sample analysis, these recent papers illustrate a new trend for CVID research. Rather than examining single-, or small numbers of discrete analytes, whole sample profiling and application of computational methods provides a more suitable approach to study undefined disorders with variable clinical phenotypes.

The heterogeneity of CVID makes it a good candidate disease for whole sample profiling; in transcriptomic and proteomic studies in CVID, thousands of immune-related RNA transcripts and proteins can be assessed to produce disease- or disease-subgroup-related signatures. Likewise, in a single experiment using biospectroscopy, the dataset obtained from each sample can contain around 3600 variables relating to wavenumber absorbance values[34]. In this study, the identification of unique spectral patterns (biochemical fingerprint) associated with samples from CVID patients has upheld our proposition - that biospectroscopy provides a novel opportunity to examine this disorder.

Whilst the heterogeneity of CVID makes it a suitable candidate for biospectroscopy, investigating this disorder as the study population also has limitations. The limiting factors were considered throughout the conception and design of the study, and where possible, steps were taken to address them. The challenges identified were: the rarity of CVID, whilst relatively common in the context of PIDs, CVID is still classified as rare disorder (affecting between 1:10,000 and 1:100,000 of the population[27,35]); the impact of treatment (replacement immunoglobulin) on the spectra of blood samples; the wide variety and shared clinical features with other pathologies (predominantly secondary immune deficiency but also gastrointestinal disorders, respiratory disorders, lymphoproliferative disease, autoimmune disorders, and granulomatous diseases); and the fact that the disorder is not molecularly well-defined, remaining a diagnosis of exclusion. These factors have been critically appraised throughout the project, as variables that could impact the interpretation of any generated results.

The therapeutic treatment given to CVID patients is replacement immunoglobulin (IG), given either intravenously (IV) or subcutaneously (SC). Therapeutic IG is a sterile preparation of immunoglobulins, produced by the fractionation and purification of pooled plasma from thousands of healthy donors. In light of this we acknowledge that CVID patients, as a cohort, face restrictions when undergoing serological assessment, for example any serological tests for microbial or autoimmune assessment that require quantification of antibodies of the IgG class are unreliable[36].

As a live blood product, the final composition will depend on the concentration, and variation of antibodies present in the donor population. Final formulations will also vary depending on the manufacturer and the production process. IgG makes up more than 90% of the final composition of therapeutic IG; the remaining portion includes small amounts of albumin, IgA, traces of other immunoglobulins, cytokines and soluble receptors[37]. It was beyond the scope of this project to determine specific wavenumber variances associated with replacement IG in samples of treated patients. However the potential impact on this study has been considered below and suggestions for further experiments have been included in the next section.

Firstly, we recognised that IVIG and SCIG, despite the formulation differences dependent on manufacturer, are made up of a finite number of constituents, the majority of which being IgG derived from healthy donors. Therefore, given that ATR-FTIR spectroscopy generates a spectrum representative of the complete composition of a sample (proteins, carbohydrates, lipids, DNA) it is unlikely that contribution from therapeutic IG will impact the entire spectrum, but will be almost completely restricted to the protein-associated peaks.

Blood plasma contains a large number of proteins, albumin accounts for approximately 65% of the total protein concentration, with gammaglobulins accounting for approximately 11% (of which IgG makes up 80%)[38,39]. As the target dose of therapeutic IG is similar to normal physiological levels, we can expect that a similar percentage of the protein contribution (11%) will potentially come from IG treatment. Taking that into consideration, we postulate that the remaining 89% of the protein-associated wavenumbers, and the regions associated with carbohydrates, lipids and DNA are unlikely to have absorbance contributions from replacement IG.

In the further consideration of the potential impact of replacement IG, we examined the demographics and immunoglobulin levels in the study participants. There were two CVID patients that were treatment-naïve, the mean serum level of IgG in 19 treated patients was 7.86 g/L (SD 2.37), and in HCs 10.48 g/L (SD 1.72). There were six CVID patients with IgG levels < 6g/l and 15 with levels > 6 g/L; 13 on SCIG and six receiving IVIG. The inclusion of the treatment-naïve patients into the study adds strength to the notion that the spectral differences resulting in the successful classification of CVID patients from HCs is independent of replacement IG therapy.

A further aim of the study was to extract spectral biomarkers contributing to the main differences in spectra from CVID patients and HCs. We identified 28 potential CVID biomarkers in wavenumbers associated with nucleic acids, collagen, lipids and carbohydrates, but only six wavenumbers associated with protein molecules (see Chapter 3 and Appendix G). This suggests that the presence of replacement IG in serum and plasma (likely associated with protein wavenumber variances) was not a major contributing factor to the discrimination of CVID spectra from HCs.

This research project was undertaken on a study population recruited from the clinical immunodeficiency service provided by Lancashire Teaching hospitals NHS Trust (LTHTR). This service is adult-based, and includes the management of approximately 60 clinically diagnosed CVID patients. Based on our local population, recruiting from this group of PID patients would provide the greatest opportunity to obtain an adequate numbers of participants to ensure sufficient statistical power during analysis.

In summary, CVID was chosen as the study cohort as it provided a novel area to apply biospectroscopy. Moreover, if successful, this technology could be developed into a new diagnostic test for a disease in which current laboratory approaches are limited.

4.2.2 ATR-FTIR as the biospectroscopy method

IR spectroscopy has been used commercially as an analytical platform for several decades. Advances in instrumentation and more importantly, in computerised analysis, have recently enabled its more wider-reaching application, including into biomedical and diagnostic research. Despite this, the majority of biospectroscopy research studies are undertaken within the academic setting, as the highly advanced, technical instrumentation required is better suited to research laboratories. To overcome these

challenges and undertake this study in the clinical setting, we built upon an existing collaborative partnership between our Immunology department at LTHTR and the University of Central Lancashire (UCLAN). This was a major strength for the success of the project. As a specialist technique, not available within any NHS pathology laboratory in the UK, it would not have been possible to undertake this biospectroscopy-based research without this clinical-academic partnership.

This project has further built upon the existing active research interaction with UCLAN. The data presented in this thesis shows the successful initial application of ATR-FTIR in the diagnostic setting of CVID, and has opened up enormous opportunities for continued biospectroscopy research between UCLAN and clinical immunology at LTHTR [40,41]. The expansion of this research group and ongoing application of biospectroscopy methods in immunology studies will be essential; not only for helping healthcare scientists and clinicians to gain an increased awareness of this powerful diagnostic platform, but towards contributing to the understanding of disease-related mechanisms underpinning complex immunological diseases.

The specific analytical method choice was guided by the experience and expertise of the biospectroscopy research group at UCLAN. The research group have published widely in the field of biospectroscopy (over 200 peer-reviewed articles and four book chapters), specifically developing Nature Protocol Tutorials for using FTIR and Raman spectroscopy to analyse biological materials and for the application of multivariate analysis to vibrational spectroscopy of biological samples [19,20,42,43].

As specified in Section 1.1, both FTIR and Raman spectroscopy can be used to examine biological samples, and as they are based on fundamentally different principles (change in a dipole moment compared to change in polarizability, respectively) they are often used as complimentary techniques. In our application for ethical approval, we included the use of both methodologies under the umbrella term of ‘vibrational spectroscopy’. However, due to the length of time allocated to acquiring competence the new methodology, understanding the principles for computational data analysis, and on the collection, processing and analysis of the serum and plasma samples, it was only possible to apply one of the techniques within the time-frame of this research project.

FTIR spectroscopy was selected due to its broader range of application than Raman spectroscopy; almost all organic molecules have a dipole moment and can become excited to undergo vibrations after absorption of IR light [44–46]. Of the three sampling modes of FTIR spectroscopy (transmission, transflection and Attenuated total reflection (ATR)), ATR was selected as it is considered the most suitable for biological samples [47,48]. Further work using Raman spectroscopy should be undertaken to provide further insight into the diagnostic application of vibrational spectroscopy, and will be discussed in the final section on suggestions for future work.

4.2.3 Sample collection and processing

Vibrational spectroscopy is a powerful tool capable of classifying samples into discrete groups based on subtle differences; however this also has limitations, being highly sensitive to the introduction of bias into the analysis process. For this reason, sample collection, storage and processing onto slides prior to FTIR analysis must be carefully controlled and standardised as far as practicable. To manage this, I attended the adult immunodeficiency clinic led by Dr Vijayadurai, Consultant Immunologist at LTHTR to recruit patients and transport study samples immediately to the laboratory where they were centrifuged, aliquoted and frozen at -80 °C until required.

As CVID is a rare disease, in order to maximise patient recruitment, letters of invitation (Appendix D) were sent out to eligible patients two weeks prior to clinic appointment, along with the patient information sheets and consent forms. Prior to opening the study, I attended the Introduction to Good Clinical Practice training course held at Royal Liverpool Hospital, where I gained the requisite skills and competency to obtain patient consent and recruit study participants.

From the total CVID cohort under the care of Clinical Immunology at LTHTR, we had anticipated to reach our target of 20 participants in six months. The recruitment target was actually met over a nine month period, with a total of 21 CVID patients recruited into the study. At the six month time point we had 18 patients and 30 HCs; the decision was made to keep recruitment open to increase the statistical power of any key findings.

In order to gain access to a larger cohort of CVID patients for future larger-scale validation studies; a multi-site recruitment protocol would be required and additional ethical approval obtained. We would approach other immunology centres through The European

Society for Immunodeficiencies (ESID) and UK Primary Immunodeficiency Network (UKPIN). Based on our current working relationships with UK centres, we would initially include Liverpool, Manchester, Cambridge and Oxford in future studies. For the size and time-scales associated within this proof-of-concept study, a single-site based protocol was deemed appropriate.

In the design of the study protocol, blood serum and plasma samples were chosen for ATR-FTIR analysis. These biofluids were chosen for a number of reasons; i) prior experience within the group, ii) suitability for batch analysis, and iii) ease of sample collection, transport and storage in the clinical setting. The significant experience of analysing these sample types within the group provided a template protocol to aid standardisation of sample collection and slide preparation prior to analysis.

The acquisition of sample spectra is a manual process most suited to small batch analysis, with each sample taking approximately one hour to analyse (two blood spot replicates, 10 spectra acquired per replicate). As patient recruitment was undertaken over nine months, samples were required to be aliquoted and frozen prior to analysis. This was vital to prevent the introduction of bias from variable aging samples stored at 2-8 °.

One of the reasons for using blood serum and plasma as opposed to whole blood peripheral blood mononuclear cells (PBMCs) was to avoid the introduction of any further variability from additional preparation processes. In the design of this study it would not have been possible to collect and process samples with 72 hours as is usually recommended for fresh peripheral blood, therefore Ficoll-Paque separation and freezing of PBMCs with subsequent thawing would have been required prior to spectra analysis.

Further, in our aim to establish ATR-FTIR as a method to analyse patient samples and develop vibrational spectroscopy into a routine laboratory test, the successful application of using ATR-FTIR on serum or plasma samples would provide a more accessible approach for the clinical setting and laboratory analysis.

The sample processing and spectral analysis was conducted over a six month time period. In total 102 samples were tested in duplicate (21 CVID patients; 30 HCs, for both serum and plasma samples). ATR-FTIR spectral acquisition was undertaken within the School of Pharmacy and Biomedical Science laboratories at UCLAN. The spectra were obtained

using a Tensor 27 FTIR spectrometer with Helios ATR attachment (Bruker Optics Ltd) operated by OPUS 5.5 software.

As previously stated, FTIR spectrophotometers designed for research laboratories are complex, state-of-the art pieces of equipment requiring specialist expertise for data collection and analysis; spectroscopy training and guidance was provided by Professor Frank Martin and post-doctoral scientists within the research group. Prior to commencing analysis of research project samples, I obtained practical experience using 'training' samples, preparing slides and acquiring spectra to ensure I had full competency to use the equipment unsupervised.

Research samples were thawed and deposited in duplicate onto IR-reflective glass slides (Low-E), air-dried on the bench and placed in a desiccator overnight. All slides were analysed the subsequent day to ensure standardisation of the protocol. Further steps to prevent bias included having one researcher performing all spectral acquisition, and the generation of random study numbers for each participant; this allowed all samples (CVID and HCs) to be analysed 'blind', again to minimise the introduction of bias at any point of the data collection process.

The acquisition of an FTIR spectrum involves collecting a 'single-beam' spectral measurement at one point within a sample. To minimize any bias associated with variance in sample thickness and sample heterogeneity, 10 spectra were collected from different point locations within the dried blood spot for each study subject. As recommended by the literature, point spectra were not taken from the peripheral edges of the blood spot to avoid interference from the well-described 'coffee ring effect' [49–51].

A total of 2040 spectra were acquired; 20 spectra per biofluid per participant, which equated to over one hundred hours of spectral analysis. Whilst steps were taken to maintain a robust methodological approach, due to the involvement of a series of manual techniques, the sample processing stage remains a potential source of error within this research project. Once samples were thawed and mixed, 50 µL were manually pipetted onto the glass slides, producing an approximately 10 mm diameter blood spot (see Chapter 2, section '2.2.2 Sample Collection and preparation' for further detail and an example image of a dried blood spot slide). Although I was the sole researcher preparing the slides, there would have been variance between the size and thickness of the blood

spots. Taking 10 replicates for blood spot and the subsequent application of pre-processing techniques (described in Section 2) were considered adequate steps to control this potential source of error. Additional steps to control the sample thickness would be to introduce an automated platform for sample preparation, as reported by Ollesch et al.[52] where the authors reported higher reproducibility of acquired spectra compared to the non-automated method.

In an alternative approach, Sala et al.[53] recently evaluated the technique of 'digital drying'; collecting spectra from 'wet' samples applied directly to the diamond internal reflection element and computationally subtracting the spectral contribution from water. The group demonstrated increased classification performance of spectra collected from digitally dried samples compared to air-dried samples, proposing that this could provide a promising technique to aid clinical translation of ATR-FTIR spectroscopy. This approach would eliminate the need to prepare slides, not only saving significant amounts of time but would also improve the standardisation of spectral acquisition.

The length of time samples spent drying may have also attributed to potential error. It would take approximately one hour to manually acquire 20 point spectra per sample; the maximum number of samples processed in a single practical session was eight samples. This would result in the final slide having an additional seven hours drying time compared to the first slide; introducing a potential source of variance in the dataset. As highlighted above, pre-processing techniques are applied to aid the elimination of any variances or 'noise' attributed to environmental factors or sampling variance.

Additional factors considered to limit the impact was the inclusion of sufficient participants, undertaking analysis over several different days, and implementing a process for blind sample preparation and spectral acquisition. In this regard it would be unlikely that all HC or all CVID slides would be grouped together and subject to extended drying conditions. Using Raman spectroscopy should be considered as an alternative spectroscopy method to avoid the drying step, as through using an immersion objective or microfluidic device liquid, biofluids can be examined directly[43,54].

4.2.4 Data Analysis and interpretation

As in other specialities, the future direction of immunology investigations will require a wider, 'systems biology' approach. This is now possible through the development of large-

scale analysis methods such as genomics, proteomics and transcriptomics, whereby an astonishing amount of data is generated from a single sample to gain a far more comprehensive view of the immune system. The multidimensional 'big data' generated using these technologies cannot be analysed using traditional or simple univariate analysis methods. Instead, the application of chemometrics, such as multivariate analysis and computational (machine-learning) methods are vital to extract the relevant information for clinical interpretation.

In light of this, interest in bioinformatics and the use of multivariate analysis techniques within the field of immunology has started to rapidly expand[10,11,18,55–58]. This suggests that these methods will become central to future immunology research. As large multidimensional datasets are generated from biospectroscopy, this project provided an opportunity to gain experience and understanding of these advanced analytical platforms. As a non-expert in bioinformatics and chemometrics, the data analysis phase of this project was a challenging task which required expert supervision and guidance to undertake. A substantial amount of time within the project was allocated to learning the fundamental chemometric processes, and how to apply appropriate pre-processing and analysis techniques to complex spectral datasets.

As chemometrics is a speciality in its own right, in order to produce robust data from biospectroscopy studies, it is not recommended for non-experts to process the data independently. The multivariate data analysis undertaken within was performed using the high-level computer programming software, MATLAB (MathWorks). This software, along with the vibrational spectroscopy-focussed computer programming toolbox (IRootLab), developed within the UCLAN research group, were both novel data analysis platforms applied to the investigation of CVID data within this project.

Supported by an expert chemometrician, Dr Camilo L. M. Morais at UCLAN, I successfully evaluated and applied pre-processing (cutting, baseline correction, smoothing, normalisation), unsupervised- (principal component analysis (PCA)) and supervised (PCA-linear discriminant analysis) multivariate analysis techniques within this project (Chapter 2). The programming associated with the final support vector machine (SVM) classification models and feature extraction methods was undertaken by Dr Morais; however in order to interpret the results and perform statistical analysis on the identified

biomarkers within Chapter 3, I gained necessary experience of performing feature extraction techniques and building classification models within MATLAB.

The requirement of complex computational analysis to interpret biospectroscopy data can be considered as both a strength and weakness within this project. The use of chemometrics has been well documented within biospectroscopy studies[19,21], however the complexity of these methods have often deterred non-experts and researchers in smaller centres from applying them to clinical immunology studies. In this regard, a major strength of this project was the collaborative partnership with academia; as within the research group at UCLAN, there was significant expertise in chemometrics, analytical chemistry and biospectroscopy.

Given that we were applying a novel technique (ATR-FTIR spectroscopy) to the investigation of CVID, a methodology that would be largely unfamiliar to Immunologists and life sciences researchers, the support of an expert chemometrician ensured the generation of robust results and added credibility to the project findings.

Importantly, it should be noted that on completion of experimental spectral analysis and construction of classification models, the application of chemometrics would not routinely be required. Instead, the platform could be used as a routine laboratory test with sample spectra put through defined classification algorithms to determine a diagnostic result. As technologies and computational programming continues to advance there is a now a drive towards the development of plug-and-play and point-of-care devices; this will increase the accessibility of biospectroscopy as an investigative tool and aid the translation of biospectroscopy into clinical use[59–61,61].

4.3 Contribution to theory and clinical practice (including suggestions for further work)

The work presented in this thesis demonstrates the successful proof-of concept of applying ATR-FTIR spectroscopy to the study of CVID. Given that this is the first time that this technology and data analysis approach has been applied to an immune deficiency disorder, the scope for future work is extensive. We, as a research group at LTHTR Immunology department, continue our interest in the application of vibrational spectroscopy to investigate immunological disease.

Suggestions for future work relating to this project are outlined below. In addition, we have commenced further research applying biospectroscopy as one of several methods used to investigate COVID-19; 'EXCOVIR' (Exploring COVID-19 specific immune responses in acute and convalescent phases of infection) opened in January 2020, a summary of the project outline has been included in Appendix H.

Clinically, and from a laboratory results perspective, secondary immune deficiency presents with many of the same features of CVID, i.e. low serum immunoglobulin levels and recurrent infection. With adults being the major population affected; this poses a diagnostic dilemma to clinicians, particularly in primary care where the awareness of primary immune deficiency in adults is lower.

Currently, the cost to determine immune deficiency is approximately £500 per patient; this current approach is hindered by the lack of specific diagnostic biomarkers for CVID, and the requirement to exclude all secondary causes before a diagnosis can be reached. With ever-increasing incidence of secondary causes of immune deficiency (due to haematological malignancy, therapeutic drugs, protein-losing disorders, transplant, infection, age)[62] the financial burden on the NHS will continue to rise. An improved approach is therefore required.

Establishing vibrational spectroscopy as novel laboratory platform could provide this improvement and contribute to clinical practice. Further validation studies are required to fully assess potential to accurately identify CVID patients and diagnose patients at an earlier time point; this would have a significant beneficial impact on patient outcome, prompting earlier initiation of treatment and minimising any disease-associated morbidity related to diagnostic delays.

A major aim of this project was to establish the methodology and data analysis techniques required to use biospectroscopy within the clinical immunology laboratory. The data presented in this thesis shows that patients with CVID can be correctly classified from HCs based on the spectral fingerprint of serum and plasma samples. For the serum data, correct classification of HCs and CVID patients demonstrated success rates of 99% and 92%, respectively; and for plasma, 96% and 92% respectively.

Following successful classification of participants, further investigations to identify CVID-associated biomarkers were undertaken, alongside the examination of CVID patient sub-

groups (infection-only versus additional complications). Following individual subject-level analysis (i.e. taking the average of the 20 spectral replicates per sample), we identified six discriminating wavenumber biomarkers in the serum demonstrating significant absorbance intensity differences between CVID patients and HCs (1115 cm^{-1}, 1034 cm^{-1}, 1528 cm^{-1}, 1759 cm^{-1}, 2932 cm^{-1} and 2862 cm^{-1}, of which one biomarker (1528 cm^{-1}) was also capable of discriminating between CVID patient subgroups. In the plasma, nine biomarkers were identified (984 cm^{-1}, 1007 cm^{-1}, 1053 cm^{-1}, 1084 cm^{-1}, 1107 cm^{-1}, 1119 cm^{-1}, 1416 cm^{-1}, 1566 cm^{-1}, and 1639 cm^{-1}), however, none of which were significantly different between the CVID patient subgroups.

Classification models built towards the discrimination of HCs, from CVID patients-without complications, and from CVID patients-with complications demonstrated mixed results dependant on the choice of biofluid and region of the spectra investigated. Greatest classification rates were achieved when using the plasma fingerprint spectral data, in which sensitives for correct grouping of HCs, CVID patients-without complication and CVID patients-with complications were 93%, 73%, and 73%, respectively, and specificities were 87%, 93%, 95%, respectively.

The lower classification rate observed for the clinical sub-group analysis would have been influenced by the reduced population numbers; eight CVID-without complications and thirteen CVID-with complications patients. As these numbers would have been reduced further in the splitting of the samples to produce training and test sets, the final classification model statistics were restricted to a much smaller test dataset (made up of two CVID-without complications and six CVID-with complications patients).

Whilst encouraging findings have been put forward in this project for subgrouping patients based on spectral data, further work with the inclusion of a larger CVID patient cohort is required to validate our findings and fully examine the potential of biospectroscopy to subgroup patients and identify those at risk of severe disease. Given the recent publications using ‘omics-based technologies to define immune profiles of CVID patients with immune dysregulation distinctly from CVID patients with infections only[17,18], further biospectroscopy studies should align with the clinical phenotypes described in these reports.

Following on from the success of this biospectroscopy study which demonstrated the discrimination of CVID patients from HCs with sensitivities and specificities of up to 97% and 95%, respectively, the next step would be to undertake a wider clinical study including additional disease control cohorts. Suggested groups to include would be patients with secondary immune deficiency, inflammatory disorders (i.e. chronic respiratory diseases) and a molecularly defined PID associated with hypogammaglobulinemia such as X-linked agammaglobulinemia.

As noted earlier, further experimental work to examine the spectral variances associated with replacement IG are warranted. We propose inclusion of a further patient group also treated with IVIG, such as neurology patients (Guillain-Barre syndrome and chronic inflammatory demyelinating polyneuropathy) or autoimmune inflammatory disease patients, and look for any shared spectral features with CVID patients, not observed in the HC group. Additional experiments would include the recruitment of additional IG treatment-naïve patients, and the ‘spiking’ of HC sera with increasing concentrations of replacement immunoglobulin products to determine any significant spectral changes associated.

In the laboratory investigation of CVID, analysis of lymphocyte surface marker by flow cytometry can contribute to diagnosis and disease severity subgrouping. However the sample preparation and work-up is time consuming, a specialised, highly skilled workforce is required and there is a high cost associated with the equipment and reagents. Raman spectroscopy could provide an alternative vibrational spectroscopy method that could be applied to the analysis of whole cells.

In contrast to flow cytometry, Raman spectroscopy can provide a chemical fingerprint and image of whole cells without fixation, lysis or use of labels which may disrupt the structure of the cells [63–66]. As differentiated cells produce specific markers and chemicals throughout development, this technique could be used on whole blood samples to define specific spectral features associated with the particular stages of B and T cell differentiation. We would look to apply Raman spectroscopy alongside flow cytometry methods using the EUROclass[67] classification of CVID patients to develop a novel, rapid, label-free and lower-cost assay for the classification of CVID patients.

As the tentative molecular assignments for FTIR and Raman Spectroscopy wavenumber libraries are continually evolving it would be of great interest to collaborate with research groups specialising in the immune profiling of CVID patients. Future work using a multi-omics approach would aid the molecular interpretation of key wavenumbers by aligning CVID-specific protein- or RNA panels with CVID-specific spectral biomarkers.

Whilst the aim of this study was to set out the proof-of-concept for using ATR-FTIR spectroscopy in the setting of CVID, we predict that future development and use of this technology could provide novel insights into aetiological and molecular mechanisms underpinning CVID.

4.4 Final thoughts and conclusion

The application of IR spectroscopy has developed substantially over the past 60 years. In the mid-1940s, early IR spectrophotometers were developed using a diffraction grating to disperse the light; these instruments required absorbance measurements to be taken from each wavelength sequentially in order to produce a complete spectral read-out. Initially, IR spectroscopy was used for the analysis of organic compounds within research and development, particularly in the petrochemical industry. Fourier Transform IR (FTIR) spectrophotometers were later developed in the 1980s, which were capable of measuring all wavelengths simultaneously through the use of the Michelson interferometer. This advancement made it possible to acquire more spectral data in a much shorter time, and increased the analytical sensitivities of the instrument. This had a far-reaching impact on the popularity and application of FTIR spectroscopy, which is now being used across several sectors including chemistry, biology, pharmaceuticals, food and drink, materials science, healthcare, and forensics, among many others.

In particular, ATR-FTIR spectroscopy has gained popularity within the field of biochemical sample analysis, as it permits acquisition of spectral measurements from almost any sample type and provides high resolution data. Further recent instrument developments include the use of higher sensitivity mercury cadmium telluride (MCT) detectors. Compared to the standard Deuterated Lanthanum α Alanine doped TriGlycine Sulphate (DLaTGS) detectors, MCT detectors can make sample measurements at higher speeds without losing any IR response, and show a better signal-to-noise ratio even when the IR signal is low. The use of MCT detectors therefore reduces spectral acquisition times

without losing any sensitivity; further increasing the applicability of vibrational spectroscopy to the clinical setting.

Alongside instrument advances, developments associated with spectral data analysis have increased exponentially in recent years. Computational advances have changed how spectroscopists analyse data; the availability of a wide range of approaches for pre-processing and multivariate analysis (chemometrics) has enabled spectroscopists to uncover subtle but relevant spectral information from large datasets in a timely manner. Some of the chemometric techniques referenced throughout this thesis (LDA, SVM) can be coined 'machine learning' processes. Machine learning is a sub-specialism within the larger field of Artificial Intelligence, and it is revolutionising how large and complex datasets within healthcare science can be analysed. Using statistical algorithms and mathematical modelling, computers can be programmed to learn how to analyse data without human intervention. These tools can uncover new information that would have previously been overlooked using conventional approaches. Used in the healthcare setting, machine learning techniques can improve the efficiency of diagnostic testing and precision medicine, resulting in a better quality of patient care.

The limitations associated with using machine learning in healthcare have been recently reviewed[68]. These include potential issues with data quality, clinical governance, and the introduction of bias. Steps are being taken to address these concerns, and ensure that the expansion and use of analytical tools with artificial intelligence in healthcare are appropriate and ethical.

Within vibrational spectroscopy, machine learning techniques have been successfully applied to diagnostic studies across numerous pathologies. A practical and detailed review has been recently published by Rehman et al. 2021[69], with the aim to bridge the knowledge gap between spectroscopists (who understand chemometrics) and data scientists (who have an advanced knowledge of machine learning). Of note, some of the machine learning techniques currently applied in pilot biospectroscopy studies are not all well-suited to large, continually expanding datasets, which would be anticipated following clinical translation into larger-scale studies. To overcome this, the use of alternative artificial intelligence algorithms such as artificial neural networks (ANNs) has been highlighted. When used as supervised machine learning methods in vibrational spectroscopy, ANNs have demonstrated improved accuracy in a prostate cancer study

compared to PCA-LDA, PCA-linear regression and SVM[70]. Further to this, biospectroscopy research also using ANNs has demonstrated predictive accuracies of 75%-90% for the identification of different bacterial species of Burkholderia[71]. These latest developments within spectroscopy and machine learning could play an integral role in the translation of vibrational spectroscopy into the clinical setting.

The advantages offered by biospectroscopy compared to traditional investigations include minimal sample preparation, rapid sample processing, the use of small sample volumes and the non-destructive nature of the tests, which allows for further investigations to be performed if required[54]. The weaknesses of this method largely relate to issues with reproducibility. Due to a lack of published guidelines and best practice documents, there remains significant protocol variability between research groups. To overcome this, the inclusion of internal quality controls, implementation of an external quality assurance programme or sample exchange scheme between laboratories, and the generation of an immunology-focussed reference library will enable researchers to assess the consistency of results.

For clinicians, having vibrational spectroscopy as an additional tool in their armoury of investigations would be ground-breaking, particularly when applied to complex diseases which may have previously required a large number of clinical and laboratory investigations, some of which highly invasive (such as tissue biopsies). Diagnostic delays could be significantly improved, cost-savings could be made, and the patient pathways for a wide repertoire of diseases could be transformed. The increasing number of studies reporting excellent sensitivities and specificities, alongside the low cost of biospectroscopy analysis has already highlighted the potential use of this platform within large-scale screening programs[54,72]. To this end, the question as to why biospectroscopy has not yet been translated into a routine clinical laboratory test remains as pertinent as ever.

Research groups in the field continue to publish extensively on this subject, with a shared aim to support the development of a standardised process and facilitate clinical translation[20,60,61,61,73–77]. The major barriers identified include stakeholder buy-in, current instrument constraints, and the inability to win the hearts and minds of clinical and laboratory users.

In terms of stakeholder buy-in, the number of different participants required is often overlooked. Collaborative working between spectroscopists, researchers, instrument manufacturers, laboratory directors, clinicians, biomedical scientists, intellectual property lawyers and health economists is vital to improve patient care whilst making cost savings for the NHS. As vibrational spectroscopy has yet to reach the healthcare market, it is viewed as a considerable risk for instrument manufacturers to produce new instruments with features that are more suited to a routine clinical laboratory (improved automation and reduced user complexity). Recently however, enterprises between spectroscopy manufacturers and biospectroscopy researchers are forming (for example, Glyconics Limited and ClinSpec Diagnostics Limited) in an attempt to overcome pre-analytical and instrumentation challenges within the clinical setting. Financial planning and health economics modelling is a key prerequisite when considering the introduction of a new technology or laboratory test. The presentation of a detailed and robust business case evidencing the improvements biospectroscopy can have on patient care, alongside substantial cost savings will also have a major influence on the clinical translation of vibrational spectroscopy[60].

Finally, due to the recent appearance of vibrational spectroscopy within clinical and biological studies, the majority of clinicians and laboratory scientists will be unfamiliar with this analytical platform. If the clinical translation of vibrational spectroscopy is to be successful, it is vital to win the hearts and minds of the clinical stakeholders and take appropriate action to improve understanding. To do this continuous communication and education is paramount; this could be achieved through holding user group meetings, one-to-one training sessions, educational seminars, participation in sample share schemes and providing user group-specific information leaflets. We predict that in the future vibrational spectroscopy data will be used and understood in an equal manner to conventional laboratory tests. For successful clinical translation we must avoid biospectroscopy falling into the ‘black box’ laboratory testing notion that would deter clinicians from using the service[78].

In summary, vibrational spectroscopy has made huge strides in becoming an analytical technique that can be translated into the healthcare setting. In line with this, and the increasing interest and application to immunology research, we envisage that

biospectroscopy methods will become a routine platform in NHS laboratories in the near future.

Within this research project, the significant variation in spectral fingerprint observed between CVID patients and HCs provide an encouraging foundation for the development of biospectroscopy into a novel diagnostic approach for CVID. Characteristic spectral features and wavenumber biomarkers have been identified, the absorbance intensity of which have demonstrated significant differences within the CVID group compared to HCs. These biomarkers may provide a future tool to monitor CVID patients over time and to identify those at risk of disease progression.

The results reported here would now require clinical validation in a larger-scale study, and further experimental work has been suggested within this appraisal chapter. The work presented throughout this thesis has accomplished the first phase of a larger aspiration, to translate the novel methodology, ATR-FTIR spectroscopy, into a diagnostic platform for CVID.

References

1. Kelly, J. G. *et al.* Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. *J. Proteome Res.* **10**, 1437–1448 (2011).

2. Martin, F. L. *et al.* Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nat. Protoc.* **5**, 1748–1760 (2010).

3. Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).

4. Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* **11**, 17–30 (2010).

5. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A Map of Recent Positive Selection in the Human Genome. *PLOS Biol.* **4**, e72 (2006).

6. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).

7. Ramos, P. S., Shedlock, A. M. & Langefeld, C. D. Genetics of autoimmune diseases: insights from population genetics. *J. Hum. Genet.* **60**, 657–664 (2015).

8. The Immunological Genome Project Consortium *et al.* The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).

9. Chu, X., Zhang, B., Koeken, V. A. C. M., Gupta, M. K. & Li, Y. Multi-Omics Approaches in Immunological Research. *Front. Immunol.* **12**, 2312 (2021).

10. Dhillon, B. K., Smith, M., Baghela, A., Lee, A. H. Y. & Hancock, R. E. W. Systems Biology Approaches to Understanding the Human Immune System. *Front. Immunol.* **11**, 1683 (2020).

11. Martino, D., Ben-Othman, R., Harbeson, D. & Bosco, A. Multiomics and Systems Biology Are Needed to Unravel the Complex Origins of Chronic Disease. *Challenges* **10**, 23 (2019).

12. Shi, X.-J., Wei, Y. & Ji, B. Systems Biology of Gastric Cancer: Perspectives on the Omics-Based Diagnosis and Treatment. *Front. Mol. Biosci.* **7**, 203 (2020).

13. Du, W. & Elemento, O. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene* **34**, 3215–3225 (2015).

14. Korsunsky, I. *et al.* Systems Biology of Cancer: A Challenging Expedition for Clinical and Quantitative Biologists. *Front. Bioeng. Biotechnol.* **2**, 27 (2014).

15. Callery, E. L. & Rowbottom, A. W. Vibrational spectroscopy and multivariate analysis techniques in the clinical immunology laboratory: a review of current applications and requirements for diagnostic use. *Appl. Spectrosc. Rev.* 1–30 (2021) doi:10.1080/05704928.2021.1958337.

16. Jiang, F. *et al.* Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.* **2**, 230–243 (2017).

17. Hultberg, J., Ernerudh, J., Larsson, M., Nilsdotter-Augustinsson, Å. & Nyström, S. Plasma protein profiling reflects TH1-driven immune dysregulation in common variable immunodeficiency. *J. Allergy Clin. Immunol.* **146**, 417–428 (2020).

18. Berbers, R.-M. *et al.* Targeted Proteomics Reveals Inflammatory Pathways that Classify Immune Dysregulation in Common Variable Immunodeficiency. *J. Clin. Immunol.* **41**, 1–12 (2021).

19. Morais, C. L. M., Lima, K. M. G., Singh, M. & Martin, F. L. Tutorial: multivariate classification for vibrational spectroscopy in biological samples. *Nat. Protoc.* **15**, 2143–2162 (2020).

20. Baker, M. J. *et al.* Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **9**, 1771–1791 (2014).

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

22. Martin, F. L. *et al.* Identifying Variables Responsible for Clustering in Discriminant Analysis of Data from Infrared Microspectroscopy of a Biological Sample. *J. Comput. Biol.* **14**, 1176–1184 (2007).

23. Paraskevaidi, M. *et al.* Differential diagnosis of Alzheimer's disease using spectrochemical analysis of blood. *Proc. Natl. Acad. Sci.* **114**, E7929–E7938 (2017).

24. Cui, L., Butler, H. J., Martin-Hirsch, P. L. & Martin, F. L. Aluminium foil as a potential substrate for ATR-FTIR, transflection FTIR or Raman spectrochemical analysis of biological specimens. *Anal. Methods* **8**, 481–487 (2016).

25. Butler, H. J., McAinsh, M. R., Adams, S. & Martin, F. L. Application of vibrational spectroscopy techniques to non-destructively monitor plant health and development. *Anal. Methods* **7**, 4059–4070 (2015).

26. Boyle, J. M. & Buckley, R. H. Population prevalence of diagnosed primary immunodeficiency diseases in the United States. *J. Clin. Immunol.* **27**, 497–502 (2007).

27. Chapel, H. & Cunningham-Rundles, C. Update in understanding Common Variable Immunodeficiency Disorders (CVIDs) and the management of patients with these conditions. *Br. J. Haematol.* **145**, 709–727 (2009).

28. Bonilla, F. A. *et al.* International Consensus Document (ICON): Common Variable Immunodeficiency Disorders. *J. Allergy Clin. Immunol. Pract.* **4**, 38–59 (2016).

29. Seidel, M. G. *et al.* The European Society for Immunodeficiencies (ESID) Registry Working Definitions for the Clinical Diagnosis of Inborn Errors of Immunity. *J. Allergy Clin. Immunol. Pract.* **7**, 1763–1770 (2019).

30. Ameratunga, R., Woon, S.-T., Gillis, D., Koopmans, W. & Steele, R. New diagnostic criteria for common variable immune deficiency (CVID), which may assist with decisions to treat with intravenous or subcutaneous immunoglobulin. *Clin. Exp. Immunol.* **174**, 203–211 (2013).

31. Resnick, E. S., Moshier, E. L., Godbold, J. H. & Cunningham-Rundles, C. Morbidity and mortality in common variable immune deficiency over 4 decades. *Blood* **119**, 1650–1657 (2012).

32. Chapel, H. *et al.* Common variable immunodeficiency disorders: division into distinct clinical phenotypes. *Blood* **112**, 277–286 (2008).

33. Cunningham-Rundles, C. The many faces of common variable immunodeficiency. *Hematol. Am. Soc. Hematol. Educ. Program* **2012**, 301–305 (2012).

34. Butler, H. J. *et al.* Optimised spectral pre-processing for discrimination of biofluids via ATR-FTIR spectroscopy. *Analyst* **143**, 6121–6134 (2018).

35. Cunningham-Rundles, C. & Bodian, C. Common Variable Immunodeficiency: Clinical and Immunological Features of 248 Patients. *Clin. Immunol.* **92**, 34–48 (1999).

36. Więsik-Szewczyk, E. & Jahnz-Różyk, K. From infections to autoimmunity: Diagnostic challenges in common variable immunodeficiency. *World J. Clin. Cases* **8**, 3942–3955 (2020).

37. Barahona Afonso, A. F. & João, C. M. P. The Production Processes and Biological Effects of Intravenous Immunoglobulin. *Biomolecules* **6**, 15 (2016).

38. Larsson, A. & Hansson, L.-O. Plasma Protein Fractions in Healthy Blood Donors Quantitated by an Automated Multicapillary Electrophoresis System. *J. Chromatogr. Sci.* **44**, 479–483 (2006).

39. Stoop, J. W., Zegers, B. J. M., Sander, P. C. & Ballieux, R. E. Serum immunoglobulin levels in healthy children and adults. *Clin. Exp. Immunol.* **4**, 101–112 (1969).

40. Morris, A. D. *et al.* A comparative analysis of different biofluids using Raman spectroscopy to determine disease activity in ANCA-associated vasculitis. *J. Biophotonics* **14**, e202000426 (2021).

41. Morris, A. D. *et al.* Distinguishing active from quiescent disease in ANCA-associated vasculitis using attenuated total reflection Fourier-transform infrared spectroscopy. *Sci. Rep.* **11**, 9981 (2021).

42. Martin, F. L. *et al.* Distinguishing cell types or populations based on the computational analysis of their infrared spectra. *Nat. Protoc.* **5**, 1748–1760 (2010).

43. Butler, H. J. *et al.* Using Raman spectroscopy to characterize biological materials. *Nat Protoc.* **11**, 664–687 (2016).

44. Bunaciu, A. A., Aboul-Enein, H. Y. & Fleschin, Ş. Vibrational Spectroscopy in Clinical Analysis. *Appl. Spectrosc. Rev.* **50**, 176–191 (2015).

45. Quaroni, L. & Zlateva, T. Infrared spectromicroscopy of biochemistry in functional single cells. *The Analyst* **136**, 3219 (2011).

46. Rehman, I., Movasaghi, Z. & Rehman, S. *Vibrational Spectroscopy for Tissue Analysis*. (CRC Press, 2012).

47. Shai, Y. ATR-FTIR studies in pore forming and membrane induced fusion peptides. *Biochim. Biophys. Acta* **1828**, 2306–2313 (2013).

48. Tatulian, S. A. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy: A Method of Choice for Studying Membrane Proteins and Lipids. *Biochemistry* **42**, 11898–11907 (2003).

49. Choi, S. & Birarda, G. Protein Mixture Segregation at Coffee-Ring: Real-Time Imaging of Protein Ring Precipitation by FTIR Spectromicroscopy. *J. Phys. Chem. B* **121**, 7359–7365 (2017).

50. Yunker, P. J., Still, T., Lohr, M. A. & Yodh, A. G. Suppression of the coffee-ring effect by shape-dependent capillary interactions. *Nature* **476**, 308–311 (2011).

51. Deegan, R. D. *et al.* Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389**, 827–829 (1997).

52. Ollesch, J. *et al.* FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification. *Analyst* **138**, 4092–4102 (2013).

53. Sala, A. *et al.* Rapid analysis of disease state in liquid human serum combining infrared spectroscopy and "digital drying". *J. Biophotonics* **13**, e202000118 (2020).

54. Mitchell, A. L., Gajjar, K. B., Theophilou, G., Martin, F. L. & Martin-Hirsch, P. L. Vibrational spectroscopy of biofluids for disease screening or diagnosis: translation from the laboratory to a clinical setting. *J. Biophotonics* **7**, 153–165 (2014).

55. Pertseva, M., Gao, B., Neumeier, D., Yermanos, A. & Reddy, S. T. Applications of Machine and Deep Learning in Adaptive Immunity. *Annu. Rev. Chem. Biomol. Eng.* **12**, 39–62 (2021).

56. Stafford, I. S. *et al.* A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *Npj Digit. Med.* **3**, 1–11 (2020).

57. Saeys, Y., Van Gassen, S. & Lambrecht, B. N. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nat. Rev. Immunol.* **16**, 449–462 (2016).

58. Pappalardo, F., Russo, G. & Reche, P. A. Toward computational modelling on immune system function. *BMC Bioinformatics* **21**, 546 (2020).

59. Shaw, R. A. & Mantsch, H. H. Infrared Spectroscopy of Biological Fluids in Clinical and Diagnostic Analysis. in *Encyclopedia of Analytical Chemistry* (John Wiley & Sons, Ltd, 2006).

60. Finlayson, D., Rinaldi, C. & Baker, M. J. Is Infrared Spectroscopy Ready for the Clinic? *Anal. Chem.* **91**, 12117–12128 (2019).

61. Cameron, J. M. *et al.* Exploring pre-analytical factors for the optimisation of serum diagnostics: Progressing the clinical utility of ATR-FTIR spectroscopy. *Vib. Spectrosc.* **109**, 103092 (2020).

62. Patel, S. Y., Carbone, J. & Jolles, S. The Expanding Field of Secondary Antibody Deficiency: Causes, Diagnosis, and Management. *Front. Immunol.* **10**, 33 (2019).

63. Ramoji, A. *et al.* Raman Spectroscopy Follows Time-Dependent Changes in T Lymphocytes Isolated from Spleen of Endotoxemic Mice. *ImmunoHorizons* **3**, 45–60 (2019).

64. Hobro, A. J., Kumagai, Y., Akira, S. & Smith, N. I. Raman spectroscopy as a tool for label-free lymphocyte cell line discrimination. *Analyst* **141**, 3756–3764 (2016).

65. Managò, S., Mirabelli, P., Napolitano, M., Zito, G. & Luca, A. C. D. Raman detection and identification of normal and leukemic hematopoietic cells. *J. Biophotonics* **11**, e201700265 (2018).

66. Ichimura, T. *et al.* Non-label immune cell state prediction using Raman spectroscopy. *Sci. Rep.* **6**, 37562 (2016).

67. Wehr, C. *et al.* The EUROclass trial: defining subgroups in common variable immunodeficiency. *Blood* **111**, 77–85 (2008).

68. Cutillo, C. M. *et al.* Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. *Npj Digit. Med.* **3**, 1–5 (2020).

69. Meza Ramirez, C. A., Greenop, M., Ashton, L. & Rehman, I. ur. Applications of machine learning in spectroscopy. *Appl. Spectrosc. Rev.* **56**, 733–763 (2021).

70. Shao, X. *et al.* Deep convolutional neural networks combine Raman spectral signature of serum for prostate cancer bone metastases screening. *Nanomedicine Nanotechnol. Biol. Med.* **29**, 102245 (2020).

71. Kochan, K. *et al.* Raman spectroscopy as a tool for tracking cyclopropane fatty acids in genetically engineered Saccharomyces cerevisiae. *Analyst* **144**, 901–912 (2019).

72. Freitas, D. L. D. *et al.* Spectrochemical analysis of liquid biopsy harnessed to multivariate analysis towards breast cancer screening. *Sci. Rep.* **10**, 12818 (2020).

73. Kim, J. A., Wales, D. J. & Yang, G.-Z. Optical spectroscopy for in vivo medical diagnosis—a review of the state of the art and future perspectives. *Prog. Biomed. Eng.* **2**, 042001 (2020).

74. Siraj, N. *et al.* Raman spectroscopy and multivariate regression analysis in biomedical research, medical diagnosis, and clinical analysis. *Appl. Spectrosc. Rev.* 1–58 (2021) doi:10.1080/05704928.2021.1913744.

75. Baker, M. J. *et al.* Clinical applications of infrared and Raman spectroscopy: state of play and future challenges. *The Analyst* **143**, 1735–1757 (2018).

76. Lovergne, L. *et al.* Investigating optimum sample preparation for infrared spectroscopic serum diagnostics. *Anal. Methods* **7**, 7140–7149 (2015).

77. Butler, H. J. *et al.* Shining a light on clinical spectroscopy: Translation of diagnostic IR, 2D-IR and Raman spectroscopy towards the clinic. *Clin. Spectrosc.* **1**, 100003 (2019).

78. Bathaee, Y. The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harv. J. Law Technol.* **31**, 50 (2018).

APPENDIX A

HIGHER SPECIALIST SCIENTIFIC TRAINING MODULE C – INNOVATION PROPOSAL

This appendix contains the innovation proposal document. This piece of work is required by all Life Sciences Higher Specialist Trainees in Cohort 3 and onwards to be included as an appendix in the thesis. Assessment of the innovation proposal has been undertaken by the author's workplace and academic supervisors.

Innovation Proposal - Pass/~~Fail~~

Establishing a new diagnostic pathway for immune deficiency disorders in the Clinical Immunology Laboratory

Executive Summary

Immunological disorders are complex, often demonstrating interaction across multiple molecular pathways which results in delayed diagnosis. The frequency of immune deficiency is rising, especially in older adults; the number of patients with secondary immunodeficiencies (SIDs) greatly exceeds those with primary immunodeficiencies (PIDs); however the prevalence of both appears to be on the rise probably because of increased access to laboratory investigations, greater use of immunosuppressive drugs and a growing aging population. Vibrational spectroscopy is being applied in many fields and here we propose an innovation project for its use in clinical immunology, specifically within the investigation of immune deficiency disorders. The most prevalent, symptomatic primary immunodeficiency, common variable immunodeficiency (CVID) is a key area to pilot this innovation project. There is currently no diagnostic test specific for CVID. This leads to patients experiencing an average diagnostic delay of 4-5 years, during which time irreversible airway damage can occur as a result of recurrent infections [1]. Moreover, due to the fact that presentation is mostly in adulthood, it can be difficult to discriminate CVID from secondary immune deficiencies, for which the clinical care pathways largely differ. CVID patients require life-long treatment with replacement intravenous immunoglobulins (IVIG), and close clinical monitoring to recognise and manage further disease-associated complications. If successful, the potential benefits of introducing vibrational spectroscopy into a routine test for immunodeficiency include; its use in disease diagnosis, monitoring and treatment response, an enhanced understanding of both genetic and non-genetic components of CVID, and the prediction of future clinical outcomes. The implementation of vibrational spectroscopy within Lancashire and Lakeland Regional Immunology Service will be the first NHS laboratory to translate this technique into routine practice for investigation of immune deficiency disorders.

Project Background and Rationale

Immune deficiency disorders are caused by a defect in the immune system which results in the absence or reduced function of vital immune components needed to fight infection. These disorders can be divided into genetically inherited primary immunodeficiency disorders (PIDs) and secondary immunodeficiency disorders, which are acquired as a result of disease or environmental factors. The diagnosis and monitoring of immunodeficiency largely involves investigation of lymphocytes (T- B- and NK-cells) and numerous proteins associated with these cells. A wide range of laboratory methods are routinely used to investigate suspected immunodeficiency including multi-colour flow cytometry, enzyme immunoassays, cell

proliferation, and nephelometric assays. The results obtained from these tests can provide an indication of the immunological defect leading to a compromised immune system. However, each method has its limitations, and consequently information obtained at the macromolecular level is restricted. An enhanced and more detailed analysis of the cells involved in the immune response would develop current understanding and add to the discovery of defective components or signalling pathways implicated in immunodeficiency.
In the 2019 update to the International Union of Immunological Societies (IUIS) classification of Inborn Errors of Immunity report, 430 distinct PIDs were characterised, ranging in prevalence from very rare to relatively common [2]. Common Variable Immune Deficiency (CVID) is the most prevalent symptomatic primary immunodeficiency disorders (PIDs)[3], with 90% of patients suffering from recurrent infections due to the failure to produce protective antibodies (called immunoglobulins). There is also an increased risk of autoimmune disorders (22% of patients) and malignancy (16% of patients) because of underlying immune dysregulation [4–7]. In light of its heterogeneous clinical presentation and lack of disease-specific biomarker, diagnosis remains one of exclusion. Current diagnostics rely heavily on the measurement of immunoglobulins alongside clinical symptoms, with both International CONsensus (ICON) and European Society for Immunodeficiencies (ESID) criteria requiring patients to have low serological levels reported (hypogammaglobulinemia)[4,8]. The caveat to this finding is the requirement to exclude all other causes of low immunoglobulins, for which there are a wide variety of causes[9], the most prevalent of which are secondary, acquired due to multiple causes including HIV, malignancy, therapeutics, malnutrition and protein loss. Due to its low clinical specificity, the use of immunoglobulins as a diagnostic biomarker for CVID is limited, highlighting the need for an improved diagnostic approach.

New discoveries and advances in technologies mean that we can explore new laboratory approaches for the investigation of complex disorders. Methods designed for 'omics-like' analysis, such as genome, transcriptome and proteome investigations are often a challenge to bring into routine laboratory use due to the high cost burden associated with these analytical platforms. There is an unmet need for a low cost and simple to process analytical platform that can be routinely accessible within the financial constraints of NHS laboratories. One potential and innovative candidate that enables sample analysis at the molecular level is vibrational spectroscopy. Vibrational spectroscopy is an umbrella term to describe the techniques used to produce a unique spectral read-out, or molecular 'fingerprint' of a sample following excitation with light[10]. The unique molecular fingerprint of a sample relates to its biomolecular constituents (i.e. proteins, lipids, nucleic acids, carbohydrates) and is generated from the vibrations of the chemical bonds in these molecules. The most important optical techniques are infrared (IR) and

Raman spectroscopy [11]; both of which are well established methods for studying sample types such as biofluids, tissues and cell cultures [10]. As the molecular fingerprint of a sample will change due to the presence of disease; vibrational spectroscopy is a well-placed candidate for the study of pathological processes and development of a novel diagnostic platform [11]. Thorough the use of computational methods and machine learning for spectral data analysis, this innovation will promote the digital revolution and potentially be the first of many to apply machine learning applications to clinical pathways. As discussed in detail within the literature in Chapter 1, the diagnostic potential of vibrational spectroscopy has already been pioneered and proven successful when applied to a wealth of proof-of concept studies. These studies have been undertaken across a wide range of medical disciplines, including disorders associated with immune dysfunction. No NHS laboratory is currently using vibrational spectroscopy within the clinical setting of immunodeficiency.

Project Objectives

The objectives for the innovation project, to embed vibrational spectroscopy into the immunology laboratory testing pathway for CVID are to:

- Further validate and assess clinical utility of individual CVID-associated wavenumber biomarkers identified in proof-of-concept study[12]. These included wavenumbers in regions indicative of nucleic acids (984 cm^{-1}, 1053 cm^{-1}, 1084 cm^{-1}, 1115 cm^{-1}, 1528 cm^{-1}, 1639 cm^{-1}), and a collagen-associated biomarker (1528 cm^{-1}) in serum and plasma samples
- Examine the proof-of concept for investigating lymphocytes with Raman spectroscopy and benchmark against existing conventional methods (flow cytometry) for immunodeficiency diagnosis within a clinical setting
- Improve diagnostic sensitivity and specificity for CVID by potentially eliminating the need to exclude all other causes of hypogammaglobulinemia prior to classification of patients as CVID
- Reduce diagnostic delay for CVID patients

Feasibility

Option 1: Do nothing

Immunodeficiency is one of the major immunological disorders investigated within the NHS clinical immunology service, investigations to determine a diagnosis costs approximately £500 per patient; there remain questions about the specificity of this conventional approach and it is a significant burden on existing resources. For this option, measurement of serum immunoglobulins and response to test vaccinations will remain the mainstay of laboratory testing for CVID. There would be no further research studies undertaken using vibrational spectroscopy methods, and there would be no additional equipment or staff training required. As previously stated, the

findings from these tests will not differentiate between primary and secondary causes of hypogammaglobulinemia. Patients presenting with symptoms suggestive of immune deficiency without an obvious secondary cause for subnormal immunoglobulins levels, will continue to be referred to secondary care services. Further investigations will then be undertaken on fresh whole blood samples; these will include flow cytometry-based testing for the enumeration of lymphocyte subsets and assessment of B-lymphocyte differentiation (presence of class-switched memory B lymphocytes). There would be no additional costs associated with this option however hypogammaglobulinemia patients would require collection of additional blood samples and outpatient clinic appointments prior to specialised investigations being undertaken. Patient care would remain unchanged but any potential cost savings through reduced hospital visits, unnecessary clinic appointments and use of phlebotomy services would be difficult. The risk for this option is low; however there would be minimal scope to improve patient care, particularly in relation to improving diagnostic delays and reducing the financial pressures associated with investigation of immunodeficiency patients.

Option 2: Introduce vibrational spectroscopy as an adjunct to current laboratory investigation for CVID

Following on from the success of our proof-of-concept study which demonstrated the ability of ATR-FTIR spectroscopy to discriminate CVID patients from healthy controls with sensitivities and specificities of up to 97% and 95%, respectively, we propose a novel laboratory diagnostic approach to CVID, which embeds vibrational spectroscopy techniques into the clinical immunology laboratory for the first time. Our main objectives within this service improvement are to use both Raman and ATR-FTIR spectroscopy techniques alongside current laboratory assays for CVID (analysis of serum immunoglobulins and B lymphocyte cell surface markers) in order to increase the diagnostic sensitivity and specificity of the current diagnostic testing pathway. This strategy will aim to reduce the diagnostic delay for CVID patients and thus progression of irreversible lung and other complications. Moreover, as CVID is considered to be a group of disorders with a spectrum severity, we aim to use disease-specific wavenumbers as biomarkers in order to assess patient risk for developing severe disease, enabling prompt and more aggressive treatment if required. Finally, with improved diagnostic specificity, it will also aid in preventing erroneous categorisation of secondary immunodeficiency patients as CVID, which would ensure that patients receive the correct clinical management, and that expensive blood-derived products such as IVIG are used appropriately.

Using Raman spectroscopy to identify individual populations of lymphocytes has potential to be developed into a new diagnostic method for identifying immune cell disorders. The Raman method could replace current investigations such as flow cytometry, and provide a lower cost

alternative that could be offered to laboratories and clinical services across the UK. Raman spectroscopy is an emerging area which can provide a chemical fingerprint of cells without fixation, lysis or use of labels which may disrupt the structure of the cells. The translation of vibrational spectroscopy into the clinical setting of immunodeficiency has the ability to impact the patient pathway and clinical management of a large volume of patients. Collaboration between partners based within a clinical setting and academia is always challenging; however we have an existing active research interaction in this area with UCLAN which opens up enormous opportunities. We are aware of interest nationally and internationally in developing these technologies. There will be minimal upfront investment required during the initial Clinical Validation phase of the project as spectroscopy-based analysis will be performed at UCLAN, where there is existing expertise in these methodologies. The specialised immunology testing which is currently performed will remain unchanged, using the facilities and expertise already present within the Immunology department of Lancashire and Lakeland Regional Immunology service.

Stakeholder engagement and further support has been offered by Regional spectroscopy network, to which the project has been presented. The network is made up of clinicians (across a wide range of disciplines) at Lancashire Teaching Hospitals NHS Trust (LTHTR), spectroscopy research scientists (Professor, post-doc and PhD level academics) at UCLAN and bioinformaticians at LTHTR and UCLAN. Consultant immunologists are supportive and currently manage a cohort of approximately 60 CVID patients. Possible barriers to implementing this option include regular access to specialised spectroscopy equipment once the test becomes a routine immunology assay, the laboratory would be required to purchase own spectrophotometer which would involve a large capital purchase or managed service contract. Ethical consideration and formal approval would be required to conduct the further research required prior to translating vibrational spectroscopy into a routine test. Gaining the funding to conduct further research would require successful grant application, and finally there would be time and costs associated with training current immunology staff to undertake the testing once the platform is fully validated.

Clinical impact & Patient care

In a typical primary care trust such as Preston, approximately 10,000 cases of suspected immunodeficiency annually are screened at a cost of £200k in a catchment area of 370k population. Rolled out nationally, this is a significant cost burden. The proposal to develop vibrational spectroscopy into a routine immunology test would provide a high throughput technology that is inexpensive with greater potential sensitivity and specificity. If proved to be

effective, vibrational spectroscopy could vastly improve individualised patient care and management. Usually, the only prompt for clinicians suspect an underlying immunodeficiency is that the patient develops recurrent or serious infections and is then investigated for potential immunological defects, by which time inflammatory changes and irreversible organ damage may have already occurred. Establishing a novel laboratory platform with the potential to accurately identify these patients at an earlier time point would have a significant beneficial impact on patient outcome, prompting earlier initiation of treatment and minimising any disease-associated morbidity related to diagnostic delays. The treatment and management of immunodeficiency patients depends on severity and whether the cause is primary- or secondary. Introduction of a new diagnostic test that could accurately discriminate between primary and secondary immune deficiency disorders would ensure patients are entered onto the correct clinical pathway and managed more effectively, alleviating financial pressures and inappropriate use of expensive blood-derived therapeutics such as IVIG.

Conclusion and Next steps

Immunodeficiency disorders are a common immune problem. There is an urgent need for translation of a new methodology that would facilitate high throughput analysis in an inexpensive fashion. Using vibrational spectroscopy has potential to be developed into a new diagnostic method for identifying immunodeficiency disorders. Following the remarkable findings from our proof-of-concept study using FTIR spectroscopy alone to successfully classify CVID patients from healthy controls (with sensitivities and specificities of up to 97% and 95%, respectively), the next steps required to implement this innovation project would involve securing funding for a Raman spectroscopy-based study of lymphocytes in immune deficiency disorders. Subsequently, in the clinical setting, we propose a multi-centre large-scale evaluation study whereby biospectroscopy will be used alongside the current laboratory investigations recommended within the CVID diagnostic criteria (measurement of immunoglobulins, vaccine responses, class-switch) to assess the impact that including this novel method could have on improving diagnostic delays and patient management. We envisage biospectroscopy would add value in the first instance by eliminating the requirement to exclude all other causes of hypogammaglobulinemia. This method could replace current investigations such as flow cytometry, and provide a lower cost alternative that could be offered to laboratories and clinical services across the UK, if deemed superior in performance characteristics (see appendix 1). This piece of work would provide the necessary evidence to streamline the diagnostic investigations for CVID; by proposing new testing algorithms and diagnostic guidelines to improve the future diagnosis and management of patients with CVID. We envisage establishing the first national FTIR- and Raman spectrochemical

diagnostic laboratory within a clinical setting; such an entity would generate employment towards lab managers, data analysts and bio-imaging personnel.

References

1. Gathmann, B. *et al.* Clinical picture and treatment of 2212 patients with common variable immunodeficiency. *J. Allergy Clin. Immunol.* **134**, 116-126.e11 (2014).

2. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee | SpringerLink. https://link.springer.com/article/10.1007/s10875-019-00737-x.

3. Boyle, J. M. & Buckley, R. H. Population prevalence of diagnosed primary immunodeficiency diseases in the United States. *J. Clin. Immunol.* **27**, 497–502 (2007).

4. Bonilla, F. A. *et al.* International Consensus Document (ICON): Common Variable Immunodeficiency Disorders. *J. Allergy Clin. Immunol. Pract.* **4**, 38–59 (2016).

5. Chapel, H. & Cunningham-Rundles, C. Update in understanding Common Variable Immunodeficiency Disorders (CVIDs) and the management of patients with these conditions. *Br. J. Haematol.* **145**, 709–727 (2009).

6. Cunningham-Rundles, C. & Bodian, C. Common Variable Immunodeficiency: Clinical and Immunological Features of 248 Patients. *Clin. Immunol.* **92**, 34–48 (1999).

7. Cunningham-Rundles, C. The many faces of common variable immunodeficiency. *Hematol. Am. Soc. Hematol. Educ. Program* **2012**, 301–305 (2012).

8. Ameratunga, R., Woon, S.-T., Gillis, D., Koopmans, W. & Steele, R. New diagnostic criteria for common variable immune deficiency (CVID), which may assist with decisions to treat with intravenous or subcutaneous immunoglobulin. *Clin. Exp. Immunol.* **174**, 203–211 (2013).

9. Perez, E. E. *et al.* Update on the use of immunoglobulin in human disease: A review of evidence. *J. Allergy Clin. Immunol.* **139**, S1–S46 (2017).

10. Naumann, D. Vibrational Spectroscopy in Microbiology and Medical Diagnostics. in *Biomedical Vibrational Spectroscopy* (eds. Lasch, P. & Kneipp, J.) 1–8 (John Wiley & Sons, Inc., 2008).

11. Bunaciu, A. A., Aboul-Enein, H. Y. & Fleschin, Ş. Vibrational Spectroscopy in Clinical Analysis. *Appl. Spectrosc. Rev.* **50**, 176–191 (2015).

12. Callery, E. L. *et al.* New approach to investigate Common Variable Immunodeficiency patients using spectrochemical analysis of blood. *Sci. Rep.* **9**, 7239 (2019).

Appendix 1

Laboratory analysis	**Basic technique**	**Advantages**	**Limitations**	**Through-put**	**Approx. cost per test (£)**
Serum Immunoglobulins & electrophoresis	Quantitation of IgG, IgA, IgM & total protein by turbidimetry or nephelometry. Electrophoresis separates proteins into 5 groups of similar size, shape, and charge using an electrical field.	Major requirement for diagnosis: decreased IgG fundamental to diagnose CVID (in both ICON and ESID criteria) Highly automated, rapid, cost-effective, long-standing, standardised assay (ERM-DA470k/ IFCC reference material)	Not Diagnostic for CVID in isolation – must exclude all secondary causes of hypogammaglobulinemia	High	10
Lymphocyte subsets (TBNK)	Flow cytometry is used to detect, identify, and count specific cells. Fluorochrome-tagged monoclonal antibodies against cell surface markers enable each subset (T- B- and NK- cells) to be enumerated in whole blood.	Informative, useful to investigate for any quantitative deficiencies in each of the lymphocyte subsets. CVID thought of as B cell defect but T cell lymphopenia noted in approx. 50% patients.	Not required for diagnosis and not diagnostic for CVID in isolation – B cells can be low/normal; T cells variable Manual assay - not fully automated Specialist skilled staff to perform and interpret results	Medium	35
Class-switch +/- Euroclass	Method as above – Fluorochrome-tagged monoclonal antibodies against B-cell specific surface markers	Included in ESID (2014) diagnostic criteria Assessment of CVID clinical phenotype and risk severity based on percentages of class-switched memory B cells, transitional cells, or CD21^{low} cells.	Not Diagnostic for CVID in isolation Not possible to perform in patients with B cell lymphopenia (<50 cells/ul) Not essential for diagnosis Specialist skilled staff to perform and interpret results	Medium	30

Appendix A

Laboratory analysis	**Basic technique**	**Advantages**	**Limitations**	**Through-put**	**Approx. cost per test (£)**
Vaccine responses Specific Microbial antibodies – Tetanus (TET)/Haemophilus influenza b (HIB) and Pneumococcal polysaccharide serotypes (PNS)	Multiplex or ELISA-based test to assess the efficiency of IgG antibody response to active immunisation (or infection) to HIB, TET or PNS in immunocompromised individuals.	Included in ESID (2014) Measurement of specific antibody production (spontaneous and post immunisation) is useful in the assessment of patients with suspected immune (immunoglobulin) deficiency (hereditary or acquired). Can assess T-dependent and T-independent pathways by using pneumococcal conjugated vaccines (PCV) or pneumococcal polysaccharide vaccine (PPV), respectively Automated assay platforms available	Not Diagnostic for CVID in isolation – must exclude other causes of poor vaccine responses. Not standardised, results using different platforms not comparable. A widely misused test – not to be used as general test of vaccination and certainly not as a test of infection. The pneumococcal antibody test (total IgG) is of limited use in testing for specific immunisation status since the new vaccines contain very different serotypes – serotype specific test required. No universal agreement as to what constitutes a 'normal' serotype-specific response to the PPV; AAAAI definition most commonly used.	High	85
Genetic Analysis	Next-generation and targeting sequencing; involves array-based massive parallel sequencing (genomic DNA is fragmented and ligated for library preparation followed by amplification and sequencing) Whole-exome or whole- genome approaches can be used. Panel of 16 genes offered by PreventionGenetics covering the coding regions of targeted genes plus ~10 bases of non-coding DNA flanking each exon	Diagnostic for CVID (genetic defects found in ~20% CVID cases) Commercial production of panels for PID-associated genes have reduced costs and therefore increased accessibility of genetic testing High throughput analysis High sensitivity Performed using nanograms of starting material	Expensive test – not appropriate as a first line investigation in NHS healthcare services Majority of patients have polygenic or multifactorial disease therefore not helpful for diagnosis Complex data - Requires highly specialist scientific staff and bioinformatics experts, Expensive set-up costs and infrastructure required Not useful if gene/variant of interest is not known	Low-Medium (high through-put analysers however complex interpretation increases turnaround time to >14 days)	650

Appendix A

Laboratory analysis	Basic technique	Advantages	Limitations	Through-put	Approx. cost per test (£)
Vibrational spectroscopy	Detect biochemical changes on the basis of spectral features. Changes in energy due to vibration of molecular bonds. These changes reflect the chemical and molecular composition of a sample which can be associated with specific pathological conditions	Potentially diagnostic in isolation - proof of concept study demonstrated sensitivity and specificity >95%; requires verification in larger-scale trials Non-invasive technique, suitable with a wide range of sample types (cells, tissues, biofluids inc. blood, urine, CSF) Rapid, label- and reagent-free methodology, non-destructive therefore allows further sample analysis if required Inexpensive Scope to become used as a point-of- care testing device – in development in commercial companies	Currently largely used as a research technique, limited clinical application – unfamiliar with clinicians and NHS laboratory scientists Lack of large-scale clinical trials to validate initial findings & support translation into the clinical setting Complex data produced – requires specialist staff and bioinformaticians to process and interpret data (chemometrics) Not yet automated or standardised – ongoing work is being undertaken to address this.	Medium (potential to become high throughput with automation of sampling and bioinformatics analysis)	0.60

APPENDIX B

PROJECT PRESENTATION - LAYMAN'S ORAL PRESENTATION FEEDBACK

Name of Trainee: Emma Callery Specialism: Clinical Immunology
Innovation Title: A novel laboratory approach using biospectroscopy methods for the investigation of CVID patients
Assessment criteria:

- Quality and clarity of explanation of the innovation for a lay audience (awareness of the use of jargon, scientific language and acronyms)
- Synthesis of relevant scientific evidence for a lay audience
- Ability to persuade a lay audience of the merits (or otherwise) of the innovation and its potential role in healthcare science services
- Style of presentation (slides, delivery; body language, eye contact, voice, confidence) and appropriateness for a lay audience
- Demonstration of values, attitudes and behaviours expected of a leader in clinical science

Final Result	**Pass**	

Summary comments for feedback:

The presentation was within time limits and delivered in a professional manner.

Emma explained the background to her project clearly in lay terms, using her slides very effectively to illustrate complex scientific information. The novelty of the project was demonstrated, and the clinical need for such novel approaches was explained well. Although the clinical need was addressed, it would have been beneficial to see more examples from the point of view of the CVID patient group, and the anticipated patient journey / impact on patient care. Overall, terminology was well explained but some areas of the presentation, and language used, could have been simplified more for the benefit of a lay audience.

Slides were of excellent quality, well planned, colourful and with a very good balance between text and images. Use of the animation tool within PowerPoint was effective to build up complex information gradually and aid understanding. Without animation, slides would have contained too much text.

Emma's presentation style seemed scripted, and sounded a little unnatural given the small size of the audience. Body language was rather stiff to begin with, but improved as she started to relax and gain confidence, and eye contact was maintained with all of the audience throughout her presentation. Voice projection, intonation and expression were all excellent. Emma's enthusiasm for the project shone through.

There was demonstration of professional values with a primary focus on diagnostic needs, and service provision. Emma answered all of the questions from the panel well; she seemed much more relaxed and fluent when communicating with the audience during the Q&A session.

Overall, the presentation was understandable to the lay representative who commented that Emma explained the science well, and was very confident in the Q&A session.

APPENDIX C

IRAS AND HRA APPROVALS

This appendix contains copies of the Integrated Research Approval System (IRAS) and Health Research Authority (HRA) approvals for this study. Permission was granted 22nd November 2016.

Miss Emma Callery
18 Crocketts Walk
Eccleston
WA10 5DUDr Anthony Rowbottom
Lancashire & Lakeland Regional Immunology Service
Royal Preston Hospital, Sharoe Green Lane,
Preston PR2 9HT

Email: hra.approval@nhs.net

22 November 2016 (reissued same day)

Dear Miss Callery

Letter of HRA Approval

Study title:	**The use of biospectroscopy for the analysis of blood plasma or serum in CVID patients: a novel approach for the investigation of immune disorders**
IRAS project ID:	**202518**
REC reference:	**16/NW/0502**
Sponsor	**Lancashire Teaching Hospitals NHS Foundation Trust**

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

IRAS project ID	202518

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from www.hra.nhs.uk/hra-approval.

Appendices

The HRA Approval letter contains the following appendices:

- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval

The document *"After Ethical Review – guidance for sponsors and investigators"*, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:

- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the *After Ethical Review* document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the HRA website, and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the HRA website.

Scope

HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found at http://www.hra.nhs.uk/resources/applying-for-reviews/nhs-hsc-rd-review/.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

User Feedback

IRAS project ID	202518

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please email the HRA at hra.approval@nhs.net. Additionally, one of our staff would be happy to call and discuss your experience of HRA Approval.

HRA Training

We are pleased to welcome researchers and research management staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

Your IRAS project ID is **202518**. Please quote this on all correspondence.

Yours sincerely

Michael Higgs
Assessor

Email: hra.approval@nhs.net

Copy to: *Mrs Gemma Whiteley, Lancashire Teaching Hospitals NHS Foundation Trust*
Dr Anthony Rowbottom, Lancashire Teaching Hospitals NHS Foundation Trust

IRAS project ID	202518

Appendix A - List of Documents

The final document set assessed and approved by HRA Approval is listed below.

Document	*Version*	*Date*
Copies of advertisement materials for research participants	1	25 April 2016
IRAS Application Form [IRAS_Form_15062016]		15 June 2016
Letters of invitation to participant	2	28 July 2016
Non-validated questionnaire [Control volunteers]	2	28 July 2016
Participant consent form [Control volunteers]	3	10 November 2016
Participant consent form [CVID]	3	10 November 2016
Participant information sheet (PIS) [Control volunteers]	3	10 November 2016
Participant information sheet (PIS) [CVID]	3	10 November 2016
Research protocol or project proposal [RD Protocol]	3	28 July 2016
Summary CV for Chief Investigator (CI) [A. Rowbottom]		
Summary CV for student [E Callery]		
Summary CV for supervisor (student research) [F Martin]		
Summary, synopsis or diagram (flowchart) of protocol in non technical language	1	25 April 2016

IRAS project ID	202518

Appendix B - Summary of HRA Assessment

This appendix provides assurance to you, the sponsor and the NHS in England that the study, as reviewed for HRA Approval, is compliant with relevant standards. It also provides information and clarification, where appropriate, to participating NHS organisations in England to assist in assessing and arranging capacity and capability.

For information on how the sponsor should be working with participating NHS organisations in England, please refer to the, *participating NHS organisations, capacity and capability* and *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* sections in this appendix.

The following person is the sponsor contact for the purpose of addressing participating organisation questions relating to the study:

Name: Emma Callery
Tel: 01772522189
Email: emma.callery@lthtr.nhs.uk

HRA assessment criteria

Section	HRA Assessment Criteria	Compliant with Standards	Comments
1.1	IRAS application completed correctly	Yes	No comments
2.1	Participant information/consent documents and consent process	Yes	Participant Information Sheets were amended to align with HRA Approval standards.
3.1	Protocol assessment	Yes	No comments
4.1	Allocation of responsibilities and rights are agreed and documented	Yes	This is a single site, same-sponsor study. Therefore no Statement of Activities, Schedule of Events or other agreement documents are expected to be used.
4.2	Insurance/indemnity arrangements assessed	Yes	Where applicable, independent contractors (e.g. General Practitioners) should ensure that the professional indemnity provided by their medical defence organisation covers the activities expected of them for this

IRAS project ID	202518

Section	HRA Assessment Criteria	Compliant with Standards	Comments
			research study
4.3	Financial arrangements assessed	Yes	The study is funded by the Sponsor.
5.1	Compliance with the Data Protection Act and data security issues assessed	Yes	No comments
5.2	CTIMPS – Arrangements for compliance with the Clinical Trials Regulations assessed	Not Applicable	No comments
5.3	Compliance with any applicable laws or regulations	Yes	No comments
6.1	NHS Research Ethics Committee favourable opinion received for applicable studies	Yes	A favourable ethical opinion of the study was given by the North West - Preston Research Ethics Committee on 16 August 2016. Changes to documents since the REC favourable opinion were submitted as non-substantial amendments.
6.2	CTIMPS – Clinical Trials Authorisation (CTA) letter received	Not Applicable	No comments
6.3	Devices – MHRA notice of no objection received	Not Applicable	No comments
6.4	Other regulatory approvals and authorisations received	Not Applicable	No comments

Participating NHS Organisations in England

This provides detail on the types of participating NHS organisations in the study and a statement as to whether the activities at all organisations are the same or different.
This is a single site study involving the collection of blood and serum samples to analyse by biospectroscopy and the completion of questionnaires. If this study is subsequently extended to other NHS organisation(s) in England, an amendment should be submitted to the HRA, with a Statement of Activities and Schedule of Events for the newly

IRAS project ID	202518

participating NHS organisation(s) in England.

The Chief Investigator or sponsor should share relevant study documents with participating NHS organisations in England in order to put arrangements in place to deliver the study. The documents should be sent to both the local study team, where applicable, and the office providing the research management function at the participating organisation. For further guidance on working with participating NHS organisations please see the HRA website.

If chief investigators, sponsors or principal investigators are asked to complete site level forms for participating NHS organisations in England which are not provided in IRAS or on the HRA website, the chief investigator, sponsor or principal investigator should notify the HRA immediately at hra.approval@nhs.net. The HRA will work with these organisations to achieve a consistent approach to information provision.

Confirmation of Capacity and Capability

This describes whether formal confirmation of capacity and capability is expected from participating NHS organisations in England.
This is a single site study sponsored by the site. The R&D office will confirm to the CI when the study can start.

Principal Investigator Suitability

This confirms whether the sponsor position on whether a PI, LC or neither should be in place is correct for each type of participating NHS organisation in England and the minimum expectations for education, training and experience that PIs should meet (where applicable).
The Chief Investigator will oversee research activities at the site. The Sponsor does not expect the CI or student to have completed any specific formal training in order to conduct the study. The Sponsor expects the CI and student to show ongoing demonstration of competency and compliance with HCPC standards of proficiency, in addition to service requirements for laboratory accreditation. GCP training is not a generic training expectation, in line with the HRA statement on training expectations.

HR Good Practice Resource Pack Expectations

This confirms the HR Good Practice Resource Pack expectations for the study and the pre-engagement checks that should and should not be undertaken
All members of the research team are employed by the site, therefore no Letters of Access or Honorary Research Contracts are required.

Other Information to Aid Study Set-up

This details any other information that may be helpful to sponsors and participating NHS organisations in England to aid study set-up.

IRAS project ID	202518

The applicant has indicated that they do not intend to apply for inclusion on the NIHR CRN Portfolio.

APPENDIX D

PATIENT INFORMATION AND CONSENT

This appendix contains study-related documents (Patient Information Sheet, Consent form, invitation to study letter) approved by the Integrated Research Approval System (IRAS) and Health Research Authority (HRA). The patient information sheets were sent out to potential study participants via post, with an accompanying invitation letter from the Consultant, to patients previously diagnosed with Common Variable Immunodeficiency 2 weeks prior to their appointments at the immunodeficiency follow- up clinic. Consent was taken either by me or by the Consultant following the clinic appointment. All verbal and written communication was captured in the electronic patient notes.

Lancashire Teaching Hospitals
NHS Foundation Trust

A Novel Approach Using Biospectroscopy for the Investigation of CVID

Immunology Department
Royal Preston Hospital,
Sharoe Green Lane,
Fulwood,
Preston.
PR2 9HT

PARTICIPANT INFORMATION SHEET

You are being invited to participate in a research study done by Lancashire & Lakeland Immunology Service at Lancashire Teaching Hospitals NHS Foundation Trust. Please read the following information carefully before you decide whether to take part in the study or not.

WHAT IS THE PURPOSE OF THIS STUDY?
This pilot study will help healthcare professionals and scientists gain a better understanding about your immune deficiency, Common Variable Immune Deficiency (CVID). For over 80% of patients the underlying causes for the production of low levels of antibody are unknown. It is also unknown why the clinical features of CVID can vary significantly between patients; from mild forms in those who suffer only from mild infections, to more severe forms, which are associated with multiple disease-related complications, such as gastrointestinal problems, autoimmune conditions and cancers. Once more is known about the causes of CVID, patients could be diagnosed more quickly, managed more effectively, and those at risk of the more severe forms, identified earlier.

WHAT IS THE DIFFERENCE IN TREATMENT?
There will be no difference in your current treatment or care.

WHY HAVE I BEEN CHOSEN?
This study is being conducted by the Immunology department at Royal Preston Hospital. You have been chosen as you are currently under the care of the Clinical Immunology Consultant, Dr Pavaladurai Vijayadurai.

WHAT WILL HAPPEN IF I TAKE PART?
If you have read and understood the information provided and wish to take part, you will be asked to sign the informed consent document. If you have any questions, these can be answered before you decide to take part, and at any point during the study. To take part you will be asked to provide 2 additional blood samples (approximately 2 teaspoons of blood), which can be collected at the same time as your routine blood test.

We will look for markers in the blood samples you provided and compare these with samples collected from participants without CVID. This may allow us to detect a 'fingerprint' of markers in CVID patients and potentially improve the diagnosis and management of future patients.

The study will be completed in accordance with the Data Protection Act (1998), the NHS Code of Confidentiality and the GMC confidentiality guidelines. Participant data will be

Lancashire Teaching Hospitals
NHS Foundation Trust

A Novel Approach Using Biospectroscopy for the Investigation of CVID

collected by Dr Vijayadurai and the Clinical Scientists within the Immunology department, who routinely access electronic patient records and immunology blood tests results. Participant information will be used in order to group patients based on clinical symptoms and immunology blood test results in conjunction with the research data generated from the study. Participant research data will be stored in anonymised form, with any patient identifiable information only accessed on NHS login-restricted computers and stored in password-protected files. Long-term storage of paperwork will be in accordance with Trust policy in secure deep storage archiving facility. Once the samples have been analysed they will be stored securely until the end of the study, at which point they will be sent for incineration according to Trust protocol and Human Tissue Authority's Code of Practice.

WHAT ARE THE POSSIBLE BENEFITS OF TAKING PART?

By taking part in this research study you will be helping Immunology scientists to better understand CVID. Important findings from the study may be published in peer-reviewed scientific journals, and may become standard medical practice in the future.

WHAT ARE THE POSSIBLE RISKS OF TAKING PART?

In taking part in this study, the possible risks include minor discomfort on the collection of additional blood samples, where possible these will be taken at the same time as routine blood collection to reduce discomfort.

WILL MY TAKING PART BE KEPT CONFIDENTIAL?

Yes, once your samples have been collected they will be given a unique identity number to ensure patient confidentiality. Only healthcare professionals within the Immunology department will have access to your results and information.

WHAT IF I CHANGE MY MIND ABOUT TAKING PART?

You can change your mind about taking part at any time, up to the donation of blood samples. Your standard of care will not be affected if you change your mind. Additionally, you can seek further information or submit any queries to the Immunology department at any point.

WHO HAS REVIEWED THIS RESEARCH

This study has been reviewed by the R&D team in the Centre for Health Research and Innovation at Lancashire Teaching Hospitals NHS Foundation Trust and has been through the Health Research Authority (HRA) NHS approval process including a National Research Ethics Committee.

Should you be unhappy with your experience or wish to make a complaint, please contact the Customer Care Team, Royal Preston Hospital, Telephone no. : 01772 522521 or email: customer.care@lthtr.nhs.uk

If you have any questions about the study, please contact: Emma Callery, Senior Clinical Scientist Immunology. Pathology Department, Royal Preston Hospital Tel: 01772 522134 or 01772 528129 Email: emma.callery@LTHTR.nhs.uk

A Novel Approach Using Biospectroscopy for the Investigation of CVID

Royal Preston Hospital,
Sharoe Green Lane,
Fulwood,
Preston.
PR2 9HT

Name of Researcher: Emma Callery (Senior Clinical Scientist Immunology)
R&D Reference Number: 2114

PARTICIPANT CONSENT FORM

Research Study: A novel laboratory approach for the Investigation of CVID patients

Patient name:

Patient Study Number:

Please initial box

1. I confirm that I have read and understand the information sheet dated 10/11/16 (version 3) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected.	☐
3. I agree to have 2 blood sample tubes collected: • 1 blood serum (approx. 5 mL) • 1 blood plasma (approx. 5 mL)	☐
4. I understand that data from the study may be looked at by regulatory authorities or by persons from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to this data.	☐
5. I understand that relevant sections of my medical notes may be looked at by key individuals involved in the study. I give permission for these individuals to have access to my records.	☐
6. I agree to take part in the above study.	☐

________________	____________	______________________
Name of Patient	Date	Signature
________________	____________	______________________
Name of Person taking consent	Date	Signature

When completed: 1 for participant; 1 for researcher site file; 1 (original) to be kept in medical notes.

IRAS212518

IRAS212518

A Novel Approach Using Biospectroscopy for the Investigation of CVID

Lancashire & Lakeland Immunology Service
Outpatients Department
Royal Preston Hospital
Sharoe Green Lane
Fulwood
Preston
Lancashire
PR2 9HT

Consultant Clinical Immunologist	Dr P Vijayadurai	01772-522131
Locum Consultant Allergist	Dr V Crump	01772-522130
Allergy/ Immunology Specialist Nurse	Kathryn Haworth	01772-522130
Staff Nurse in Allergy/ Immunology	Carole Strickland	01772-522404
Staff Nurse in Allergy/ Immunology	Jenny Cottam	01772-522404
Secretary	Rebecca Brierley	01772-522130
Fax 01772 522021		

Dear Mr/Mrs [Surname],

As a patient at Lancashire Teaching Hospitals NHS Foundation Trust, you will be aware that we continually strive to expand and improve our specialist services, whilst always providing excellent care with compassion. One of the ways in which we can do this is through research and development of diagnostic services for a wide range of conditions, including Common Variable Immune Deficiency (CVID).

We would like to invite you to take part in an upcoming research study done by Lancashire & Lakeland Immunology Service at Lancashire Teaching Hospitals NHS Foundation Trust. This study will help healthcare professionals and scientists gain a better understanding about CVID.

Please find the enclosed patient information sheet describing the study. If you feel you would like to take part and have read and understood the information provided, we will answer any questions and collect the enclosed consent form at your next immunology outpatient appointment.

Should you not wish to take part, your standard of care will not be affected, please disregard the additional information enclosed.

Yours sincerely

Dr P Vijayadurai
Consultant Clinical Immunologist

Appendix E

This appendix includes the final marks for Module A, PGDip Leadership and Management in the Healthcare Sciences. A description of the five units and the assignment requirements has been included in the table below. For the life sciences students, the A units are usually undertaken over a five year period however the author completed the five units in the first two years of the Higher Specialist Scientific Training (HSST). This thesis forms Module C of the HSST; additional assessed requirements within this module include a literature review, a lay presentation, an innovation proposal and two online modules, i) Research Integrity, ii) Export Control.

A Units and C Credits for DClinSci Thesis

Alliance Manchester Business School (AMBS) A Units		
Unit Title	**Credits**	**Assignment Word Count**
A1: Professionalism and Professional Development in the Healthcare Environment	30	A1 – Assignment 1 – 2500 words (40%) A1 – Presentation – 10 mins (10%) A1 – Assignment 2 – 3000 words (50%)
A2: Theoretical Foundations of Leadership	20	A2 – Assignment 1 – 3000 words (50%) A2 – Assignment 2 – 3000 words (50%)
A3: Personal and Professional Development to Enhance Performance	30	A3 – Assignment 1 – 1500 words (30%) A3 – Assignment 2 – 4000 words (70%)
A4: Leadership and Quality Improvement in the Clinical and Scientific Environment	20	A4 – Assignment 1 – 3000 words (50%) A4 – Assignment 2 – 3000 words (50%)
A5: Research and Innovation in Health and Social care	20	A5 – Presentation – 15 mins (25%) A5 – Assignment – 4000 words (75%)

PGDip Leadership & Management in the Healthcare Sciences
Unit marks ratified by Board of Examiners, November 2018

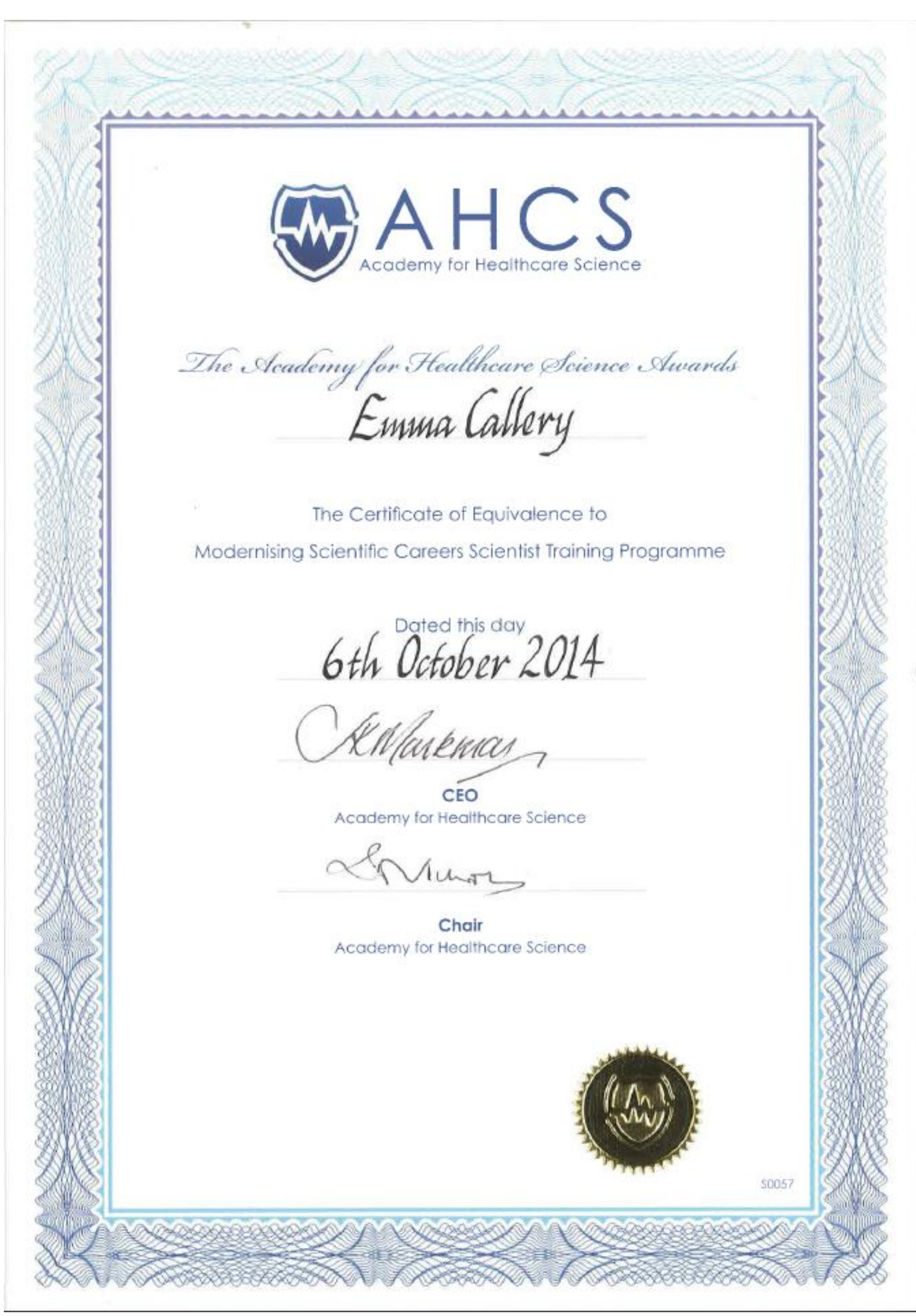
Trainee name: Emma Callery
Student ID: 5805812
Award: PG Credit

Unit	Unit Title	Mark	Credits
BMAN73511	Unit A1 Professionalism and Professional Development in the Healthcare Environment	65% **Pass**	**30**
BMAN73522	Unit A2 Theoretical Foundations of Leadership	78% **Pass**	**20**
BMAN73531	Unit A3 Personal and Professional Development to Enhance Performance	62% **Pass**	**30**
BMAN73542	Unit A4 Leadership and Quality Improvement in the Clinical and Scientific Environment	72% **Pass**	**20**
BMAN73550	Unit A5 Research and Innovation in Health and Social Care	68% **Pass**	**20**
			120 / 120

APPENDIX F

QUALIFICATIONS

The achievement of this doctoral degree will result in completion the Higher Specialist Training Programme (HSST) with DClinSci. The project presented in this thesis has been accepted by The Royal College of Pathology; it forms the research component of the Part Two examinations for eligibility to become a Fellow of The Royal College of Pathology. In order to complete the HSST, applicants on the life sciences pathway must have achieved Fellowship of the Royal College of Pathology, received official confirmation of the academic qualification from the University and have all Standards of Proficiency signed off within the e-portfolio. The author is a registered Clinical Scientist with the Health and Care Professions Council, has had all five domains of the Standards of Proficiency signed off, has passed The Royal College of Pathology Part One examinations in Spring 2019 and Part Two Examinations (Practical and Oral) in Spring 2021. Evidence of qualifications are included in this appendix.



AHCS
Academy for Healthcare Science

The Academy for Healthcare Science Awards

Emma Callery

The Certificate of Equivalence to
Modernising Scientific Careers Scientist Training Programme

Dated this day
6th October 2014

CEO
Academy for Healthcare Science

Chair
Academy for Healthcare Science

S0057

College Reference No: **20010082**
Candidate No: **642**

Emma Callery

24 May 2019

Dear Miss Callery

FRCPath Part 1 Examination in Immunology – Spring 2019

I am pleased to inform you that you have satisfied the Examiners in the Part 1 Examination. You have therefore completed this part of the examination and you are now eligible to become a Diplomate of the College.

Council will formally ratify the nominations to Diplomateship on **1 August 2019**, after which date you may use the post-nominal DipRCPath. Your Diplomateship certificate will also be sent out to you after this date.

Please complete and return the enclosed consent form and address slip to the Membership Department, in the envelope provided, as soon as possible and no later than **Friday 19 July 2019**. The subscription year runs from January and once you have returned your consent form an invoice will be sent out for your subscription due. This will take into account any fee you may have already paid and pro-rata rates for the year. For current rates please refer to our website (https://www.rcpath.org/discover-pathology/membership/subscription-fees.html).

Please check the order and spelling of your full name, as displayed at the top of this letter. This is how your name will be presented on your Diplomateship certificate. If there is an error please notify the Examinations Department (exams@rcpath.org) in writing as soon as possible.

If you do not wish to become a Diplomate of the College please indicate on the enclosed form or e-mail the Membership Department on membership@rcpath.org.

Details of the Part 2 Examinations can be found on the Examinations page of our website.

Congratulations on your success in this examination.

Yours sincerely

Dr Andrew Day
Clinical Director of Examinations

Enc: Consent form and address slip
Envelope addressed to the Membership Department

College Reference Number: **20010082**
Candidate Number: **156**

Emma Callery

4 June 2021

Dear Miss Callery

FRCPath Part 2 Examination in Immunology – Spring 2021

I am pleased to inform you that you have satisfied the Examiners in the Part 2 Examination.

However, as you are aware, you are not yet eligible to become a Fellow of The Royal College of Pathologists as your Part 2 Project has not yet been approved.

We look forward to receiving the project in due course. If you have any queries about your project, please contact exams@rcpath.org.

Congratulations on your success in this examination.

Yours sincerely

Dr Sanjiv Manek
Clinical Director of Examinations

Appendix F

Standards of Proficiency - Summary of Onefile e-portolio

Onefile

8 0 Emma Callery
Trainee

Unit	Progress	Status
SoP Domain 1 - Professional Practice Standard 1 - Practise with the professionalism of a Consultant Clinical Scientist Standard 2 - Ensure professionalism in working with peers and with service users Standard 3 - Ensure professionalism in areas of governance and service accreditation Standard 4 - Direct the education and training of others Read about this unit	100%	Unit Signed Off
SoP Domain 2 - Scientific Practice Standard 5 - Lead scientific services Standard 6 - Direct scientific validation and evaluation Standard 7 - Assure safety in the scientific setting Read about this unit	100%	Unit Signed Off
SoP Domain 3 - Clinical Practice Standard 8 - Ensure clinical relevance of scientific services provided Standard 9 - Deliver effective clinical services Read about this unit	100%	Unit Signed Off
SoP Domain 4 - Research, Development and Innovation Standard 10 - Lead research, development and innovation in clinical priority areas Standard 11 - Evaluate research, development and innovation outcomes to improve scientific service provision Standard 12 - Promote a culture of innovation Standard 13 - Assure research governance Read about this unit	100%	Unit Signed Off
SoP Domain 5 - Clinical Leadership Standard 14 - Ensure strategic leadership Standard 15 - Ensure clinical scientific leadership Standard 16 - Resource management Read about this unit	100%	Unit Signed Off

APPENDIX G

SUPPLEMENTARY INFORMATON

This appendix contains the supplementary information associated with the Nature Scientific Reports manuscript in Chapter 3. This information is also available online:

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-019-43196-5/MediaObjects/41598_2019_43196_MOESM1_ESM.docx (recent date accessed 11/10/2021)

New approach to investigate Common Variable Immunodeficiency patients using spectrochemical analysis of blood

Emma L. Callery,*[a] Camilo L. M. Morais,[b] Maria Paraskevaidi,[b e] Vladimir Brusic,[c] Pavaladurai Vijayadurai,[a] Ariharan Anantharachagan,[a] Francis L. Martin, [b] & Anthony W. Rowbottom[*a d]

Affiliations

*Corresponding authors

[a] Department of Immunology, Lancashire Teaching Hospitals NHS Foundation, Preston PR2 9HT, UK

[b] School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, UK

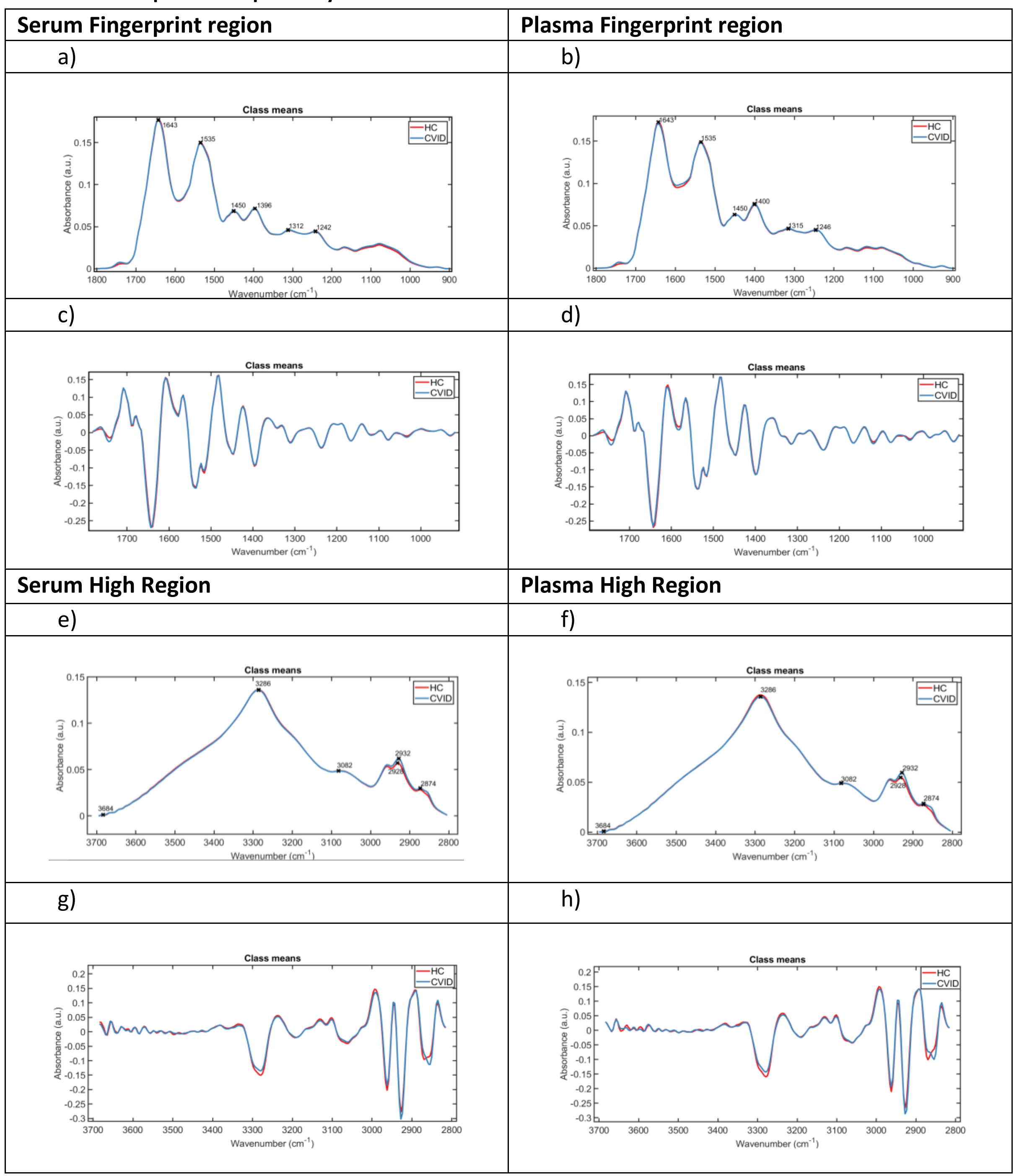
[c] University of Nottingham Ningbo China, Ningbo 315100, China

[d] School of Medicine, University of Central Lancashire, Preston PR1 2HE, UK

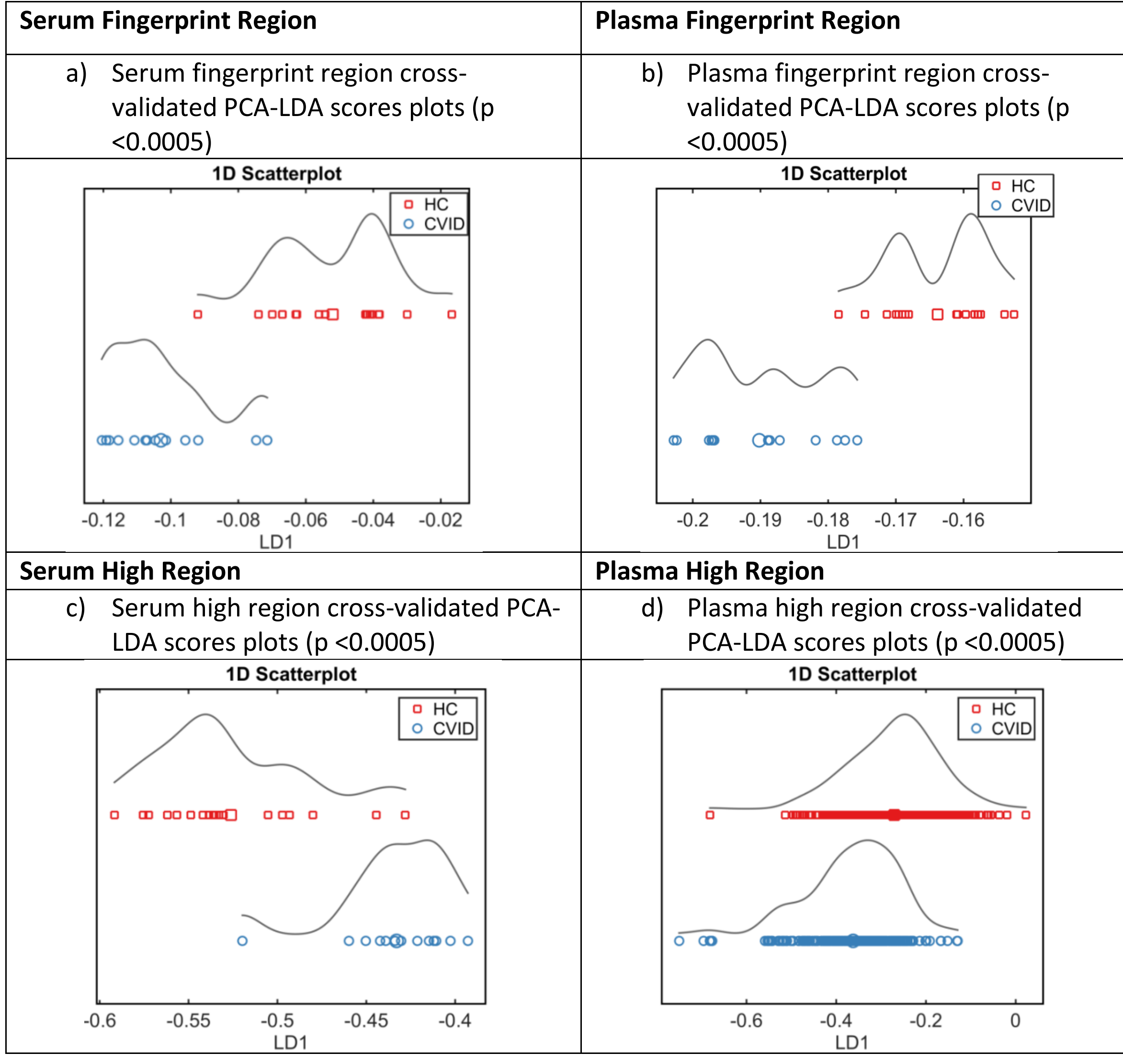
[e] Department of Surgery and Cancer, Imperial College London, W12 0HS

Supplementary Information

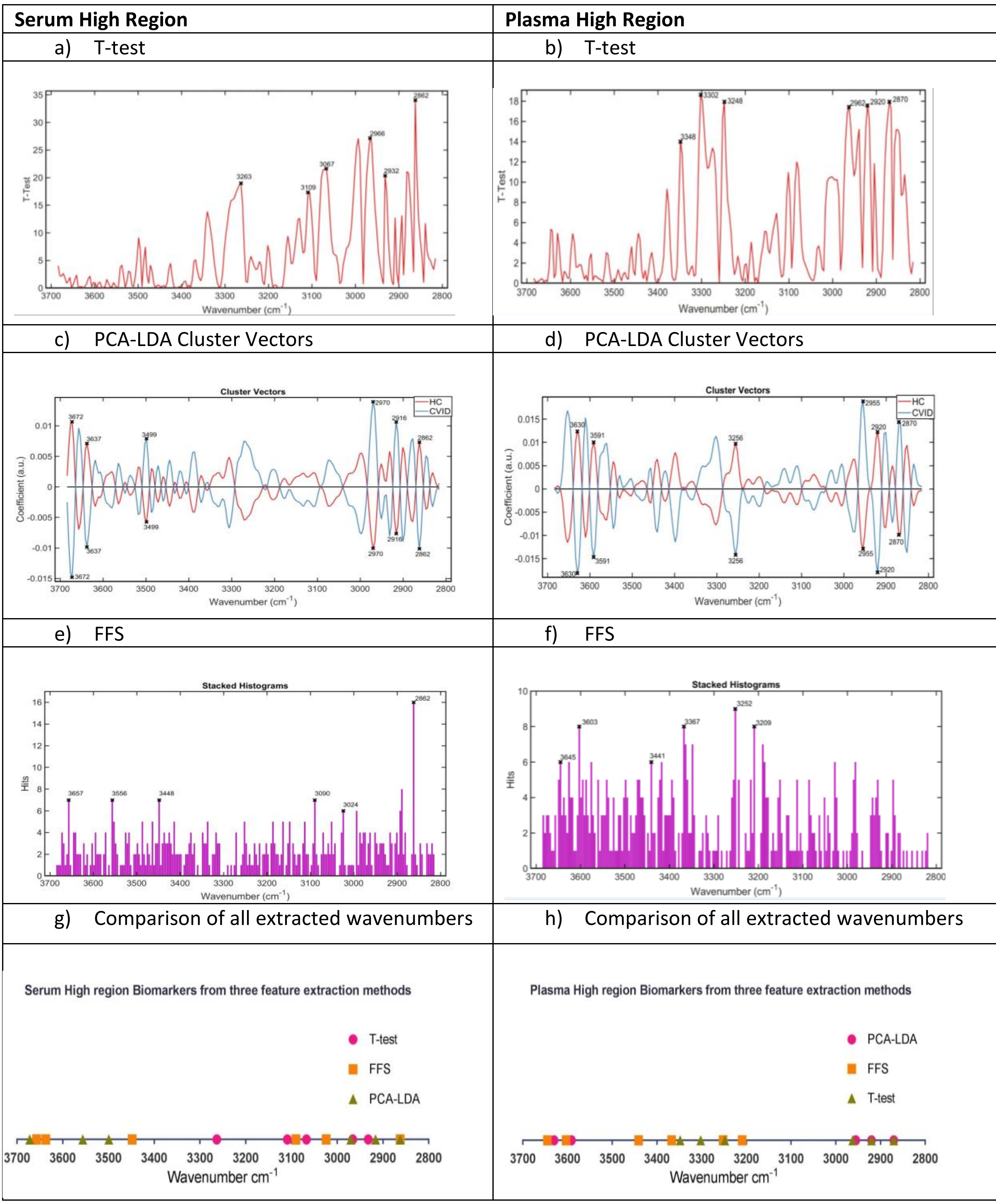
Supplemental Figure 1 – Improving robustness, accuracy and interpretability of the data. Class means of pre-processing techniques applied to all spectra in training dataset (CVID *n*=13 (260 spectra); HC *n*=18 (360 spectra)) to correct for experimental variation and to improve resolution of peak overlap. For each subject 20 individual spectra were acquired per biofluid (10 replicates of each dried blood spot, in duplicate). a-d Fingerprint region (900-1800 cm^{-1}). a, b, Rubber-band corrected, vector normalised spectra for serum and plasma respectively. c, d, Second order differentiated, vector normalised spectra for serum and plasma respectively. e-h, High region (2800-3700 cm^{-1}). Rubber-band corrected, vector normalised spectra for serum and plasma respectively. g, h, Second order differentiated, vector normalised spectra for serum and plasma respectively.



Supplemental Figure 2 - Supervised multivariate analysis techniques (PCA-LDA) successfully segregate classes on a subject level basis (CVID vs HC). a, b, Fingerprint region (900-1800 cm^{-1}); 1D scores plots (LD1) after cross-validated PCA-LDA of the training dataset (CVID *n*=13; HC *n*=18) for serum and plasma respectively. c,d, High region (2800-3700 cm^{-1}); 1D scores plots (LD1) after cross-validated PCA-LDA of the training dataset (CVID *n*=13; HC *n*=18) for serum and plasma respectively.



Supplemental Figure 3. - Serum and plasma High region biomarkers identified using three feature extraction methods. a, Serum T-test. b, Plasma T-test. c, Serum FFS. d, Plasma FFS. e, Serum cross-validated PCA-LDA. f, Plasma cross-validated PCA-LDA. i, j, Visual representation of wavenumber location for extracted biomarkers from each method for serum and plasma respectively. FFS Forward Feature Selection.



Supplemental Table 1. Patient demographics

Study No.	Sex	Age (y)	B cells	MemB	CSM	IV/ SC	Dose (g)	last infused	Days post infusion	monthly/ weekly	Pre-tx IgG	Dx (years)	Immunoglobulin level at recruitment			Spenomegaly	Autoimmunity	Bronciectasis	Malignancy	ENT	GI	Stable/ Improved/Deteriorating
65639	M	22	576	28	8	SC	5.25	26/02/2017	2	weekly	2.6	3	9.8	0	0.3	No	No	No	No	No	No	Stable
39591	M	21		- Not done -		IV	30	03/01/2017	21	3-weekly	0.52	8	6.7	0	0.01	Yes	Diabetes	No	No	No	No	Stable (passed away)
75983	F	62	267	2	0	IV	25	03/01/2017	28	3-weekly	0.38	13	10.4	0	0.03	No	No	No	No	Yes	No	Stable
53349	M	70	200	10	0	SC	8	18/01/2017	6	weekly	0.77	5	7.3	0	0.38	Splenectomy	No	Yes	No	Infecti-ons	Diarrhoea	Stable
81275	F	51	295	15	2	N/A	-	-	-	-	4.53	3	3.8	0.17	2.22	Splenectomy 1988	Haemolytic anaemia 1999	No	Bladder	No	No	Stable
70913	F	26	423	3	0	SC	6	19/02/2017	2	weekly	0.2	6	5.3	0	0.02	No	No	No	No	No	No	Stable
90802	M	78	24	42	33	SC	12	02/05/2017	6	weekly	2.35	12	7.5	0.87	0.18	No	No	No	No	No	No	Stable
87088	F	53	27	5	0	SC	6	20/09/2017	6	weekly	0.34	16	5.9	0	0.01	Yes	Thrombocyt-openia & neutropenia	Yes	Squamous cell (skin)	No	Yes	Deteriorated
10415	F	56	240	15	1	IV	25	19/04/2017	6	3-weekly	unknown	>10	11.4	0	0.03	No	No	Yes	No	Yes	No	Stable
19759	F	20	50	16	7	N/A	-	-	-	-	1.33	2	1.3	0	0.02	No	No	No	No	Yes	Diarrhoea	Improved
98037	M	50	287	23	6	IV	35	unknown	unknown	4-weekly	0.53	11	7.2	0	0.02	No	No	No	No	No	No	Stable
13582	F	79	64	18	10	IV	25	01/02/2017	28	4-weekly	3.25	7	7.1	2.07	0.28	No	No	Yes	Basal Cell	No	No	Deteriorating (Passed away)
48004	M	33	28	83	79	SC	10	23/05/2017	0	weekly	1.47	3	7.8	0	0.06	No	No	No	No	No	No	Stable
73741	M	56	28	83	79	SC	9.6	21/05/2017	2	weekly	10.65	Pre 2005	13.8	0.09	0.01	No	No	No	No (MGUS)	No	No	Stable
97511	M	80	39	15	6	SC	8	22/02/2017	6	weekly	3.0	3	6.47	0.67	0.33	No	No	Yes	No	Sinus infecti-ons	No	Deteriorating
50735	F	53	11	4	0	SC	9	23/07/2017	2	weekly	0.3	3	5.07	0	0	No	No	Yes	No	No	No	Stable
90326	M	53	69	33	17	IV	40	04/06/2017	9	3-weekly	unknown	>20	11.16	0	0	No	No	Yes	No	No	Yes	Deteriorating
29891	M	31	23	0	0	SC	9.9	unknown	unknown	weekly	0.3	3	7.89	0	0.01	Yes	No	No	No	No	No	Stable
36952	M	48	28	22	0	SC	6.5	16/01/2017	1	weekly	1.9	10	5.61	0	0.05	Yes	Cytopenia	Yes	Oropharyngeal	No	No	Stable
31108	M	61	173	5	0	SC	10.4	24/03/2017	4	5-6 daily		1	6.24	0.04	0.03	No	Primary biliary cirrhosis	Yes	No	No	No	Improved
61236	F	72	125	4	2	SC	6	unknown	unknown	weekly	4.7	12	6.71	0.44	0.05	No	No	Yes	Breast (2008)	No	No	Stable

Supplemental Table 2. Parameters for the SVM models. Classification of HC vs CVID.

	Serum		**Plasma**	
Grid-search parameters	*c*	γ	*c*	γ
Fingerprint region	1 x 10^7	1x 10^-1	1 x 10^7	1x 10^-1
High region	1 x 10^3	1x 10^1	1 x 10^3	1x 10^1

Supplemental Table 3. Parameters for the SVM models following sub-classification of CVID patients. HC vs CVID-non complications vs CVID-complications.

	Serum		**Plasma**	
Grid-search parameters	*c*	γ	*c*	γ
Fingerprint region	1 x 10^-1	1x 10^-7	1 x 10^3	1x 10^-1
High region	1 x 10^7	1 x 10^-2	1 x 10^5	1 x 10^-1

Supplemental Table 4. Serum & plasma fingerprint wavenumbers.

Wavenumber (cm^{-1})	Biofluid	Tentative Assignments	Reference	Method	P-value (T-Test)	↑↓ In CVID
933	Serum	Z type DNA	[64]	T-Test	0.0203*	
984	Plasma	Phosphodiester region (900-1300 cm^{-1})	[51,65]	FFS	0.000469***	↑
1007	Plasma	Ring stretching vibrations mixed strongly with CH in-plane bending (1000-50 cm^{-1})	[66]	T-Test	3.15E-25****	↑
1034	Serum	Collagen	[5]	T-Test & PCA-LDA	1.31E-12****	↑
1053	Plasma	vC - O & dC - O of carbohydrates, Shoulder of 1121 cm^{-1} (Symmetric phosphodiester stretching band) DNA (nucleic acids and phospholipids)	[5–7]	FFS, T-test	2.64E-69****	↑
1084	Plasma	DNA (band due to PO_2^- vibrations), Nucleic acid–Phosphate band	[3,6–11]	PCA-LDA	0.007575*	↑
1107	Plasma	v(CO), v(CC), ring (polysaccharides, pectin)	[12]	FFS	2.74E-05****	↑
1107	Serum	v(CO), v(CC), ring (polysaccharides, pectin)	[12]	FFS	2.53E-20****	
1115	Serum	Symmetric stretching P – O – C	[13]	T-Test	2.02E-26****	↑
1119	Plasma	C - O stretching mode	[14]	T-Test	6.94E-30****	
1312	Serum	Amide III band components of proteins	[15,16]	T-Test	1.49E-08****	
1393	Serum	CH_2 wagging vibration of the acyl chains (phospholipids) (1250-400 cm^{-1})	[5]	FFS	4.23E-06****	
1416	Plasma	Deformation C–H, N–H, stretching C–N	[13]	PCA-LDA	8.34E-13****	↑
1420	Serum	Ring stretching vibrations with CH in-plane bending (1400-500 cm^{-1})	[17]	PCA-LDA	8.73E-07****	
1435	Plasma	Ring stretching vibrations with CH in-plane bending (1400-500 cm^{-1})	[17]	T-Test	1.43E-28****	
1528	Serum	C=N guanine, adenine, cytosine	[11,13]	PCA-LDA	5.57E-12****	↓
1566	Plasma	Amide II (1540-650 cm^{-1})	[18]	FFS	0.004467**	↑
1589	Serum	Amide II (1540-650 cm^{-1})	[18]	FFS	0.00109**	
1639	Plasma	C=C thymine, adenine, N-H guanine Amide I	[13,15]	T-Test	3.08E-39****	↓
1651	Serum	Amide I: (mainly protein C=O stretching), α-helical structure	[1,19]	T-Test	0.0462*	
1732	Plasma	C=O stretching band mode of the fatty acid ester (1725-45 cm^{-1})	[20]	T-Test	1.09E-19****	
1759	Serum	C=O vibrations of esters (triglycerides)	[10,21]	FFS	7.61E-12****	↑
1763	Plasma	Fatty acid esters (1700-1800 cm^{-1})	[20]	FFS	1.30E-13****	

*Significant to *P <0.05; **P <0.005; ***P 0.0005; ****P <0.00005.*

Supplemental Table 5. Serum and Plasma High region biomarkers.

Wavenumber (cm^{-1})	Biofluid	Tentative Assignments	Reference	Method	P-value
2862	Serum	Fatty acids	[19]	T-Test/PCA-LDA/FFS	1.19E-13****
2870	Plasma	CH_3 symmetric stretching: protein side chains, lipids, with some contribution from carbohydrates and nucleic acids	[1]	T-Test, PCA-LDA	1.13E-18****
2916	Serum	vibrations of CH_2 and CH_3 of phospholipids, cholesterol and creatinine	[22]	PCA-LDA	2.79E-10****
2920	Plasma	C–H, Lipid region, CH_3, CH_2-lipid and protein (2800-3000 cm^{-1})	[5,23,24]	T-Test, PCA-LDA	2.64E-18****
2932	Serum	C-H, C-H stretching bands in malignant tissues	[23,25]	T-Test	3.99E-12****
2962	Plasma	CH_3 asymmetric stretching	[26]	T-Test	3.72E-18****
2970	Serum	ν_{as} CH_3, lipids, fatty acids	[12]	PCA-LDA	1.53E-05****
3024	Serum	C-H stretching vibrations of methyl (CH_3) and methylene (CH_2) groups and olefins (2800–3100 cm^{-1})	[25]	FFS	0.0003***
3067	Serum	C-H stretching vibrations of methyl (CH_3) and methylene (CH_2) groups and olefins (2800–3100 cm^{-1})	[25]	T-Test	0.00714*
3248	Plasma	Symmetric and asymmetric vibrations attributed to water. So, it would be better not to consider this region for detailed analysis (3200-550 cm^{-1})	[22]	T-Test	1.14E-18****
3302	Plasma	As above	[22]	T-Test	2.31E-19****
3348	Plasma	As above	[22]	T-Test	0.000206***
3441	Plasma	As above	[22]	FFS	5.05E-05****
3448	Serum	As above	[22]	FFS	2.72E-08****
3499	Serum	As above	[22]	PCA-LDA	0.0146*
3556	Serum	OH bonds(3500–600 cm^{-1})	[27]	PCA-LDA	0.00108**
3591	Plasma	OH bonds(3500–600 cm^{-1})	[27]	PCA-LDA	8.48E-05****
3630	Plasma	O-H stretching (water)	[5]	PCA-LDA	1.10E-05****
3637	Serum	O-H stretching (water) (3000–700 cm^{-1})	[5]	FFS	0.00019**
3645	Plasma	O-H stretching (water) (3000–700 cm^{-1})	[5]	FFS	4.60E-06****
3672	Serum	O-H stretching (water) (3000–700 cm^{-1})	[5]	PCA-LDA	0.032*

*Significant to *P <0.05; **P <0.005; ***P <0.0005; ****P <0.00005.*

References

1. Mostaço-Guidolin, L. B., Murakami, L. S., Batistuti, M. R., Nomizo, A. & Bachmann, L. Molecular and chemical characterization by Fourier transform infrared spectroscopy of human breast cancer cells with estrogen receptor expressed and not expressed. *Journal of Spectroscopy* https://www.hindawi.com/journals/jspec/2010/498719/abs/ (2010) doi:10.3233/SPE-2010-0466.

2. Sindhuphak, R. *et al.* A new approach for the detection of cervical cancer in Thai women. *Gynecologic Oncology* **90**, 10–14 (2003).

3. Zelig, U. *et al.* Pre-screening and follow-up of childhood acute leukemia using biochemical infrared analysis of peripheral blood mononuclear cells. *Biochim. Biophys. Acta* **1810**, 827–835 (2011).

4. Manoharan, R. *et al.* Raman spectroscopy and fluorescence photon migration for breast cancer diagnosis and imaging. *Photochem. Photobiol.* **67**, 15–22 (1998).

5. Fabian, H. *et al.* A comparative infrared spectroscopic study of human breast tumors and breast tumor cell xenografts. *Biospectroscopy* **1**, 37–45 (1995).

6. Fung, M. F. K., Senterman, M. K., Mikhael, N. Z., Lacelle, S. & Wong, P. T. T. Pressure-tuning fourier transform infrared spectroscopic study of carcinogenesis in human endometrium. *Biospectroscopy* **2**, 155–165 (1998).

7. Fukuyama, Y., Yoshida, S., Yanagisawa, S. & Shimizu, M. A study on the differences between oral squamous cell carcinomas and normal oral mucosas measured by Fourier transform infrared spectroscopy. *Biospectroscopy* **5**, 117–126 (1999).

8. Yang, Y. *et al.* Study of tumor cell invasion by Fourier transform infrared microspectroscopy. *Biopolymers* **78**, 311–317 (2005).

9. Yano, K. *et al.* Direct measurement of human lung cancerous and noncancerous tissues by fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool? *Anal. Biochem.* **287**, 218–225 (2000).

10. Caine, S., Heraud, P., Tobin, M. J., McNaughton, D. & Bernard, C. C. A. The application of Fourier transform infrared microspectroscopy for the study of diseased central nervous system tissue. *Neuroimage* **59**, 3624–3640 (2012).

11. Mossoba, M. M. *et al.* Printing microarrays of bacteria for identification by infrared microspectroscopy. *Vibrational Spectroscopy* **38**, 229–235 (2005).

12. Wong, P. T. T., Papavassiliou, E. D. & Rigas, B. Phosphodiester Stretching Bands in the Infrared Spectra of Human Tissues and Cultured Cells. *Applied Spectroscopy* **45**, 1563–1567 (1991).

13. Thumanu, K. *et al.* Diagnosis of liver cancer from blood sera using FTIR microspectroscopy: a preliminary study. *J Biophotonics* **7**, 222–231 (2014).

14. Huang, Z. *et al.* Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int. J. Cancer* **107**, 1047–1052 (2003).

15. Wang, X. *et al.* The study of a single BGC823 cell using Fourier transform infrared microspectroscopic imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **79**, 1660–1662 (2011).

16. Junhom, C., Weerapreeyakul, N., Tanthanuch, W. & Thumanu, K. FTIR microspectroscopy defines early drug resistant human hepatocellular carcinoma (HepG2) cells. *Exp. Cell Res.* **340**, 71–80 (2016).

17. Li, X. *et al.* Raman spectroscopy and fluorescence for the detection of liver cancer and abnormal liver tissue. *Conf Proc IEEE Eng Med Biol Soc* **1**, 212–215 (2004).

18. Prabhakar, S., Jain, N. & Singh, R. A. Infrared spectra in monitoring biochemical parameters of human blood. *J. Phys.: Conf. Ser.* **365**, 012059 (2012).

19. Lucassen, G. W., Veen, G. N. A. V. & Jansen, J. A. J. Band analysis of hydrated human skin stratum corneum attenuated total reflectance Fourier transform infrared spectra in vivo. *JBO, JBOPFO* **3**, 267–281 (1998).

20. Dovbeshko, G. I., Gridina, N. Ya., Kruglova, E. B. & Pashchuk, O. P. FTIR spectroscopy studies of nucleic acid damage. *Talanta* **53**, 233–246 (2000).

21. Laska, J. & Widlarz, J. Spectroscopic and structural characterization of low molecular weight fractions of polyaniline. *Polymer* **46**, 1485–1495 (2005).

22. Khanmohammadi, M. & Garmarudi, A. B. Infrared spectroscopy provides a green analytical chemistry tool for direct diagnosis of cancer. *TrAC Trends in Analytical Chemistry* **30**, 864–874 (2011).

23. Andrus, P. G. & Strickland, R. D. Cancer grading by Fourier transform infrared spectroscopy. *Biospectroscopy* **4**, 37–46 (1998).

24. Ke, Y., Li, Y. & Wang, Z.-Y. The changes of fourier transform infrared spectrum in rat brain. *J. Forensic Sci.* **57**, 794–798 (2012).

25. Stone, N., Kendall, C., Smith, J., Crow, P. & Barr, H. Raman spectroscopy for identification of epithelial cancers. *Faraday Discuss.* **126**, 141–157; discussion 169-183 (2004).

26. Liu, Y. *et al.* Detection of cervical metastatic lymph nodes in papillary thyroid carcinoma by Fourier transform infrared spectroscopy. *Br J Surg* **98**, 380–384 (2011).

27. Lee, J. H., Kim, H. G. & Lee, W. J. Characterization and tissue incorporation of cross-linked human acellular dermal matrix. *Biomaterials* **44**, 195–205 (2015).

APPENDIX H

ADDITIONAL RESEARCH CONDUCTED DURING HSST

Exploring COVID-19 specific immune responses in acute and convalescent phases of infection - EXCOVIR

This appendix contains a summary of a COVID-19 research project 'EXCOVIR' (Exploring COVID-19 specific immune responses in acute and convalescent phases of infection) which the author is currently leading on alongside Professor Anthony Rowbottom. The work associated with this project commenced in May 2020, taking priority over any non-COVID-19 research. Ethical approval was granted on 12th June 2020. The EXCOVIR project is an Immunology research study looking at the cellular responses (T cells) found in patients during- and post-infection with coronavirus. Currently, much of the attention is directed to the ability of the immune system to produce antibodies however the evidence for long term protection and immunity is sparse. A better understanding of the role and longevity of T cell responses are essential. EXCOVIR aims to assess T cell responses in the early phases of SARS-COV-2 infection (before day twelve) and will continue to follow up patients for twelve months to determine longer-term immunity.

As part of this project we have collected serum and saliva sample for analysis with biospectroscopy. We intend to apply this novel platform to explore samples for unique spectral features associated with COVID-19 and with varying severities of disease. The findings from the biospectroscopy analysis will provide a complementary methodology which may offer new knowledge and molecular insight into the pathological mechanisms of COVID-19.

EXCOVIR is being conducted by Lancashire and Lakeland Immunology department (LTHTR), supported by UCLAN Clinical Trials Unit and the NIHR Lancashire Clinical Research Facility (LCRF) at Royal Preston Hospital. The role of the author has included writing the research protocol, completing the IRAS application for ethical approval, creation of study documentation, organisation and chairing of clinical meetings with Consultants and subsequent bi-weekly meetings with research nurses, submission of funding applications, preparing and presenting the Site Initiation Visit (SIV), developing novel methods for investigating SARS-COV-2 specific T cell responses, leading a team of laboratory researchers, analysing samples and performing subsequent data analysis. At the time of writing this thesis, patient follow-up sampling, data collection and data analysis is ongoing, with the aim to submit the findings from the first phase for publication to a peer-reviewed journal before the end of 2021.

The new tests developed within this study will enable us to measure immune cells specific to SARS-CoV-2 in COVID-19 patients following recovery from infection. There are still many unknowns associated with COVID-19, particularly with regard to 'Long-COVID'. The tests developed within this project will help to unravel the immune phenotypes associated with COVID-19 patients, and could potentially be translated into clinical practice to improve patient care. This appendix includes a description of the study (project aims, plan of investigation, immunology investigation, statistical power) taken from the study protocol; a research project communications summary produced for the LTHTR media webpage; example study documentation; and copies of the HRA and Health and Care Research Wales (HCRW) approvals for this study.

Title:

Exploring COVID-19 specific immune responses in acute and convalescent phases of infection - EXCOVIR

Aims:

The primary aim is to characterise host adaptive immune responses in COVID-19 patients over time adding knowledge to the overall immune response following infections with SARS-CoV-2.

There are currently no tests to predict how severe the disease will be in individual patients. Our secondary aims are to identify predictors of disease severity through detection of circulating levels of immune signalling molecules, evaluation of immune cell surface markers and specific immune cell populations in people with varying severities of COVID-19 disease (critical, severe and moderate/mild). If successful, this can be applied to clinical practice as a sensitive and objective marker of disease activity, and, potentially serve as a predictive determinant of disease severity in SAR-CoV-2 infected patients. These findings could inform the future acute care of COVID-19 patients, rapidly identifying those at risk of developing a severe disease requiring immediate treatment and management. We will investigate a new testing method to identify whether people had been previously infected with SARS-CoV-2. This would look at SARS-CoV-2 specific immune cells, a different approach to the SARS-CoV-2 antibodies test. This will be particularly useful in the future for those patients and staff members with a clinical history consistent with COVID-19 but not supported by antibody levels.

Further, we aim to provide additional evidence towards identifying an appropriate correlate of immunity to SARS-CoV-2. By analysing immune responses at future time points (1 month, 6 months, 12 months) after convalescence, we aim to determine the longevity of a potentially protective immune response to SARS-CoV-2 post infection.

Plan of investigation:

This study will involve collecting serial samples from COVID-19 patients over 2 phases of the disease course; an acute infection phase (day 1, day 7 and day 12), and a convalescent phase (day 28, 6 months and 12 months). The acute phase will compare the immune responses of patients admitted to hospital with symptoms of COVID-19 in different disease severity groups.

All patients admitted with COVID-19 will be approached to participate. We will take informed consent directly from the patient if possible and from a consultee if not. We will take saliva and/or blood samples at day 1, day 7, day 12, day 28, 6 months and 12 months. We will use these samples to look at how each individual's immune system responds to the infection in the acute stage and determine if there any specific results that may predict which patients may become very ill. If it can be determined which patients may develop more severe symptoms then this would facilitate an earlier more appropriate treatment plan and also allow for smarter resource and capacity planning.

Recruited patients will be retrospectively classified into severity groups depending on the clinical course of their disease. We anticipate to recruit patients into the following groups i) patients with active COVID-19 disease (moderate symptoms) ii) patients with active COVID-19 disease (severe), and iii), patients with active COVID-19 disease (critical symptoms). Hospital admissions for asymptomatic patients and those with mild symptoms will be rare therefore we will not evaluate the acute infection response in these groups. We will however evaluate the immune responses of recovered mild and asymptomatic COVID-19 patients, with enrolment into the convalescent sampling phase commencing at day 28 post infection.

We will follow the participants for 12 months to measure how the immune system responds and if it sustains (memory) immunity in the longer term. Participants recruited during convalescence will follow the same schedule as above commencing at 28 days. A healthy donor population will be

recruited from volunteers following advertisement at recruiting sites. Healthy donors will provide blood samples at a single time point for use as reference measurements for the study tests in the non- SARS-CoV-2 infected population.

We will investigate a number of immunological parameters in both serum and whole blood samples using established techniques. The primary outcome will be to determine immune cell population characteristics and cytokine profile changes within and between disease severity groups at each time point during acute (active) infection, measured with flow cytometry (cell surface marker staining analysis and TCR Vβ repertoire quantification test), Bio-Plex Multiplex Immunoassays and biospectroscopy. Secondary outcomes will include evaluation of predictive disease severity biomarkers. Differences in any immunological markers (cytokine quantity or profile, immune cell populations, TCR Vβ repertoire quantification) correlating with disease severity will be assessed at each time-point. Correlates of immunity will be explored through identification and quantification of SARS-CoV-2 specific immune response, through specific cell marker analysis and functional testing of these immune cells post recovery from COVID-19 at 28 days, 6 months and 12 months post infection. Results from the tests in this study will be compared and correlated with SARS-CoV-2 serology (antibody) tests.

Immunology Study Investigations

Cellular Assays:

- SARS-COV-2 & Influenza matrix protein A MHC Class I Pentamer analysis
- HLA typing (to assess Pentamer compatibility)
- Immunophenotyping T Cell Subpopulations (TSUB) 10 cell-marker panel
- Immunophenotyping CD4+ T regulatory cells (TREG) 8 cell-marker panel
- Immunophenotyping Lymphocyte activation and plasmablasts (LYAS) 10 cell-marker panel
- SARS-COV2 specific lymphocyte activation (peptide stimulation and intracellular cytokine staining)
- CD4+ TCR VB Family expression (24 Vb families)

Serological assays:

- 17-plex cytokine analysis (G-CSF, GM-CSF, IFN-γ, IL-1β, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12 (p70), IL-13, IL-17A, MCP-1 (MCAF), MIP-1β, TNF-α)
- SARS-COV2 antibody responses (spike and nucleocapsid)
- Vibrational Spectroscopy Analysis (Serum and Saliva samples at Lancaster University)

Power & Statistical analysis

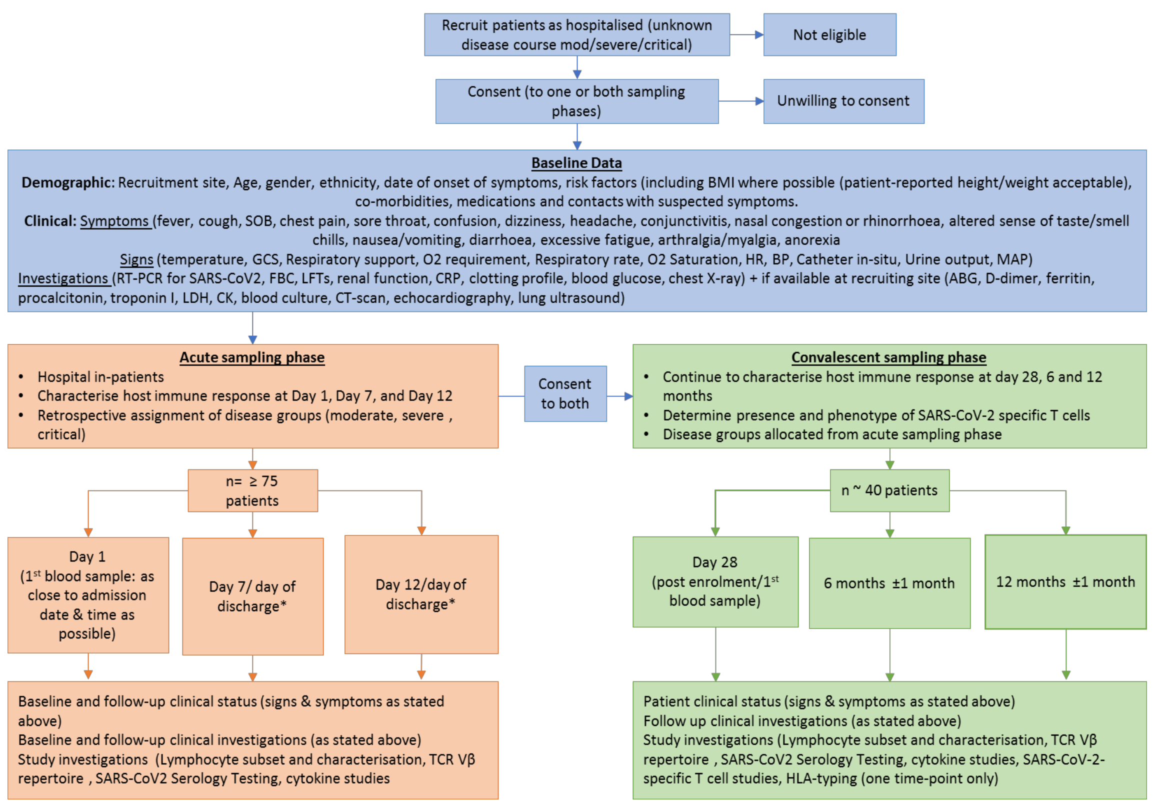
As this is a feasibility study there is no requirement for a power calculation. However, with at least 20-25 per group, no single individual has the potential to dominate the descriptive statistics for that group. Within a group, a percentage of 32% would have a 95% confidence interval of 14%-50%. A comparison between two groups would have 80% power to detect large differences in group percentage of around 40% as statically significant at the 5% level.

There will be a descriptive analysis of the data collected in the acute phase and convalescent phase and the disease severity groups within each of these phases. Rates of recruitment per month per centre will be described. Proportions falling into each disease severity group will be presented, where appropriate with 95% confidence intervals. Given the very large number of blood and serum measurements per patient there is the potential for a very large number of statistical comparisons, typically of measured data, using independent t-tests, ANOVA or non-parametric equivalents. Despite the great potential for type 1 errors due to multiple testing, these comparisons are not expected to have the power to provide definitive answers to research questions, but will instead provide estimates of differences between means in different groups and standard errors for planning a larger more definitive study.

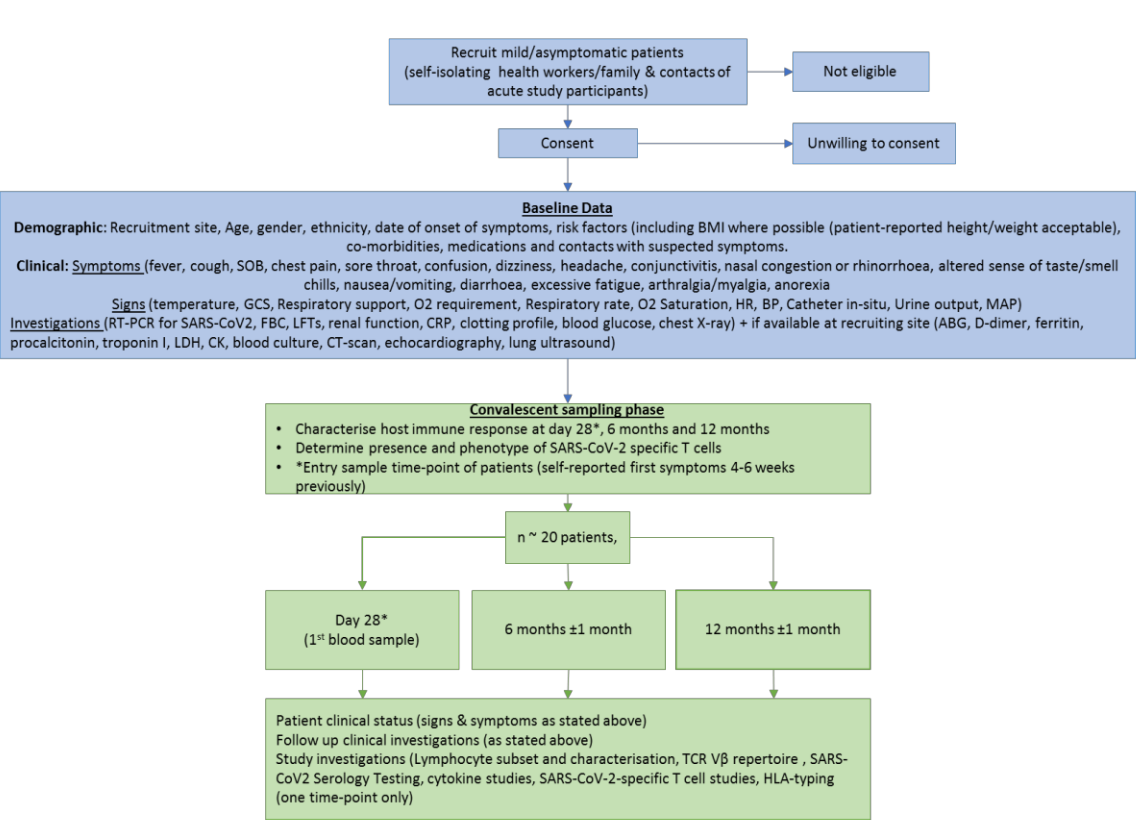
Appendix H

Protocol Flow Charts:

Hospitalised Patient Study Flow Chart

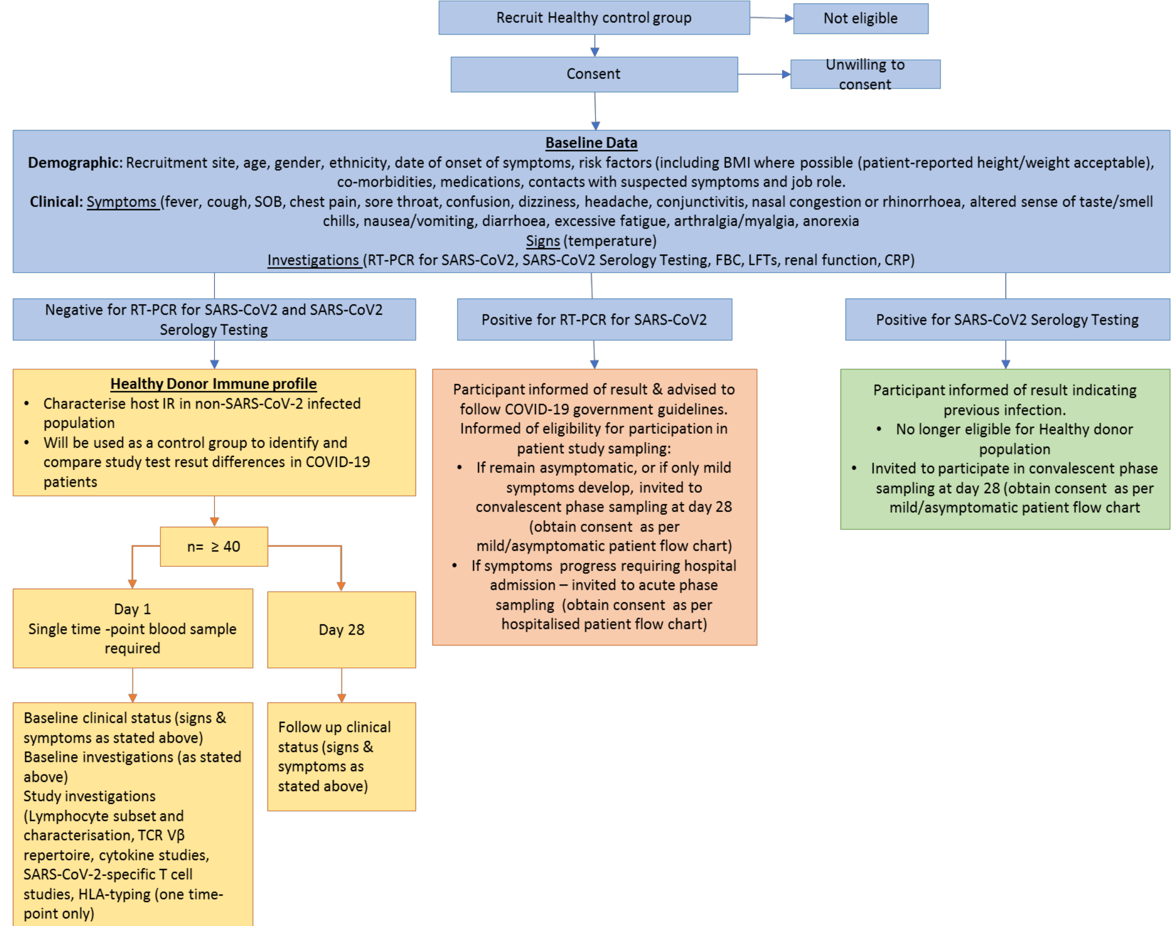


Mild/asymptomatic Patient Study Flow Chart



Appendix H

Healthy Donor Study Flow Chart



EXCOVIR (Exploring COVID-19 specific immune responses in acute and convalescent phases of infection)

What is the aim of the study?
The aim of this study is to further understand the immune response to SARS-CoV2 infection (the virus that causes COVID-19) and how it affects individuals in different ways.

The EXCOVIR project is an Immunology research study looking at the cellular responses found in patients during- and post-infection with coronavirus. Currently, much of the attention is directed to the ability of the immune system to produce antibodies however the evidence for long term protection and immunity is sparse. The EXCOVIR immunology research study will assess T cell responses to coronavirus and add knowledge to the overall immune following infections with SARS-CoV-2. T cells have a range of roles from direct killing of viruses to being a key regulator of the cytokine storm. The disease can have dramatically different effects on different people, with some patients having very mild or no symptoms, and others suffering from severe lung problems requiring hospital admission and intensive care treatment. There are currently no tests to predict how severe the disease will be in individual patients, or to determine whether patients develop protective immunity following recovery from COVID-19.

One of the outcomes from this study will be to better understand this process highlighting opportunities for the application of new treatments. The Immunology Department have developed a number of novel methods and will be one the first NHS laboratories to be exploring the clinical benefit from measuring T cell responses. These new tests will enable us to measure immune cells specific to SARS-CoV-2 in COVID-19 patients following recovery from infection. A better understanding of the role and longevity of T cell responses is essential and will be translated into clinical practice to improve patient care, especially once a vaccine is widely available. The project has gained ethical approval as a multi-centre trial and is supported by the NIHR Clinical Research Facility at Lancashire Teaching Hospitals.

Who can participate?
We aim to recruit 75 participants into the study as soon as they become unwell and hospitalised with COVID-19, doing swabs and blood tests at 3 time-points to look at early changes in the immune response. We will continue to follow them up, doing swabs and blood tests at 3 further time-points for 12 months to determine longer-term immunity.

We will advertise the study locally to NHS staff, friends and family to recruit 40 Healthy volunteers and 20 individuals who have tested positive for COVID-19 but had mild or no symptoms of infection. Healthy volunteers will have a single swab and blood test. Mild asymptomatic participants will have 3 swabs and blood tests taken, with the first at least 28 days after the infection.

How do I Volunteer?
For further information or to view a participant information sheet contact the research team via Kina.Bennett@lthtr.nhs.uk or Emma.Callery@lthtr.nhs.uk, or by ringing us on 01772 522031.

The Research group
The study led by Professor Anthony Rowbottom and Miss Emma Callery will be supported by clinicians Dr Kirsty Challen (PI), Dr Munavvar, Dr Vyas, Dr Laha, Dr Gudur, Dr Southworth, Dr Elbashir and Dr Anantharachagan, the Clinical Research Facility and Clinical Trials Unit at

UCLAN. The Immunology Department have developed a number of novel methods and will be one the first NHS laboratories to be exploring the clinical benefit from measuring T cell responses.

The clinical, scientific and laboratory elements of the study will be undertaken within the Immunology Department (Pathology) at Lancashire Teaching Hospitals NHS Foundation Trust. Our experience and expertise in the techniques used to routinely assess immune response within a number of different immune disorders drove us to rapidly put forward a research study proposal in light of the emergence and spread of COVID-19. The research group have published extensively and have demonstrated previous success in identifying disease-specific T cells in cancer, autoimmunity and viral infection.

The study is sponsored by LTHTR and will be supported by our NIHR Lancashire Clinical Research Facility (LCRF) which was developed to offer patients of Lancashire and South Cumbria access to early phase clinical research in order to offer options in their disease treatment and management. Along with the Centre for Research & Innovation, the NIHR LCRF has well developed processes, an experienced and motivated workforce and well established relationships both with clinical colleagues, support services and external collaborators to enable us to safely deliver this study. Additional support and co-ordination of the study is being provided by the Lancashire Clinical Trials Unit at the University of Central Lancashire (UCLAN).

Professor Anthony Rowbottom
Consultant Clinical Scientist, Associate Divisional Medical Director, Deputy Director For Research & Innovation
Lancashire Teaching Hospitals NHS Trust
Immunology Department
Royal Preston Hospital
Sharoe Green Lane
PR2 9HT

Email: approvals@hra.nhs.uk
HCRW.approvals@wales.nhs.uk

12 June 2020

Dear Professor Rowbottom

HRA and Health and Care Research Wales (HCRW) Approval Letter

Study title:	**Exploring COVID19 specific Immune Responses in acute and convalescent phases of infection (EXCOVIR)**
IRAS project ID:	**283457**
Protocol number:	**n/a**
REC reference:	**20/IEC08/0020**
Sponsor	**Lancashire Teaching Hospitals NHS Foundation Trust**

I am pleased to confirm that HRA and Health and Care Research Wales (HCRW) Approval has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications received. You should not expect to receive anything further relating to this application.

Please now work with participating NHS organisations to confirm capacity and capability, in line with the instructions provided in the "Information to support study set up" section towards the end of this letter.

How should I work with participating NHS/HSC organisations in Northern Ireland and Scotland?
HRA and HCRW Approval does not apply to NHS/HSC organisations within Northern Ireland and Scotland.

If you indicated in your IRAS form that you do have participating organisations in either of these devolved administrations, the final document set and the study wide governance report

(including this letter) have been sent to the coordinating centre of each participating nation. The relevant national coordinating function/s will contact you as appropriate.

Please see IRAS Help for information on working with NHS/HSC organisations in Northern Ireland and Scotland.

How should I work with participating non-NHS organisations?
HRA and HCRW Approval does not apply to non-NHS organisations. You should work with your non-NHS organisations to obtain local agreement in accordance with their procedures.

What are my notification responsibilities during the study?

The standard conditions document "*After Ethical Review – guidance for sponsors and investigators*", issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

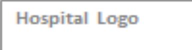
The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

Who should I contact for further information?
Please do not hesitate to contact me for assistance with this application. My contact details are below.

Your IRAS project ID is **283457**. Please quote this on all correspondence.

Yours sincerely,
Matt Rogerson
Approvals Specialist
Email: approvals@hra.nhs.uk

Copy to: *Dr Kina Bennett, Lancashire Teaching Hospitals NHS Foundation Trust*



EXCOVIR

EXploring COVID19 specific Immune Responses in acute and convalescent phases of infection

Participant Information Sheet

This hospital is taking part in a COVID-19 study and we would like to invite you to take part. Before you decide, it is important for you to understand why the research is being done and what taking part means for you. Please take time to read the following information carefully.

What is the purpose of the study?

When a new virus emerges, it is often difficult to predict how our immune systems will react to it, both immediately and over time. Understanding the immune response over a range of patients during different stages in their illness and recovery is key to controlling and fighting the virus. We will also be looking at samples from healthy volunteers to see if it is possible to identify if some people are more susceptible to severe illness associated with COVID-19.

This study is exploring;

- How the immune system responds to COVID-19 in the acute stage and in the convalescence period.
- Is it possible to identify which patients are at risk of developing severe disease and therefore be able to set priorities in care earlier?
- Is it possible to develop a new testing strategy that may determine how long patients retain immunity?

Why have I been invited?

You have been invited to participate as either, you currently have COVID-19, have tested positive for COVID-19 within the last 4 – 6 weeks or had the symptoms of COVID-19 (but not been tested), within the last 4 – 6 weeks.

Do I have to take part?

It is up to you to decide whether to take part or not. We will describe the study and go through this information sheet with you. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you have difficulty with writing or your eyesight, a witness will listen to your questions to the research team in relation to the information sheet and attest to your verbal consent to participate. The witness may be your relative (if they are able to be present) or a member of staff.

If you decide to take part, you are still free to withdraw at any time and without giving a reason and it will not affect the care you receive.

What will happen to me if I take part?

If you join the study whilst you are an inpatient, we will take saliva and a small blood sample (20 mls., equivalent to 4 teaspoons) on day of admission, day 7, day 12, day 28 (around 4-6 weeks), 6 months and 12 months. We will use these samples to look at such things as your antibody levels and your immune cells. Samples taken whilst an inpatient will usually be taken at the same time as routine tests ensuring no extra discomfort and if you are discharged before day 7, day 12 or day 28, we may adjust these timings slightly to be more convenient for you.

If you join the study as an outpatient, we will take samples as above on day 28 (4 - 6 weeks), 6 months and 12 months.

Hospital Logo

What are the possible disadvantages of taking part?
Your care will be exactly the same regardless of a decision to take part or not. There may be some inconvenience to returning for follow-up testing but no other foreseen disadvantages.

What are the possible benefits of taking part?
There is no direct benefit to you for taking part, but your data is vitally important and will contribute to understanding more about COVID-19 and how we might treat patients in the future.

Will my taking part in the study be kept confidential?
Your participation will be known to the research team and we will also inform your GP that you are taking part. All data will be kept confidential, accessible only to the research team and only to enable study procedures and follow-up.
The research is being conducted by experienced research staff at your recruiting NHS Trust and Lancashire Clinical Trials Unit. Lancashire Teaching Hospitals Trust is the Sponsor of this study and will act as the data controller. Your personal data will be kept for 5 years as per regulatory guidelines and then securely destroyed. Research data will be anonymised and given a unique identifier. At the end of the study the link between the two will be destroyed meaning all clinical and research data is permanently anonymised.

How will we use information about you?
We will need to use information from you and your medical records for this research project.
This information will include your initials, NHS number, name, contact details and details of your care. We will use this information to do the research or to check your records to make sure that the research is being done properly.
People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead.
We will keep all information about you safe and secure.

What are your choices about how your information is used?

- You can stop being part of the study at any time, without giving a reason, we will keep information about you that we already have, but please be assured it will be anonymised and only a study number will be used. It will not be possible to identify any one person from this data.
- If capacity is lost during the study, we would seek to obtain advice from a nominated consultee regarding your continued participation in the study. If your participation is stopped, data collected up to that point would be kept (as described above).
- If you agree to take part in this study, you will agree for your anonymised data and samples to be included within Lancashire Teaching Hospitals Trust Biobank and used for future research.
- The study is not designed to detect other illnesses, however if at any point we see evidence of other underlying disorders we will inform your GP in order for this to be followed up.

What will happen if I don't want to carry on with the study?
You are still free to withdraw at any time without giving a reason.

What will happen to the results of the research study?

Hospital Logo

The results of the study will be published as a full report and findings will be disseminated through publications/presentations and used to inform the evidence base for COVID-19 and future studies. All publications will use anonymised data and it will not be possible to identify any individual participant. Anonymised data will be used to contribute to other projects within this research theme and other relevant topics.

Who is organising and funding the research?
The study is collaboration between Lancashire Teaching Hospitals Trust and Lancashire Clinical Trials Unit.

Who has reviewed the study?
All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee and by the Health Research Authority, to protect your safety, rights, well-being and dignity. This study has been reviewed and given a favourable opinion by the Social Care Research Ethics Committee.

Who should I contact if I am unhappy with my treatment and wish to make a complaint?
If you have a specific concern or query about the research you can contact the study team on the details below. For a more independent contact, you can contact the Research Governance lead within the Centre for Health Research and Innovation at Lancashire Teaching Hospitals by contacting 01772 522031. You may also wish talk to the hospitals Patient Advice and Liaison Service (PALS) which provides support to patients, families and visitors. Hopefully, in most cases they will be able to sort out your concerns very quickly. However if you are not satisfied with the response that you receive you can make a complaint in writing. Please contact the Sponsor's Trust Customer Care department on 01772 522521 or email customer.care@lthtr.nhs.uk who can they will assist you with your complaint.

Where can you find out more about how your information is used?
You can find out more about how we use your information

- At www.hra.nhs.uk/information-about-patients/
- Our leaflet available from *www.hra.nhs.uk/patientdataandresearch*
- By asking one of the research team
- By sending an email to Kina.Bennett@lthtr.nhs.uk or emma.callery@lthtr.nhs.uk, or
- By ringing us on 01772 522031

Further information
For further information regarding the EXCOVIR study please contact the research team who provided you with this information sheet.

Local contact details:
Name:
Title:
Institution Name:
Address:
Telephone: