



AUTARQUIA ASSOCIADA À UNIVERSIDADE DE SÃO PAULO

**SKIN CANCER DIAGNOSIS USING INFRARED MICROSCOPY IMAGING AS A
MOLECULAR PATHOLOGY TOOL**

CÁSSIO APARECIDO LIMA

**Tese apresentada como parte dos
requisitos para obtenção do Grau de
Doutor em Ciências na Área
de Tecnologia Nuclear - Materiais**

**Orientadora:
Profa. Dra. Denise Maria Zezell**

**São Paulo
2019**

INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Autarquia associada à Universidade de São Paulo

**SKIN CANCER DIAGNOSIS USING INFRARED MICROSCOPY IMAGING AS A
MOLECULAR PATHOLOGY TOOL**

CÁSSIO APARECIDO LIMA

**Tese apresentada como parte dos
requisitos para obtenção do Grau de
Doutor em Ciências na Área
de Tecnologia Nuclear - Materiais**

**Orientadora:
Profa. Dra. Denise Maria Zezell**

Versão Corrigida
Versão Original disponível no IPEN

São Paulo
2019

Fonte de Financiamento: Fapesp (CEPOF 05/51689-2), CNPq (INCT: 465763/2014-6, PQ: 309902/2017-7, PhD: 141629/2015-0) e CAPES (PROCAD: 8888.068505/2014-1, PDSE: 88881.132771/2016-01).

Autorizo a reprodução e divulgação total ou parcial deste trabalho, para fins de estudo e pesquisa, desde que citada a fonte

Como citar:

LIMA, C. A. **SKIN CANCER DIAGNOSIS USING INFRARED MICROSPECTROSCOPY IMAGING AS A MOLECULAR PATHOLOGY TOOL**. 2019. 91 p. Tese (Doutorado em Tecnologia Nuclear), Instituto de Pesquisas Energéticas e Nucleares, IPEN-CNEN/SP, São Paulo. Disponível em: <www.teses.usp.br> (data de consulta no formato: dd/mm/aaaa)

Ficha catalográfica elaborada pelo Sistema de geração automática da Biblioteca IPEN/USP, com os dados fornecidos pelo(a) autor(a)

Lima, Cassio Aparecido
SKIN CANCER DIAGNOSIS USING INFRARED MICROSPECTROSCOPY
IMAGING AS A MOLECULAR PATHOLOGY TOOL / Cassio Aparecido
Lima; orientadora Denise Maria Zezell. -- São Paulo, 2019.
91 p.

Tese (Doutorado) - Programa de Pós-Graduação em Tecnologia Nuclear (Materiais) -- Instituto de Pesquisas Energéticas e Nucleares, São Paulo, 2019.

1. Microespectroscopia FTIR. 2. Câncer de pele. 3. Diagnóstico precoce. 4. Segmentação de imagens. 5. PCA-LDA. I. Zezell, Denise Maria, orient. II. Título.

Nome: Lima, Cássio Aparecido

Título: Skin cancer diagnosis using Infrared microspectroscopy imaging as a molecular pathology tool

Tese apresentada como parte dos requisitos para obtenção do Grau de Doutor em Ciências na Área de Tecnologia Nuclear - Materiais.

Aprovado em: ____/____/____.

Banca Examinadora

Prof. Dr.: _____

Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____

Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____

Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____

Instituição: _____

Julgamento: _____ Assinatura: _____

Aos meus pais (Lenino e Idelma) e irmã (Leidiane), que mesmo de longe e sem entender o que eu faço, me apoiaram nessa jornada e acreditaram em meus sonhos!

ACKNOWLEDGEMENTS

Ao **Instituto de Pesquisas Energéticas e Nucleares (IPEN)** e **Centro de Lasers e Aplicações (CLA)** pela infraestrutura concedida em todos estes anos de pós-graduação.

À Profa. Denise M. Zezell pela paciência, atenção e carinho nos últimos 6 anos como minha orientadora de mestrado e doutorado. Obrigado por sempre apoiar minhas ideias e incentivar meu crescimento pessoal e acadêmico.

À Profa. Luciana Correa do Departamento de Patologia da FOU SP pela contribuição fundamental no desenvolvimento deste trabalho.

Ao Prof. Hugh J. Byrne e demais funcionários da Technological University of Dublin (DIT) pela calorosa recepção e amplo apoio dentro e fora da universidade.

Aos meus amigos Daniel Flávio e Ulisses Marques pelo apoio incondicional durante todos os episódios dessa longa jornada.

Aos meus amigos e moradores da república Zapata (Uiran, Thiago, Leto, Paxo e Riboli) que sempre me acolheram em sua casa em diversos momentos.

Aos meus amigos e ex-moradores da república Sombreiro Verde (Victor, Mateus, Dimitri, Pedro, Zé e Fabio) pelos muitos momentos de alegria e descontração.

Aos atuais colegas do laboratório de Biofotônica e CLA (Gabriela, Daísa, Cacau, Ariane, Mariliza, Nathalia, Nielsen, Pedro, Amanda, Tânia, Armando, Fernanda, Camila) e aos muitos outros que passaram por aqui desde que cheguei há alguns anos.

À todos os funcionários do CLA e da Secretaria de Pós-Graduação do Programa de Tecnologia Nuclear.

Às agências de fomento que financiaram a execução desta pesquisa, Fapesp (CEPOF 05/51689-2), CNPq (INCT: 465763/2014-6, PQ: 309902/2017-7, PhD: 141629/2015-0) e CAPES (PROCAD: 8888.068505/2014-1, PDSE: 88881.132771/2016-01).

“It is truly an amazing feeling when you know that you have built
something that no one else ever has”

- Donna Strickland -

Nobel Prize in Physics 2018

RESUMO

LIMA, C. A. *Diagnóstico de câncer de pele usando imagens de microespectroscopia no infravermelho como ferramenta de patologia molecular*. 2019. 91 p. Tese (Doutorado em Tecnologia Nuclear), Instituto de Pesquisas Energéticas e Nucleares, IPEN-CNEN/SP, São Paulo.

Nas últimas décadas, a microespectroscopia de absorção no infravermelho por transformada de Fourier (FTIR) tem surgido como potencial ferramenta para complementar a Histopatologia no estudo e diagnóstico de doenças teciduais. Ao contrário do exame histológico, que se baseia na inspeção visual de amostras coradas visando avaliar as alterações morfológicas que as doenças ocasionam no tecido, o imageamento químico obtido pela técnica de FTIR baseia-se nas características bioquímicas da amostra sem o uso de colorações. Apesar da vasta literatura comprovando a eficácia da espectroscopia FTIR em detectar alterações biológicas causadas por doenças com altos níveis de sensibilidade e especificidade, a implementação do método na prática clínica tem sido relativamente lenta devido ao alto custo dos substratos transparentes no infravermelho que são necessários para aquisição de dados. Diante disso, o objetivo principal do presente trabalho é avaliar a capacidade diagnóstica de imagens hiperespectrais coletadas de amostras em lâminas de vidro como substratos alternativos para a espectroscopia FTIR. Camundongos Swiss foram submetidos a um protocolo de carcinogênese química, no qual lesões cutâneas cancerosas e não-cancerosas foram obtidas variando-se o tempo de exposição dos animais aos fatores carcinogênicos. Imagens hiperespectrais FTIR foram adquiridas no modo de transmissão na região do infravermelho médio a partir de amostras de tecido depositadas em substratos transparentes no infravermelho (fluoreto de cálcio - CaF₂) e vidro convencional. Na primeira fase de nosso estudo, os dados espectrais foram segmentados usando as técnicas estatísticas *k*-means (KMCA) e Análise Hierárquica de Clusters (HCA) como algoritmos de agrupamento para reconstruir as imagens hiperespectrais com o objetivo de avaliar a capacidade dos mapas de cores falsas em reproduzir as estruturas histológicas das amostras de tecido. As imagens foram segmentadas por cada técnica de agrupamento variando-se o substrato usado para colocar as amostras (CaF₂ ou vidro convencional) assim como os métodos de tratamento utilizados para pré-processamento dos dados. As regiões de impressão digital (1000-1800 cm⁻¹) e de altos números de onda (3100-4000 cm⁻¹) das imagens coletadas em CaF₂ foram usadas separadamente como dados de entrada

para a reconstrução das imagens, enquanto apenas a faixa de altos números de onda foi utilizada no caso de amostras colocadas em vidro. Ao fim do processo de segmentação os mapas de cores falsas obtidos foram comparados com a Histopatologia padrão a fim de avaliar a qualidade e consistência das imagens. Os resultados obtidos pela técnica de KMCA foram ligeiramente superiores com relação a HCA na identificação de pixels dos mapas morfo-químicos correspondentes às estruturas histológicas da amostra. No entanto, nossos achados indicam que a escolha do substrato, dados de entrada, métodos de pré-processamento e preparação de amostras têm mais influência nos resultados finais do que o algoritmo de agrupamento usado para reconstruir as imagens. Na segunda fase do nosso estudo, a Análise de Componentes Principais (PCA) foi empregada para comparar os dados do grupo saudável aos animais expostos aos produtos carcinogênicos por 8, 16 e 48 semanas a fim de avaliar as alterações bioquímicas induzidas pela carcinogênese química. O desempenho da classificação em cada comparação pareada foi calculado usando um teste de classificação binária baseado na Análise de Discriminante Linear associada à técnica de PCA (PC-LDA). O método obteve discriminação satisfatória (acima de 80%) comparando tecido saudável com as amostras que foram classificadas como papiloma (16 semanas) e carcinoma espinocelular invasivo (48 semanas) independentemente do substrato usado para colocar as amostras. A comparação de pele saudável com animais expostos aos fatores carcinogênicos por 8 semanas (livres de malignidade de acordo com as evidências clínicas e morfológicas) apresentou figuras de performance cujos valores variaram entre 35-78%, indicando que a habilidade da técnica de PC-LDA em classificar corretamente dados espectrais de lesões cancerosas e pré-cancerosas variam com o estágio da doença durante o processo de tumorigênese. Diante disso, como uma prova de conceito, demonstramos a viabilidade da espectroscopia FTIR na avaliação dos eventos biológicos desencadeados pelo câncer usando uma metodologia que não requer colorações e substratos caros, assim como não interrompe/altera o fluxo de trabalho atual do patologista. Este é um passo importante na implementação da tecnologia no ambiente clínico, uma vez que o método pode ser usado para complementar o processo de diagnóstico do câncer como uma alternativa não-subjetiva e que não requer procedimentos trabalhosos e demorados, nem sondas caras como biomarcadores.

Palavras-chave: Microespectroscopia FTIR; câncer de pele; diagnóstico precoce; segmentação de imagens; PCA-LDA; KMCA, HCA, RMieS-EMSC.

ABSTRACT

LIMA, C. A. *Skin cancer diagnosis using infrared microspectroscopy imaging as a molecular pathology tool*. 2019. 91 p. Thesis (Ph.D. in Nuclear Technology), Nuclear and Energy Research Institute, IPEN–CNEN/SP, São Paulo.

Over the past decades, Fourier Transform Infrared (FTIR) microspectroscopy has emerged as a potential candidate to complement Histopathology in the study and diagnosis of tissue diseases. Contrary to the histological examination, which relies on the morphological tissue alterations assessed by visual inspection of stained samples, FTIR chemical imaging is a rapid and label-free tool that provide simultaneously information about histological structures as well as the localisation and magnitude of basic molecular units that compose tissue sections (proteins, nucleic acids, lipids, and carbohydrates). Despite the many proof-of-concept studies demonstrating the effectiveness of FTIR spectroscopy in detecting biological disorders with high levels of sensitivity and specificity, translation into clinical practice has been relatively slow due to the substantial cost of infrared transparent substrates required to collect the images. Thus, the main objective of this research is to evaluate the diagnostic potential of infrared chemical images collected from samples placed on conventional histology glass slides as alternative substrates for FTIR spectroscopy. Swiss mice were submitted to a well-established chemical carcinogenesis protocol, in which cancerous and non-cancerous cutaneous lesions were obtained by varying the exposure time of the animals to carcinogenic factors. FTIR hyperspectral images were acquired in transmission mode over the mid-infrared region from tissue specimens placed on conventional infrared substrates (calcium fluoride - CaF_2) and glass slides. In the first phase of our study, spectral datasets were segmented using *k*-means (KMCA) and Hierarchical Cluster Analysis (HCA) as clustering algorithms to reconstruct the hyperspectral images aiming to evaluate the ability of the false-color maps in reproducing the histological structures of tissue specimens. The images were segmented by each clustering technique using several different combinations varying parameters including the substrate used to place the samples (CaF_2 or conventional glass) and the methods employed to preprocess the datasets. Fingerprint ($1000\text{-}1800\text{ cm}^{-1}$) and high wavenumber ($3100\text{-}4000\text{ cm}^{-1}$) regions from images collected on CaF_2 were separately used as input for image reconstruction and only the high wavenumber range was employed in the case of samples placed on

glass. All pseudocolor maps were compared to standard histopathology in order to evaluate the quality and consistency of images after segmentation. KMCA presented slightly superior ability in correctly assigning the pixels of morphochemical maps to the histological structures of the specimen, nevertheless, our findings indicate that the choice of the substrate, input data, preprocessing methods, and sample preparation have more influence in the final results than the clustering algorithm used to reconstruct the images. In the second phase of our study, Principal Component Analysis (PCA) was employed to compare datasets from healthy group to animals exposed to chemicals for 8, 16, and 48 weeks in order to evaluate the biochemical changes induced by chemical carcinogenesis. The performance of classification in each pairwise comparison was calculated using a binary classification test based on Linear Discriminant Analysis associated to PCA (PC-LDA). The method achieved satisfactory discrimination (over 80%) comparing healthy tissue to samples that were classified as papilloma (16 weeks) and invasive squamous cell carcinoma (48 weeks) regardless of the substrate used to place the samples. Statistical measurements obtained comparing healthy skin to animals exposed to carcinogenic factors for 8 weeks (free of malignancy based on the morphological and clinical evidence) ranged from 35-78%, indicating that the ability of PC-LDA in correctly classifying spectral data from cancerous and pre-cancerous lesions vary with the stage of the disease during the tumorigenesis process. Thus, as a proof-of-concept, we demonstrate the feasibility of FTIR spectroscopy in evaluating the biological events triggered by cancer using a label-free methodology that do not rely on expensive substrates and do not disrupt the pathologist workflow. This is a major step forward towards clinical application, since the method can be used to complement the diagnostic process of cancer as a non-subjective alternative that do not require laborious and time-consuming procedures nor expensive probes as biomarkers.

Keywords: FTIR microspectroscopy; skin cancer; early diagnosis; image segmentation; PCA-LDA; KMCA, HCA, RMieS-EMSC.

CONTENTS

| | |
|---|----|
| 1. INTRODUCTION | 13 |
| 1.1 Nonmelanoma Skin cancers | 13 |
| 1.2 Infrared Spectral Histopathology | 15 |
| 1.3 Objectives | 16 |
| 2 CHEMICAL CARCINOGENESIS | 18 |
| 2.1 Background | 18 |
| 2.2 Methodology | 19 |
| 2.2.1 Animal experiment..... | 19 |
| 2.2.2 Sample preparation..... | 20 |
| 2.3 Results | 21 |
| 2.3.1 Macroscopic evaluation..... | 21 |
| 2.3.2 Histopathological evaluation..... | 22 |
| 3 FTIR DATA ACQUISITION AND PREPROCESSING | 26 |
| 3.1 Fourier Transform Infrared (FTIR) Spectroscopy | 26 |
| 3.2 FTIR data acquisition | 26 |
| 3.2.1. Data collection..... | 26 |
| 3.3 Data preprocessing | 30 |
| 3.3.1 The RMieS-EMSC algorithm..... | 30 |
| 3.3.2 Removal of paraffin bands..... | 32 |
| 3.3.3 Spectra and Second derivatives..... | 34 |
| 3.4 Data analysis | 36 |
| 3.4.1 Principal Component Analysis (PCA)..... | 36 |
| 4. FTIR IMAGE SEGMENTATION | 39 |
| 4.1 Background | 39 |
| 4.2 Methodology | 40 |
| 4.2.1 Image segmentation..... | 40 |
| 4.2.2 Tissue specimen..... | 41 |

| | |
|--|----|
| 4.3 Results | 42 |
| 4.3.1 Reconstructing FTIR image collected from tissue placed on CaF2 substrate (fingerprint region)..... | 44 |
| 4.3.2 Reconstructing FTIR image collected from tissue placed on CaF2 substrate (high wavenumber region)..... | 46 |
| 4.3.3 Reconstructing FTIR image collected from tissue placed on glass slide (high wavenumber region)..... | 47 |
| 4.4 Discussion | 49 |
| 4.5 Conclusions | 51 |
| 5 STAGING NONMELANOMA SKIN CANCER USING INFRARED SPECTRAL HISTOPATHOLOGY | 52 |
| 5.1 Background | 52 |
| 5.2 Methodology | 53 |
| 5.2.1 Tissue samples..... | 53 |
| 5.2.2 FTIR data acquisition and spectral analysis..... | 53 |
| 5.3 Results | 56 |
| 5.3.1 Tissue specimens placed on CaF2 substrates (fingerprint spectral region)..... | 56 |
| 5.3.2 Tissue specimens placed on CaF2 substrates (high wavenumber spectral region)..... | 59 |
| 5.3.3 Tissue specimens placed on glass substrate (high wavenumber spectral region)..... | 63 |
| 5.4 Discussion | 66 |
| 5.5 Conclusions | 69 |
| BIBLIOGRAPHY | 70 |
| APPROVAL FOR COMMITTEE ON ANIMAL RESEARCH AND ETHICS | 80 |
| LIST OF PUBLICATIONS | 81 |
| MATLAB SCRIPTS USED TO ANALYZE DATA | 84 |

1 Introduction

1.1 Nonmelanoma skin cancers

Tumour or neoplasia (from greek “*neo*”, new and “*plassein*”, form) is the generic designation for the uncontrolled growth and proliferation of abnormal cells. The formation of tumours results from mutations on genes responsible by cell cycle regulation, apoptosis and DNA repair, which enable that mutated cells proliferate faster than healthy cells¹. According to their kinetics of growth and morphology, tumours are classified into benign or malignant. Benign or non-cancerous tumours generally have a slower growth rate and the cells are usually more differentiated (normal features). Cancer is the generic term of any malignant tumour and characterizes more than 100 diseases whose cells are able to rapidly reproduce and spread through the body (metastasis) in a way that affects the functioning of other organs^{2; 3}.

Cancer is the second leading cause of death worldwide, accounting for an estimated 9.6 million deaths in 2018. The International Agency for Research on Cancer (IARC) estimated 18.1 million new cases of cancer in 2018, in which non-melanoma skin cancers (NMSC) are the fifth most common type (Figure 1)⁴.

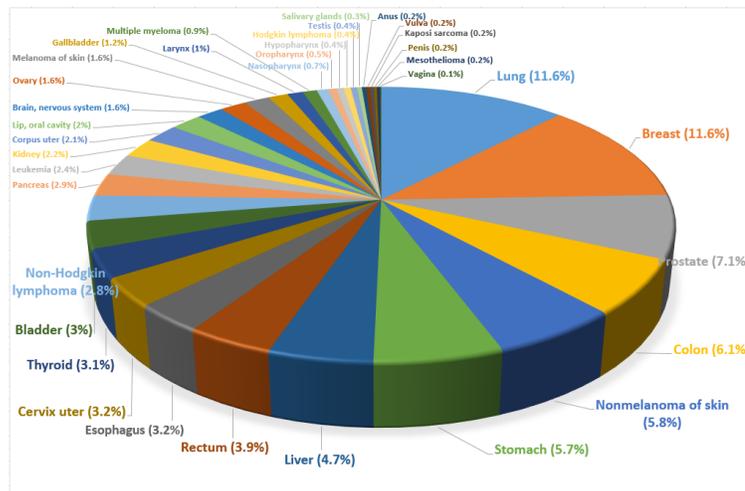


Figure 1. Global estimative for 36 types of cancers in 185 countries.

In Brazil, NMSC are the leading type of malignancy. The Brazilian National Institute of Cancer (INCA) estimated the occurrence of approximately 566.220 new cases of cancer in 2018,

in which 30% are related to NMSC (Figure 2)⁵. Basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) are the most common types of NMSC. BCC is responsible for 70% of diagnosed tumours and arise from cells located at the basal layer of epidermis. Although its high incidence, BCC has a slow growth, it is locally invasive and rarely presents metastasis^{6; 7}. SCC originates from keratinocytes of epithelium layer and corresponds to 25% of the NMSC cases recorded^{6; 7}. Contrary to BCC, SCC presents a destructive and metastatic growth pattern.

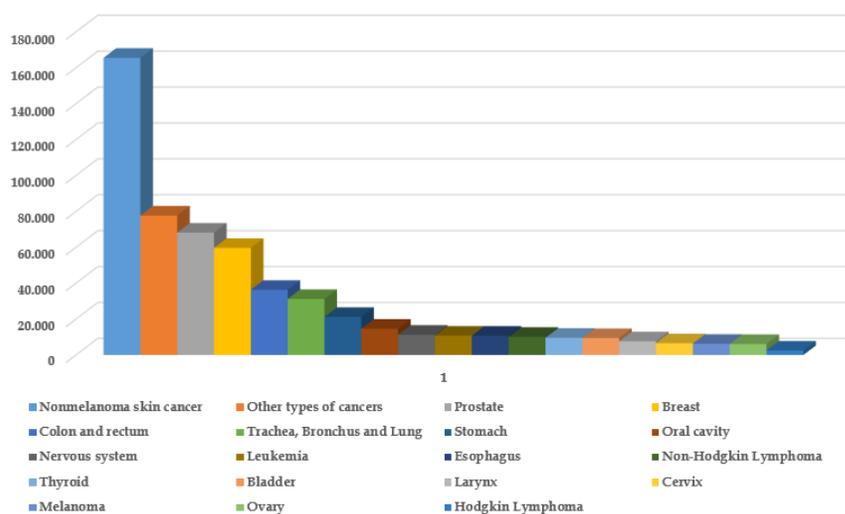


Figure 2. Estimative of cancer in Brazil (2018)

NMSC are neoplasms with good prognosis and high cure rates if early detected. In clinical practice, the diagnostic process begins with the visual inspection of a cutaneous lesion by a physician, which may extract the lesion via biopsy in order to obtain accurate diagnosis through histopathological examination^{8; 9}. This approach is also the current gold standard diagnosis for other types of cancer, which in general begin with genetic, biochemical, and structural changes in cells/tissues that lead to morphological and clinical manifestations (signs and symptoms) of the disease (Figure 3)¹⁰.

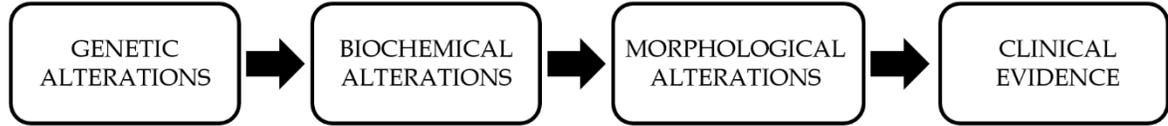


Figure 3. Biological events that occur during the cancer development.

Histopathology is based on the morphological changes and abnormalities that occur on cells and tissues during the transition from healthy to diseased state. The examination is performed by a pathologist that analyses abnormalities on tissue architecture, presence of cells at different locations, appearance of inflammatory cells, abnormalities on cell morphology, unusual number of mitotic figures, etc¹¹. The association of both clinical and histopathological examination has been extraordinarily successful as a first-line diagnostic tool in the last years, however, intra- and interobserver discrepancies regarding disease diagnosis are often obtained due to the subjectivity of the method that depends on the experience and training of both physician and pathologist. Thus, new technologies based on genetic and biochemical changes, so-called molecular pathology, have been investigated over the past years in order to complement the diagnostic process and reduce the reliance on morphological and clinical features¹¹.

The main focus of molecular pathology is identifying molecular alterations during the disease progression involving techniques developed in biochemistry, cell and molecular biology, proteomics and genetics such immunohistochemistry (IHC), immunofluorescence (IF), in situ hybridization (ISH), microRNA (miRNA) analysis, polymerase chain reaction (qPCR), DNA microarray, in situ RNA sequencing, DNA sequencing, etc¹². Despite the high sensitivity and specificity for accurate diagnosis, disease classification, treatment response and disease progression, the methods used in molecular pathology are laborious, time-consuming and rely on the use of expensive labeled antibodies/nucleic acid probes^{13; 14}.

1.2 Infrared Spectral Histopathology

During the last years, Raman and Infrared (IR) spectroscopy have been proposed as molecular pathology tools¹¹. These methods interrogate the chemical profile of a sample based on its chemical vibrations, enabling the evaluation of biochemical information using spectral

signatures instead of probing for the presence of targeted markers¹¹. Although Raman and IR spectroscopy are based on different modes of interaction between the incident light and the sample¹⁵, both technologies provide biochemical information about the analyzed sample monitoring the internal motion of atoms in molecules¹¹. Conventionally, IR spectroscopy is widely used to analyze large sample areas (tissue samples) due to the relatively short time required to perform the measurements, whereas Raman spectroscopy is preferred for single cell analysis¹⁵.

In the view of spectroscopic methods, cells and tissues are complex samples. The collected spectrum is the superposition of all spectra of the individual molecular units that compose the specimen (nucleic acids, carbohydrates, lipids, and proteins). Thus, the analysis of biological samples is much more complex than the simple explanation of the features of single component vibrations, in which the effect of each structure may interact with the others and result in the amplification or reduction of a specific signature¹⁶. In light of this, the currently challenge is to decode the complex spectral signatures retained by spectral data and use this information for medical diagnosis^{11; 17}. Results from a number of laboratories worldwide indicate that IR spectroscopy has been fully succeeded as a non-subjective diagnostic tool for cancerous (Cervix¹⁷, lung^{18; 19; 20}, breast^{21; 22; 23}, skin^{16; 24; 25}, brain^{26; 27}, prostate^{28; 29; 30}, colon^{31; 32}, thyroid³³, etc.) and non-cancerous diseases (Alzheimer's³⁴, diabetes^{35; 36}, etc.) as well as to monitor the effects of drug treatments on disease progression/regression^{13; 17; 37; 38; 39; 40}. Despite the many proof-of-concept studies demonstrating the effectiveness of IR spectroscopy in detecting biological disorders with high levels of sensitivity and specificity, the techniques still remains in the research environment and translation into clinical practice has been relatively slow mainly due to the substantial cost of infrared transparent substrates required to collect chemical images.

1.3 Objectives

The main objective of this research is to evaluate the diagnostic ability of infrared chemical images collected from samples placed on conventional glass slides as alternative substrate for FTIR spectroscopy. Specific objectives include:

- Evaluating the influence of sample preparation, clustering algorithms, substrate to place the samples, and computational preprocessing methodologies to segment FTIR hyperspectral images for histopathologic recognition;
- Using information retained by spectra in discriminating healthy tissue from cancerous and pre-cancerous specimens obtained from the skin of animals exposed to chemical carcinogenesis;

2 Chemical carcinogenesis

2.1 Background

Using animal models for the study of skin carcinogenesis have been widely employed in preclinical and research tests due to their molecular similarities with the human tumourigenesis. Different methodologies can be used to induce cutaneous cancers including ultraviolet radiation (UVR), tumour cells transplantation and chemical carcinogenesis. Skin cancers induced via UVR provides a more physiologically relevant method for assessing the mechanisms of cutaneous tumorigenesis due to the analogy with the carcinogenic agents that cause the disease in humans. However, the biological events that lead to cancer apply only to tumours arising on skin, and thus do not describe the mechanisms in other types of cancers. Furthermore, UV radiation is more effective in hairless mice, which complicates the ability to evaluate genetically manipulated mouse models produced in common strain backgrounds⁴¹. Injection of human cancer cell lines into mice has also been extensively used in several studies. However, such cell lines undergo multiple adaptations to cell culture, which compromises the representativeness of the induced cancer compared to the disease that arise spontaneously in a patient⁴².

Chemical carcinogenesis in mouse skin has been used for several decades and is based on the administration of carcinogenic agents to the skin of animals. The chemical carcinogens may be classified as initiators and promoters according to their role in the process of carcinogenesis. Initiating agents are mutagenic chemicals that induce mutations in the key genes of epidermal keratinocytes⁴³. The damaged DNA must be replicated to occur the initiation, therefore cells exposed to the carcinogen must undergo at least one cycle of proliferation¹⁰. This event is irreversible; however, no visible tumours will appear until ‘promoted’ by the application of a promoting agent⁴³, which induce biochemical rather than genetic alterations. In the same way as the initiators, promoting agents are not able of inducing tumours when applied alone.

The protocols used in the chemical carcinogenesis can be classified into ‘complete’ or ‘two-stage’. On the complete protocol, a single high dose of a carcinogenic factor (or repeated applications of a lower dose) that acts as both initiating and promoting components is applied in the mice skin in order to develop the tumours⁴³. On the two-stage protocol, animals are first submitted to initiators followed by additional treatments with promoting agents. The

interpretation of the results obtained by the complete protocol may be complicated due to the inability to distinguish events associated to the tumour initiation and promotion stages. On the other hand, in two-stage protocol these phases can be distinctly separated both operationally and mechanistically⁴³.

All chemical initiators are electrophiles (electron-deficient molecules) that can react with nucleophilic (electron-rich molecules) sites in cells¹⁰. The polycyclic aromatic hydrocarbon 7,12-dimethyl-benzanthracene (DMBA) is one of the most common agents used as initiator in two-stage skin carcinogenesis. DMBA results in mutations to regulatory genes in epidermal keratinocytes by interacting with DNA and inducing an A→T transversion in codon 61 of the *Hras1* gene. This mutation can be observed in the epidermis 1 week after DMBA application and results in the production of permanently activated Ras proteins, which regulate pathways responsible by controlling cell proliferation, differentiation, and survival⁴³. Currently, 12-O-tetradecanoylphorbol-13-acetate (TPA) is one of the most used tumour promoting agents. Although the exact mechanisms whereby TPA induces skin tumours are not fully understood, it is known that TPA stimulate cell signaling, increase production of growth factors and generate oxidative stress and inflammation that promote tumour growth⁴¹. The result of the promotion stage is the development of clonal outgrowths of skin referred as papillomas, which may be converted into carcinomas according the genetic background of mice and the doses of carcinogenic factors^{41; 43}. During the conversion, additional genetic alterations may occur independent of continued treatment with tumor promoting agents including trisomies of chromosomes 6 and 7, as well as mutations in Trp53⁴³.

The main goal of the present study is to evaluate the ability of FTIR spectroscopy in discriminating cancerous and non-cancerous tissue obtained via two-stage chemical carcinogenesis protocol. This section is focused in describing the animal experiment, sample preparation, macroscopic and microscopic results obtained by the protocol.

2.2 Methodology

2.2.1. Animal experiment

After approval by the Committee on Animal Research and Ethics (164/15-CEUA-IPEN/SP) (attached), 77 female Swiss mice were submitted to a well-established two-stage

chemical carcinogenesis protocol⁴³. In the first phase (induction), the shaved backs of the mice were submitted to a single dose of DMBA (50 µg diluted in 200 µl of acetone). One week later, the animals underwent twice a week to applications of TPA (5 µg diluted in 200 µL acetone). Control group (healthy) only received topical applications of acetone during the experiment and the remaining animals were exposed to TPA for 8, 16 and 48 weeks as described in Table 1.

Table 1. Experimental mice groups

| Group | Period of exposure to promoting agent (weeks) |
|------------------|--|
| G1 (n=12) | - |
| G2 (n=12) | 8 |
| G3 (n=12) | 16 |
| G4 (n=41) | 48 |

In the first phase of the experiment, animals were submitted to general anesthesia using ketamine (0.001 ml/g) associated to xylazine (0.0005 ml/g) intraperitoneally. The carcinogens were applied in a laminar flow to ensure the safety of the manipulator. Animals were euthanized at the end of the experiment using overdose of anesthetics and the specimens were extracted with a scalpel.

2.2.2. Sample preparation

Specimens extracted from animals were formalin-fixed and paraffin-embedded (FFPE). Two adjacent histological slices (5 and 10 µm thickness each) were obtained using a microtome. The first slice (5 µm thickness) was prepared in a conventional glass slide, dewaxed and stained with hematoxylin and eosin (H&E) according to the international standards and coverslipped. The second slice (10 µm thickness) was placed on calcium fluoride (CaF₂) substrates (Crystran Ltda., Poole, Dorset, UK). No chemical dewaxing or staining was performed on the second slice and it was not coverslipped.

2.3 Results

2.3.1 Macroscopic evaluation

The first tumours appeared between the 12th and 14th weeks of application of the promoting agent. Initially, such lesions appeared as a keratinized nodule (arrow in Figure 4.B) that subsequently evolved into highly vascularized wart-like lesions (Figure 4.C). The formation of papillomatous lesions was heterogeneous in the animals, so that it was possible to observe the formation of only one or multiple papillomas in the same animal. After 48 weeks applying TPA, the animals presented ulcerated lesions in the basal part and extensive area of necrotic tissue on the surface (Figure 4.D).

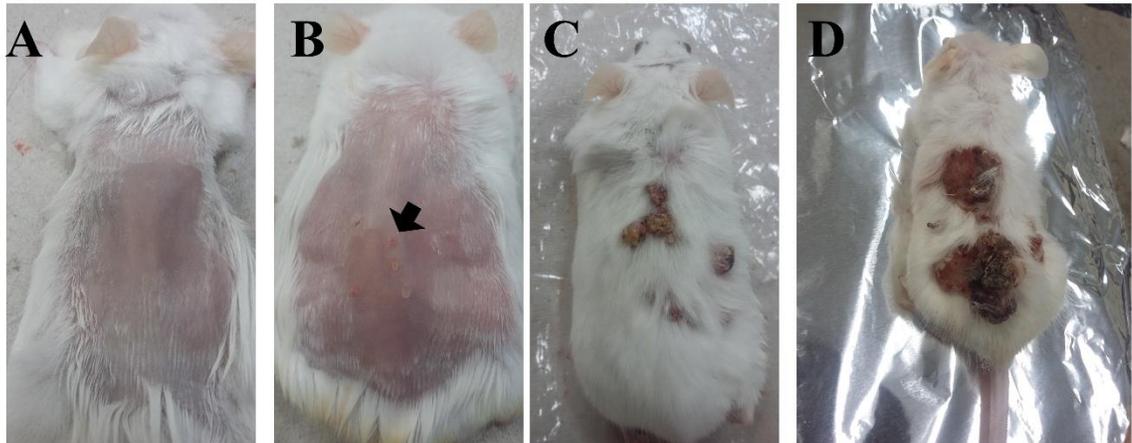


Figure 4. Panels describing the appearance and progression of cutaneous lesions typically obtained in chemical carcinogenesis. A: Animal with no evidence of lesions after 8 weeks of application of the carcinogenic factors. B: Keratinized nodule (arrow) typically obtained after 12-14 weeks. C: Highly vascularized wart-like lesions obtained after 32 weeks. D: Animal presenting ulcerated lesions in the basis part and extensive area of necrosis on the surface (48 weeks).

The longtime of experiment added to the debilitated state induced by the carcinogenesis protocol resulted in a significant loss of the initial number of animals. Figure 5.A shows the survival rate of animals during the experiment. It is possible to observe a reduction of approximately 50% in the number of animals exposed to the experiment during 48 weeks (G1 and G4), 25% of animals from G2 and 35% for G3 group.

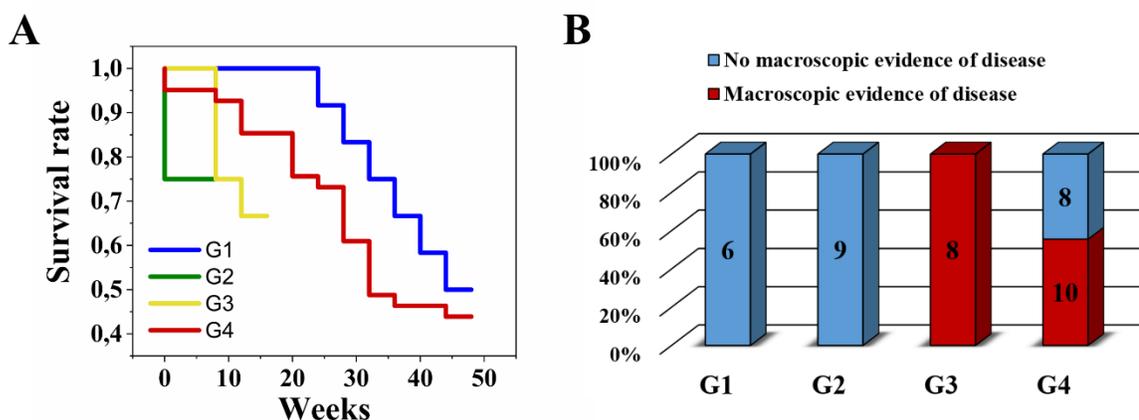


Figure 5. A: Survival rate of animals during the experiment. B: Animals that developed the disease at the end of chemical carcinogenesis.

The formation of papillomas and their progression into squamous cell carcinoma occurs stochastically and depends on several factors including genetic characteristics of the lineage used as animal model, doses of chemicals and period that mice were exposed to carcinogenic agents⁴¹. Thus, some animals may not develop the disease. Figure 5.B shows the number of animals that survived to the experiment as well as the presence of macroscopic features indicating the development of the disease. Mice from G1 (healthy skin) were not exposed to carcinogenic agents, therefore presented no clinical signs of the disease. Skin lesions were not observed in animals of group G2, suggesting that at this moment, if the disease is under development, it is still in the primary stages (genetic, biochemical or morphological changes). All mice from G3 group presented visible tumour nodules similar to those depicted by arrow in Figure 4.B. For G4 group, 45% of the animals that survived to the experiment did not develop the disease and 55% presented clinical evidence indicating the formation of neoplasms.

2.3.2 Histopathological evaluation

Histopathological evaluation is the gold standard method for the classification of cutaneous tumours, therefore the histological characteristics of the specimens were preserved and evaluated by a pathologist. Tumours induced on animals of G3 and G4 groups showed high inter- and intra-animal variability for both macroscopic and microscopic aspects. Thus, each

tumour (specimen) was considered as a sample unit instead of the number of animals per group (Table 2).

Table 2. Number of sample units (specimens) obtained per group

| Group | Number of mice that survived to the experiment | Number of sample units (specimens) collected from animals |
|-------|--|---|
| G1 | 6 | 6 |
| G2 | 9 | 9 |
| G3 | 8 | 12 |
| G4 | 10 | 25 |

The specimens were classified into healthy skin, papilloma, superficial and invasive SCC according to their histological aspects. Figure 6 shows the frequency of the histological patterns observed in the specimens collected from mice groups.

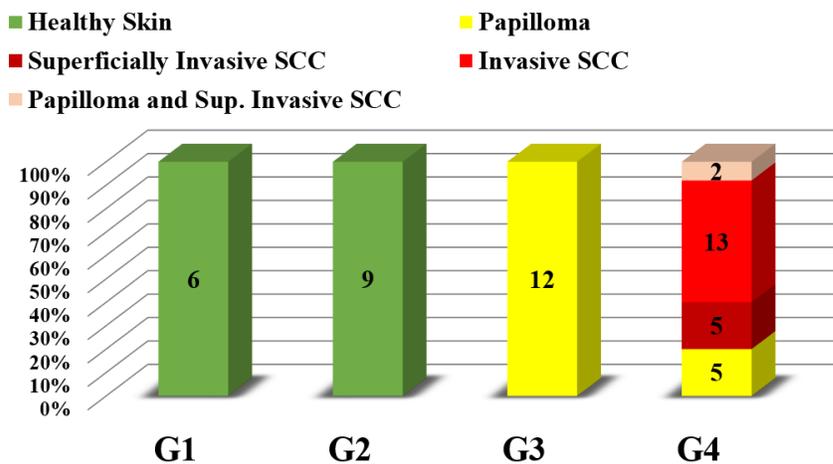


Figure 6. Frequency of the histological patterns observed in the specimens collected from mice groups.

Microscopic aspects of healthy tissue were evidenced in the specimens collected from animals of G1 (Figure 7.A) and G2. Both groups presented intact epidermis composed by 3 or 4 layers of keratinocytes covered by a thin layer of keratin (stratum corneum). The dermis presented dense connective tissue with hair follicles and organized hypodermis. Inflammation

or changes in vascularization were not observed. Some specimens of G2 presented areas with histological aspects indicating microscopic alterations induced by DMBA/TPA (Figure 7.B).

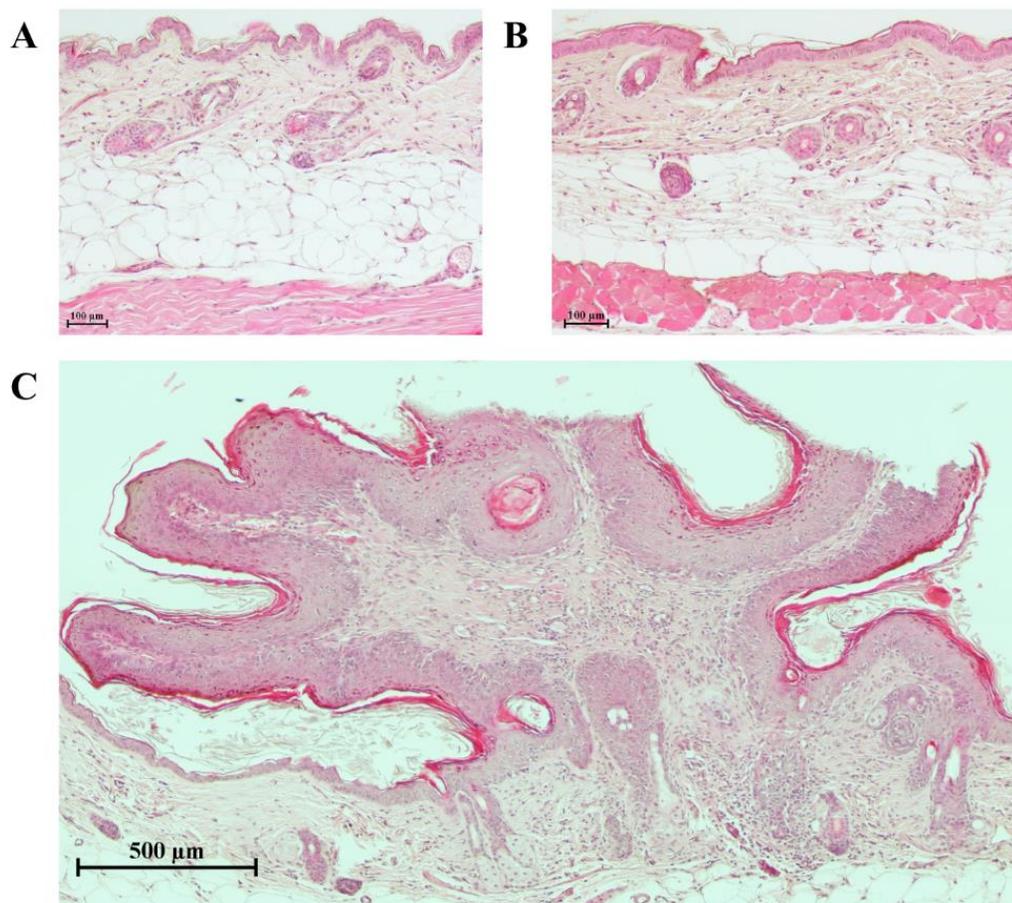


Figure 7. Representative photomicrographs of histological aspects observed in the histological specimens of G1 (A), G2 (B) and G3(C).

Papillomatous lesions were observed in 100% of the histological sections of animals from G3 and in 5 specimens from G4 group (Figure 7.C). The lesions presented intense proliferation of keratinocytes (epithelial hyperplasia) in exophytic pattern covered by a thick keratin layer (hyperkeratinization), moderate connective tissue with inflammation underlying the papillomatous lesions and, in some cases, intense cellular atypia.

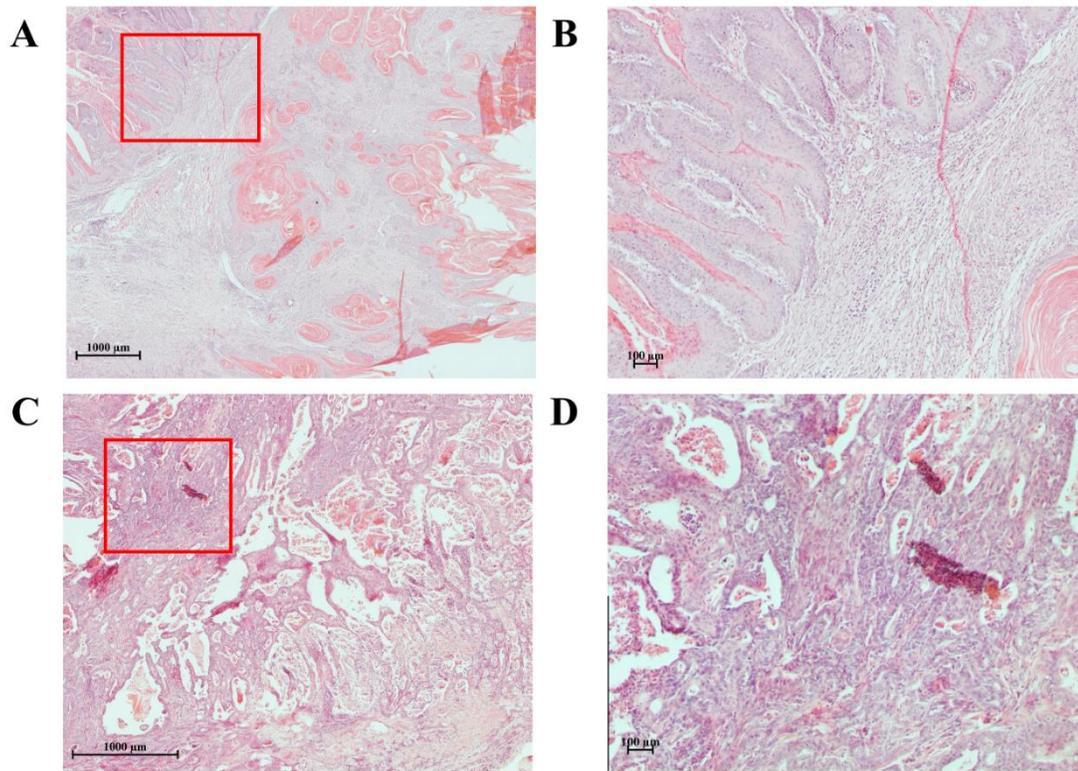


Figure 8. Histological aspects of specimens classified as superficial (A) and invasive SCC (C). B and D represent higher magnification of regions delimited by red rectangles in A and C.

Figure 8.A shows a photomicrograph representing the lesions observed in 20% of specimens from G4 that were classified as superficial SCC. The lesions presented hyperkeratinization, poor connective tissue with intense inflammation and epithelial hyperplasia with superficial invasion into the dermis. Ulcerated connective tissue and areas of necrosis were observed in some cases. Specimens of G4 that were classified as invasive SCC (52% of sample units) presented intense epithelial hyperplasia with a solid invasiveness pattern, poor connective tissue with intense inflammation, extensive and moderate necrotic areas and intense atypia (Figure 8.C). Some sample units collected from animals exposed to carcinogenic factors for 48 weeks presented histological aspects of both papillomas and superficial SCC in the same histological slide.

3 FTIR data acquisition and preprocessing

3.1 Fourier Transform Infrared (FTIR) Spectroscopy

Spectroscopy is defined as the set of analytical techniques that are able to obtain physical and/or chemical information based on the interaction of electromagnetic radiation with a sample⁴⁴. In Fourier transform Infrared (FTIR) spectroscopy, a multi-frequency beam of mid-infrared radiation (wavelengths range from 1 to 10 μm) is focused onto the material and the profile of the transmitted/reflected radiation is collected and compared against a non-absorbing background^{45; 46}. The photons of infrared radiation are absorbed by the sample based on the interaction of the incident electromagnetic field with the molecular electric dipoles created between atoms that are vibrating in the sample. The frequency of this vibration is characteristic and unique for different functional groups relating to unique frequencies of light being absorbed. By this way, it is possible to identify which molecules or chemical bonds are present in the analyzed sample⁴⁶. The absorbed frequencies are usually presented as an absorption spectrum, which relates the frequencies (in units of wavenumbers) with the intensity of absorption (in arbitrary units)⁴⁵. Due to its ability to identify macromolecules of major biochemical importance, FTIR spectroscopy has become particularly interesting in the study of biological tissues^{47; 48}.

3.2 FTIR data acquisition

3.2.1. Data collection

The most used experimental setups for acquiring FTIR spectral data include transmission, transfection and Attenuated Total Reflection (ATR). Each sampling mode offers convenience for some samples and challenges for others⁴⁷.

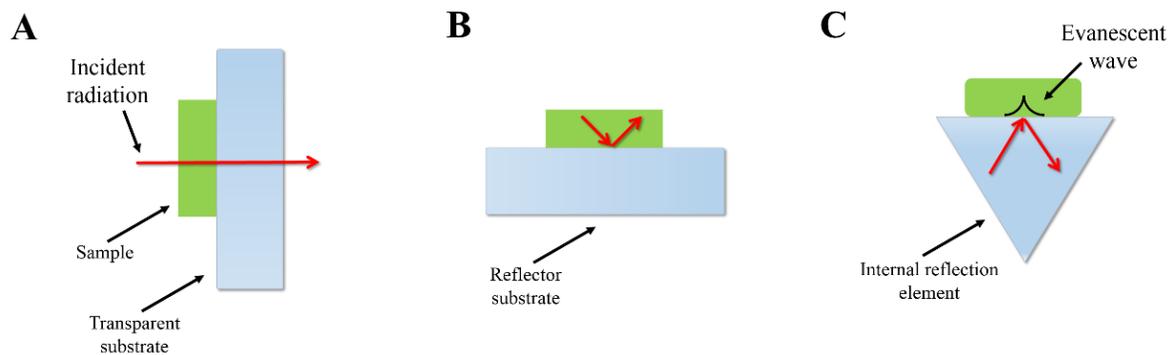


Figure 9. Schematic representation of the three main sampling modes for FTIR spectroscopy: transmission (A), Transflectance (B), and Attenuated total reflection (C).

In ATR sampling mode, the sample is placed on an internal reflection element (IRE) with a refractive index higher than the analyzed material, inducing total internal reflection of the infrared beam that reaches the sample by passing through the reflection element. The radiation is attenuated and penetrates into the sample as an evanescent wave (Figure 9.C)⁴⁹. The IRE commonly used are made by diamond, ZnSe, Germanium or Silicon⁴⁷. The penetration depth of the evanescent wave (d_p) into the sample vary according the incident wavelength (λ), angle of incidence of radiation (θ) and the refractive indices of sample (n_2) and IRE (n_1) as shown in equation 1^{49; 50}.

$$d_p = \frac{\lambda / n_1}{2\pi \sqrt{\left[\text{sen}\theta - \left(\frac{n_1}{n_2} \right)^2 \right]}} \quad (1)$$

FTIR measurements acquired on ATR mode require little or no sample preparation method and do not present spectral distortions as the commonly obtained in transmission and transflection⁵¹. The main disadvantage of using ATR is that sample may be destructed during the measurements due to the pressure required to achieve good contact between the sample and the IRE⁴⁷.

In transflection mode, the radiation interacts with the sample and it is reflected back by a reflector substrate (Figure 9.B). The most common material used for this purpose is glass coated with a layer of Ag/SnO₂, which are known as low-emissivity (low-e) slides. These substrates

are cheaper than those required in transmission, but are still expensive in comparison to the price of the conventional glass slides employed in standard histopathology. Recently, aluminum foil has been proposed as a cheap and versatile substrate to collect spectral data in transflection mode and has a great potential to promote a transition towards accessible substrates that can be readily implemented in either research or clinical practice⁵². The main disadvantage of acquiring FTIR spectra in transflection mode is the variety of physical effects that occur simultaneously to absorption that may give rise to undesired spectral distortions that must be corrected computationally prior any analysis^{47; 53; 54}. Besides the common artifacts induced by light scattering, an electric-field standing wave may be induced in the surface of the substrate for samples with thickness less than $2.3 \mu\text{m}$ ^{55; 56; 57; 58}.

In transmission mode, spectra are collected considering the non-absorbed radiation that pass through the sample (transmitted radiation). For this, the samples are placed in mid-infrared transparent substrates (Figure 9.A) such potassium bromide (KBr), barium fluoride (BaF_2), calcium fluoride (CaF_2) or zinc selenide (ZnSe). In the case of biofluids, cells and soft tissue samples, KBr is unsuitable due to its water solubility. Contrary to the obtained in transflection, standing-wave artifacts are not observed in spectra collected in transmission. However, spectral distortions induced by scattering are commonly observed, therefore must be corrected before spectral analysis.

The main disadvantage in acquiring spectra on transmission mode is the high cost of the substrates required to place the samples. CaF_2 slides are by far the most used substrate to acquire FTIR images by light transmission. Despite the satisfactory results obtained, translation into clinical practice has been relatively slow due to the substantial cost of mid-infrared transparent substrates. In addition, these substrates are fragile and require that sample preparation has to be carried out manually instead of the conventional automated equipment used in the clinical practice. Bassan *et al* demonstrated that standard glass histology slides have the potential to be used for infrared chemical imaging of tissue²³. Figure 10 shows the transmittance spectra of CaF_2 and glass substrates, in which it is possible to observe that spectral region over the range $1000\text{-}1800 \text{ cm}^{-1}$ is transparent to CaF_2 but it is fully absorbed by glass. This region is commonly referred as fingerprint due to the high number of vibrational modes associated to the main

molecular units that compose biological samples (lipids, nucleic acids, proteins, and carbohydrates).

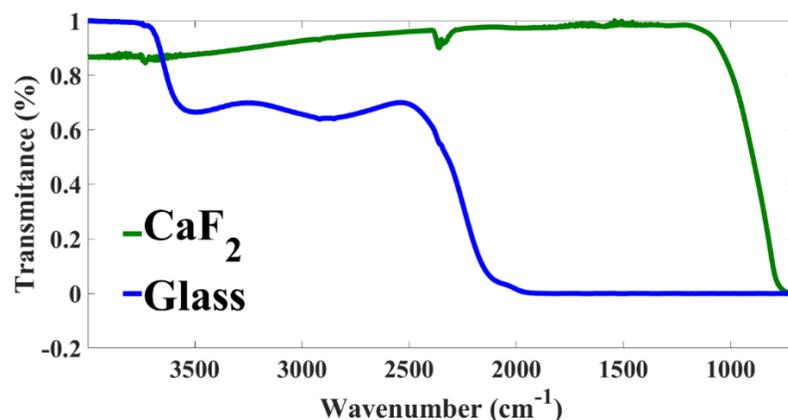


Figure 10. Transmittance spectra of conventional glass slides and CaF_2 substrates over the mid-infrared spectral region.

Figure 11 shows spectra collected in transmission mode from healthy cutaneous tissue placed on CaF_2 and glass substrates. Despite the opacity of glass in the fingerprint region, it does have a narrow transmission window covering the regions occurring at $2500\text{--}4000\text{ cm}^{-1}$ that is commonly referred as high wavenumber region.

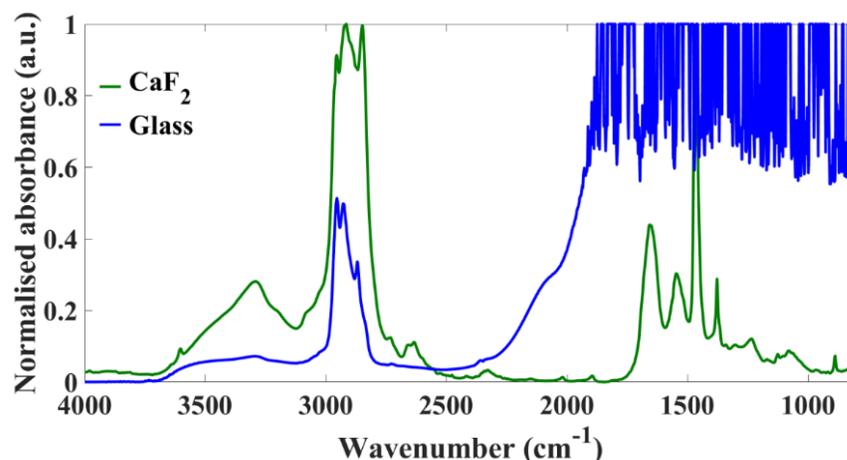


Figure 11. Absorbance spectra of healthy cutaneous tissue placed on glass slides and CaF_2 substrates over the mid-infrared spectral region.

The fingerprint region have been largely employed by different studies evaluating the biomedical applications of FTIR spectroscopy and few studies have investigated the diagnostic

potential of high wavenumber region. In light of this, the present study evaluated the ability of both spectral ranges to discriminate healthy from diseased tissue (Chapter 5) as well as to segment FTIR images in order to identify the histological structures of the skin (Chapter 4).

Samples were placed onto glass and CaF_2 as described in section 2.2.2. FTIR hyperspectral images were collected on transmission mode using a Spotlight 400N FTIR imaging system (Perkin Elmer, Waltham, MA, USA) equipped with an AutoImage microscope system operating with a $\times 40$ Cassegrain objective and a liquid nitrogen cooled mercury cadmium telluride (MCT-A) line detector. Spectral data were recorded with a pixel size of $6.25 \times 6.25 \mu\text{m}$ at a spectral resolution of 4 cm^{-1} and 8 scans per pixel. Background measurements were acquired on a region free of tissue with 120 scans per pixel. Spectral measurements were acquired over the range $750\text{-}4000 \text{ cm}^{-1}$ for samples placed on CaF_2 and $2500\text{-}4000 \text{ cm}^{-1}$ for samples on glass slides. The data collection was performed at the Technological University of Dublin (TUD) in collaboration with Prof. Hugh J. Byrne.

3.3 Data preprocessing

3.3.1 The RMieS-EMSC algorithm

FTIR spectra collected in transmission and transfection modes are subjected to undesired spectral contributions non-related to the phenomenon of light absorption. Among these, Mie scattering has been identified as the main responsible for inducing spectral distortions, which may be presented as resonant or non-resonant.

In non-resonant Mie-type scattering (or classical), the most prominent spectral distortion is a broad sinusoidal oscillation in the baseline upon which the absorption bands are superimposed. From theory, it is known that the interaction of electromagnetic radiation with a dielectric sphere of dimensions close to that of the wavelength of light will induce strong scattering⁵⁹. Thus, the origin of classical Mie scattering relates to the similar length-scale of the cellular components from biological samples to the wavelengths used in FTIR spectroscopy ($1\text{-}10 \mu\text{m}$)⁵⁷. In Figure 12.A, the baseline alterations induced by classical Mie scattering are depicted by dashed red line.

In resonant Mie scattering (when there is simultaneous scattering and absorption), the most obvious effect is the so-called ‘anomalous dispersion’ or ‘dispersion artefact’, which has been recently demonstrated to have origin in contributions of the real component of the refractive index⁵⁷. The artifact is presented in the spectrum as a dip in the absorption profile on the high wavenumber side of the band peaking between 1600-1700 cm^{-1} (amide I) (red arrow in Figure 12.A)⁵³. Still as result of the artefact, alterations in intensity and position of the amide I band can also be observed (in Figure 12.A, the Amide I is incorrectly peaking at 1644 cm^{-1}). The non-correction of dispersion artifacts can lead to erroneous interpretation and conclusions of the biochemical features retained by spectral data. Thus, the spectral datasets acquired in transmission must be corrected prior any analysis⁵⁹.

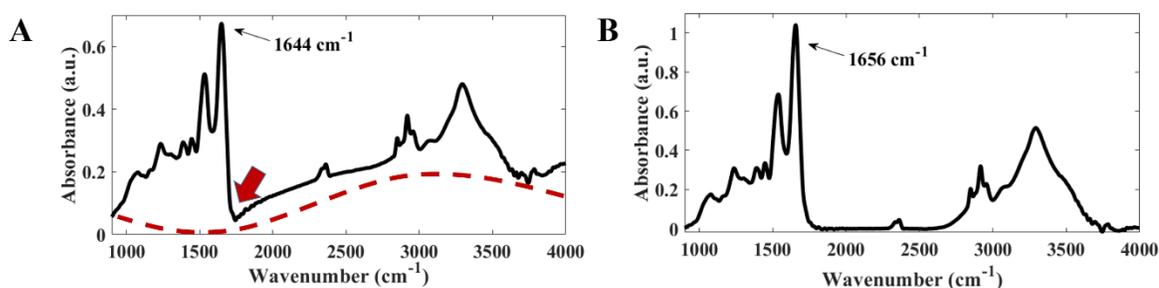


Figure 12. **A:** Raw FTIR spectrum with spectral distortions induced by Mie scattering. Sinusoidal oscillations in the baseline are represented by dashed red line, whereas the dip in the high wavenumber side of Amide I (1644 cm^{-1} , shifted of its original position) is highlighted by the red arrow. **B:** FTIR spectrum free of scattering distortions after correction using RMieS-EMSC algorithm.

The so-called “Resonant Mie Scattering - Extended Multiplicative Scatter Correction (RMieS-EMSC)” protocol has been fully succeeded in correcting spectral distortions induced by Mie scattering^{53; 54; 57; 59; 60}. Figure 12.B shows the same spectrum of panel A after correction using RMieS-EMSC, in which the amide I is shifted for its correct position (1656 cm^{-1}) and the baseline alterations as well as the anomalous artifact are completely removed. The RMieS-EMSC algorithm has been well succeeded in vibrational spectroscopy in removing contributions from a given interfering signal, baseline corrections, offsets, and normalizing spectral data.

The algorithm operates taking into account the mathematical equation describing the resonant Mie scattering⁵⁹, while the multiplicative part of the algorithm compensates for the

differences in the optical path length and normalizes the spectra^{57; 59}. In order to satisfactorily correct the spectral distortions, 4 variables play an important role as input parameters:

1. Diameter of the scattering particles;
2. Refractive index;
3. Reference spectrum;
4. Number of iterations;

In our study, we considered scattering particles with diameters between 2-10 μm and refractive index values between 1.1 and 1.5, which are recommended values for biological tissues.

The spectrum used as reference can be the average spectrum of each image or a spectrum with similar characteristics. Many studies have used the spectrum of Matrigel, a gelatinous protein mixture secreted by Engelbreth-Holm-Swarm (EHS) mouse sarcoma membrane (Corning®, NY, USA), as reference for the algorithm since its composition contains all the key components expected in tissue (lipids, proteins, nucleic acids, and carbohydrates). However, the FTIR images used in our study were obtained from samples paraffin-embedded, and thus present additional vibrational modes related to paraffin that are not considered in Matrigel. Thus, in order to evaluate the influences of the reference spectrum in the final correction, both Matrigel and averaged spectrum were separately used as input for the algorithm.

The RMieS-EMSC algorithm takes the scatter-free reference spectrum (Z_{Ref}) in order to reconstruct the raw spectrum (Z_{Raw}). At the end of the first iteration/correction, the corrected spectrum (Z_{Corr}) presents similarities to both Z_{Ref} and Z_{Raw} . In the following, Z_{Corr} is used as the new reference spectrum in a new iteration in order to obtain more accurate representation of the pure absorbance spectrum. According to Bassan *et al*, the number of iterations used as input in the RMieS-EMSC depends on the experiment being performed. The raw spectral data must be corrected to the point where the biochemical differences between classes are greater than the scattering differences to allow accurate discrimination between the groups⁵³. In our study, satisfactory data discrimination were only achieved using 10 iterations.

3.3.2 Removal of paraffin bands

The acquisition of FTIR spectra via transmission requires samples with thickness not exceeding 12 μm in order to avoid non-linear responses in the detector⁴⁷. For this, tissue samples are placed in a more rigid fixative, in a process known as embedding, and subsequently are submitted to microtome in order to precisely obtain histological sections with specific thickness. In the present study, the samples were paraffin embedded, whose molecule is composed by long hydrocarbon chains that are infrared active. Figure 13.A shows the FTIR spectrum of paraffin and B represents a spectrum commonly obtained from biological tissue paraffin-embedded.

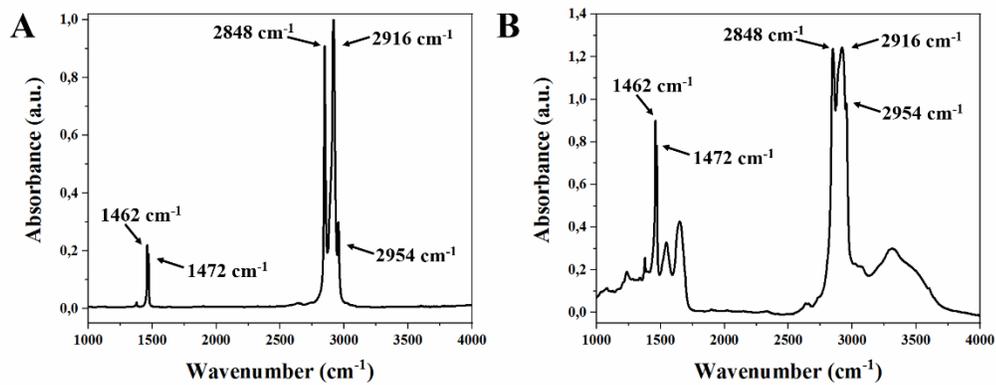


Figure 13. FTIR spectrum of paraffin (A) and biological tissue paraffin-embedded (B).

The bands identified at 1462, 1472, 2848, 2916 and 2954 cm^{-1} are assigned to paraffin and must be excluded from spectra prior any analysis^{30; 61}. The removal can be performed by using chemicals or computationally. In the first case, the samples are subjected to paraffin solvents (usually xylenes or hexane) prior the acquisition of data. Despite the satisfactory results obtained by this method, the reagents may also induce chemical alterations in tissue as well as increase the non-resonant Mie scattering effects by the formation of empty spaces that act as scattering centers. Furthermore, keeping tissue in paraffin creates some degree of refractive index matching that reduces the spectral distortions due to resonant Mie scattering²³. In the case of digital dewaxing, computational algorithms are employed to neutralize the paraffin contribution³¹ or simply by truncating the spectra and excluding regions of paraffin bands.

In our study, tissue samples placed on CaF_2 were left embedded in paraffin and the regions containing absorption bands of wax (1350-1490 cm^{-1} and 1800-3100 cm^{-1}) were manually

removed from spectra. For tissue prepared in conventional glass slides, the samples were chemically dewaxed and stained by H&E as previously described in the sample preparation section.

3.3.3 Spectra and Second derivatives

The information retained by spectral datasets can be assessed by different approaches in terms of input data. The most commonly used are the spectra and their second derivatives. Figure 14 show the fingerprint (A) and high wavenumber (B) regions of absorbance spectra collected from healthy cutaneous tissue placed on CaF₂.

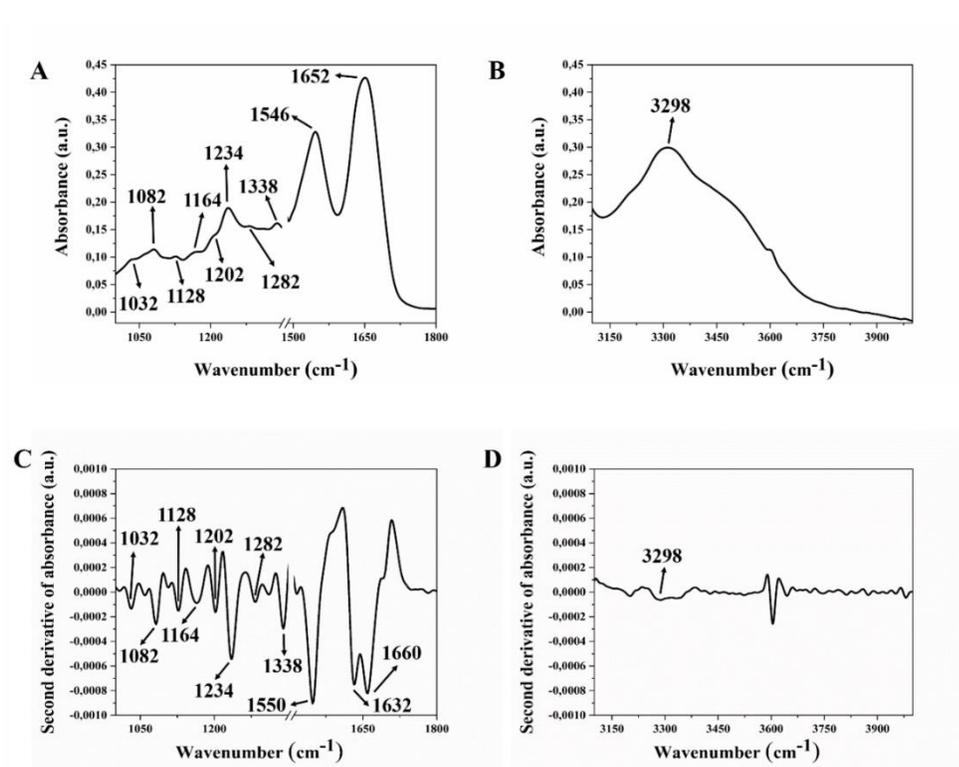


Figure 14. Fingerprint (A) and high wavenumber (B) regions of FTIR spectra collected from healthy cutaneous tissue placed on CaF₂ substrate. C and D depict the second derivatives calculated from spectrum in A and B.

In spectra, the bands associated to distinct vibrations from sample are presented in the graphic as maxima, which may overlap to each other according to the proximity of bands peak position. In the case of overlapping, the information provided by bands with lower absorption

intensity may not be accessible due to the stronger intensity of bands nearby. In such cases, second derivative filters represent a good alternative to enhance the resolution of infrared spectral data⁶². In our study, spectra and their second derivatives were employed in different occasions, as will be discussed in Chapters 4 and 5.

Figure 14.C and D show the second derivatives calculated from spectra depicted in panels A and B, in which the bands identified in spectra as maxima are converted into minima. The biological assignments of vibrational modes identified in Figure 14 are summarized in Table 3.

Table 3. Biological Assignments for Vibrational Modes Observed in the Second Derivative of Averaged Spectra

| Band peak (cm ⁻¹) | Biological assignment ^{63; 64} |
|-------------------------------|---|
| 1032 | Carbohydrates |
| 1082 | PO ₂ symmetric, Carbohydrates |
| 1164 | Carbohydrates |
| 1202 | Fibrous proteins |
| 1234 | PO ₂ asymmetric, Fibrous proteins |
| 1282 | Fibrous proteins |
| 1338 | Fibrous proteins |
| 1550 | Amide II (proteins) |
| 1632 | β-sheet structure of Amide I (proteins) |
| 1660 | α-helix structure of Amide I (proteins) |
| 3298 | Amide A |

FTIR spectra of cells and tissues are a complex superposition of the individual spectra of basic molecular units (carbohydrates, nucleic acids, proteins, and lipids) that compose biological samples. Bands peaking at 1032, 1082, and 1164 cm⁻¹ are assigned to C–O/C–C stretching

vibrations of carbohydrates⁶⁵. Vibrational modes at 1202, 1234, 1282 and 1338 cm^{-1} are attributed to the fibrous structure of proteins, mainly collagen and elastin in the case of skin samples^{64; 66; 67}. Bands in 1082 and 1234 cm^{-1} are also attributed to PO_2^- symmetric and asymmetric stretching vibrations from phosphodiester bonds of nucleic acids.

Spectral range from 1500–1800 cm^{-1} provide information about peptide bonding in proteins and their secondary structures. Band peaking at 1550 cm^{-1} is reported to amide II vibration and results from the combination out-of-phase of the N–H in plane bend and the C–N stretching vibration with smaller contributions from the C–O in plane bend and the C–C and N–C stretching vibrations^{68; 69}. The sub-bands of Amide I peaking in 1632 cm^{-1} (β -sheet) and 1660 cm^{-1} (α -helix) are associated to secondary structures of proteins and can only be identified taking the second derivatives instead of spectra. These vibrations result from the C=O stretching with minor contributions from the out-of-phase C–N stretching, the C–C–N deformation and the N–H in-plane bend⁶⁸. Besides the bands peaking in the fingerprint region, a vibrational mode associated to proteins can also be observed in the high wavenumber region peaking at 3028 cm^{-1} . This band is commonly referred as Amide A and arise from N–H of peptide bonding.

3.4 Data analysis

3.4.1 Principal Component Analysis (PCA)

In biospectroscopy, the differences between spectral data collected from healthy and diseased groups may not be immediately obvious upon visual inspection. Thus, sophisticated statistical methods may be required in order to assess subtle differences that a human observer cannot easily see. PCA is often the first tool employed to gain insight into any patterns within the data as well as for exploratory analysis (i.e. with no prior knowledge of the samples).

PCA is an unsupervised multivariate statistical method used to reduce the dataset dimensionality. The procedure is an orthogonal transformation that constructs the principal components (PCs) as linear combinations of the original variables weighted by their contribution to the original overall variance⁷⁰.

The starting point for the PCA technique is the original data matrix X :

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1J} \\ x_{21} & x_{22} & \cdots & x_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{IJ} \end{bmatrix}$$

Each sample is represented by a line-vector and each variable by a column-vector. In this case, each sample (vector-line) corresponds to one spectrum and each variable (vector-column) is represented by the absorbance values of the wavenumbers associated to spectra. The mathematics behind PCA aims to decompose our data matrix $X(I \times J)$ into two matrices, which are the scores, T , and loadings, L , so that:

$$X = TL^T$$

the scores retain the relation between samples, while the loadings indicate the relation between the variables. This decomposition is equivalent to a change of basis of the space of the original variables, J , to the space of principal components. Each one of the J columns of matrix L define the direction of a new axis, *i.e.* of a principal component in the new basis. If there is correlation between the original variables, not all J principal components are required to represent the data, but only a subset of them. The matrix of scores is constructed so that the new variables (PCs) retain the variance of the original data in a decreasing way, *i.e.*, the first PC (PC-1) retains the most variance from the data set while the last component retains the least variance^{45; 71}.

The interpretation of PCA results must be carried out using information from both matrices of scores and loadings. In order to exemplify how to evaluate the PCA results, consider the situation in which 100 FTIR spectra of healthy cells are compared to 100 spectra of cancer cells in a hypothetical experiment. The average spectrum obtained for healthy and neoplastic groups are shown in Figure 15.A, in which it is possible to identify bands absorbing at the wavenumbers represented by A, B, C, D, E, F, G and H. No significant changes are observed comparing the averaged spectra from both groups. Therefore, the data matrix containing all 200 spectra is decomposed into the matrices of scores and loadings using PCA.

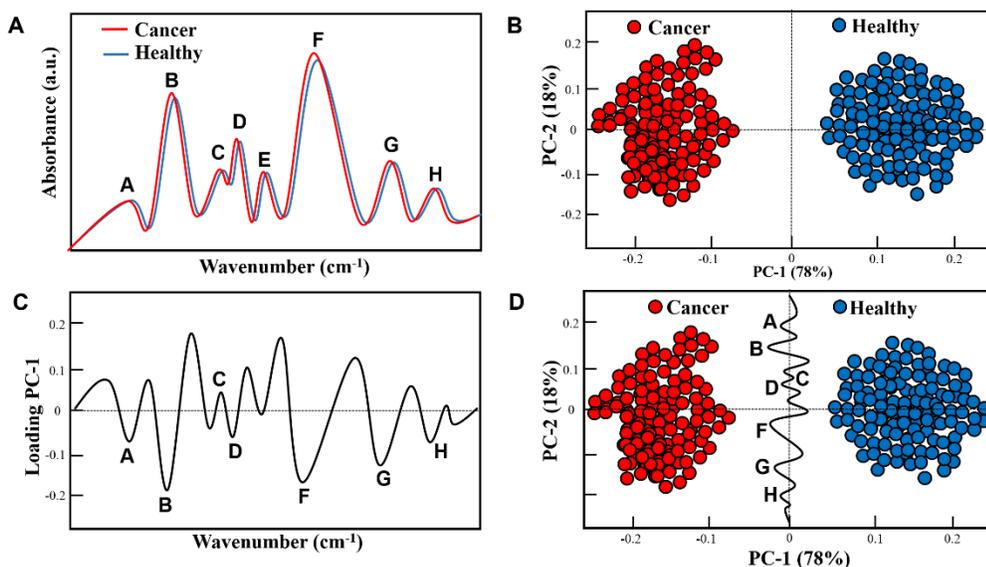


Figure 15. Example of results obtained by PCA applied to FTIR spectra (A) in a hypothetical experiment. B: Scores plot using PC-1 and PC-2. C: Loadings plot of PC-1. D: Superimposition of loading and scores plots.

The projection of the original observations (absorbance associated to each wavenumber) into the PC space is graphically represented using the so-called scores plot (Figure 15.B). In the present case, the scores plot plane was constructed using the first two columns of the matrix of scores (PC-1 and PC-2) and retain 96% of the overall variance of the original data (78%+18%).

In Figure 15.B, the scores associated with healthy spectra are grouped on the positive side of PC-1, whilst scores related to cancer cells are on the negative side. Thus, we conclude that the first principal component is able to satisfactory discriminate both groups. Figure 15.C shows the loadings plot of PC-1 and provide information about the variables responsible by data discrimination in the scores plot. Negative loadings (A, B, D, E, F, G and H) indicate bands presenting higher absorption intensity in the scores grouped in the negative side of PC-1, whilst the positive loadings are related to bands grouped in the positive side. Figure 15.D shows the loading plot superimposed to the score plot in order to facilitate the understanding of the technique. The vibrational modes A, B, D, E, F, G and H are directed towards to the region where the scores of cancer cells are grouped, indicating that these vibrational modes present higher absorption intensity in cancer cells compared to the healthy group.

4 FTIR image segmentation

4.1 Background

For decades, histopathology with routine hematoxylin and eosin staining has been and still remains the gold standard in studying the morphological features of tissue samples⁷². However, new imaging technologies that are able to provide simultaneously information about histological structures as well as the localisation and magnitude of molecular units (proteins, nucleic acids, lipids, and carbohydrates) have been explored over the past decades as potential tools in studying tissue sections⁷³. Among the current methods used for imaging, FTIR microspectroscopy shows great promise as a tool to assess the tissue architecture⁷⁴. The combination of spectroscopy and microscopy technologies enabled wide-field scanning of a sample, providing a hyperspectral image with tens of thousands of spectra in a few minutes⁴⁷. Infrared spectroscopic imaging provide morphochemical maps that are able to provide both spatial and compositional information based on signatures of vibrations containing molecular interactions occurring between biomolecules¹³.

Figure 16.A shows a non-processed FTIR hyperspectral image collected from healthy cutaneous tissue, in which it is not possible to identify the morphological structures of skin, i.e., dermis, epithelium, and stratum corneum. In general, raw datasets acquired on FTIR imaging mode are not able to provide information about morphology. However, the images can be segmented/reconstructed by distinct approaches in order to generate pseudocolor maps that reveal relevant information about histological structures. Different clustering algorithms have been proposed as segmentation methods aiming to increase the information content of FTIR images^{62; 75; 76; 77}. For this, spectral data are used as input for some clustering method in order to partitionate the dataset into sub-groups (clusters) of spectra with similar spectral features. The differences between the data within each cluster are minimised, whereas the differences between clusters are maximised. At the end of the segmentation process, each cluster represent regions of the image with similar biochemical profiles⁴⁵. Figure 16.C shows the same FTIR image of panel A after segmentation into 4 clusters using K-means cluster analysis (KMCA), in which the structures of skin (EP, epithelium; D, dermis; *, stratum corneum, and #, regions free of tissue) can be identified.

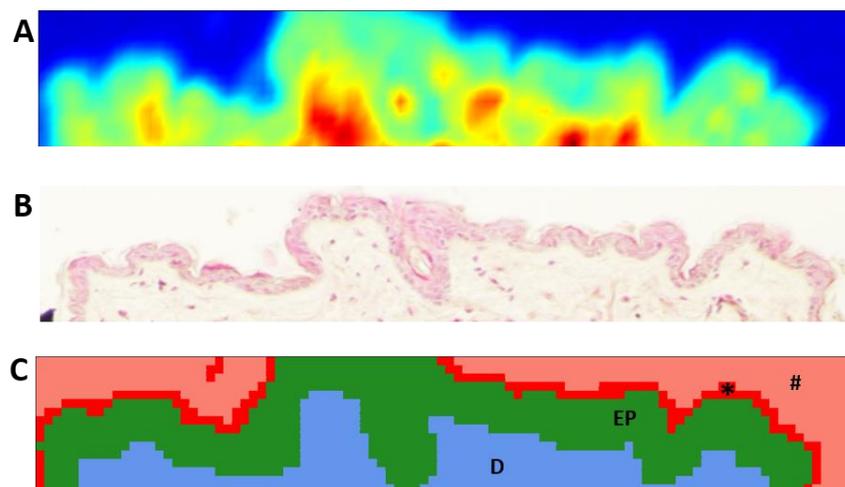


Figure 16. Segmentation of FTIR image using K-means clustering analysis. A: Non-processed FTIR image. B: Photomicrograph of specimen H&E stained. C: FTIR image partitioned into 4 sub-groups (D, dermis; EP, epithelium; *, stratum corneum; #, regions free of tissue).

Most studies about segmentation of FTIR images have focused on evaluate the influence of distinct algorithms used to partitionate the datasets. However, the final result achieved post-segmentation may be influenced by several other parameters including the spectral data used as input, preprocessing methodologies, the substrate used to place the specimens prior image acquisition, and sample preparation. Systematic comparative tests of such parameters as well as their influence on image segmentation are still missing. Thus, the present section aims to evaluate the ability of two well-established pattern recognition methods (KMCA and Hierarchical cluster analysis, HCA) as segmentation algorithms to reconstruct FTIR hyperspectral images varying the substrate used to place the samples as well as the methodologies used to preprocess the spectral data.

4.2 Methodology

4.2.1 Image segmentation

Figure 17 shows the workflow with the different methodologies used in this study to reconstruct the FTIR hyperspectral images. As mentioned in the section 2.2.2 (sample preparation), 2 sequential histological slices were obtained from FFPE tissue blocks and placed

on CaF₂ (non-stained) and glass substrates (H&E stained). FTIR images were collected from both substrates according to the parameters described in the section 3.2.1 (data collection).

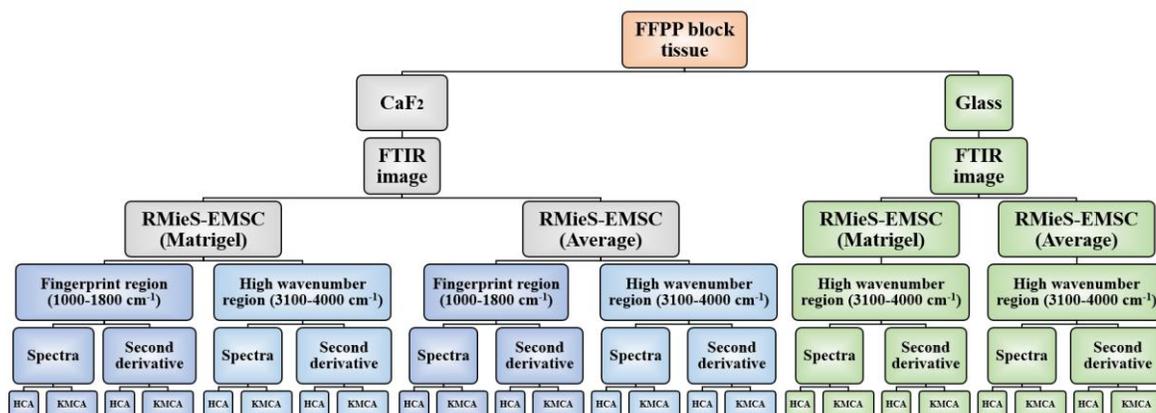


Figure 17. Workflow for image segmentation using distinct methodologies

The raw images were submitted to RMieS-EMSC algorithm in order to remove dispersion artifacts induced by scattering as well as to normalize the data. The spectrum used as reference for the RMieS-EMSC algorithm was varied (Matrigel or average spectrum) aiming to evaluate the influence of the reference spectrum in the image segmentation. In the following, the fingerprint (1000-1800 cm⁻¹) and high wavenumber (3100-4000 cm⁻¹) regions from images collected on CaF₂ substrates were truncated and used separately as input for image reconstruction. For samples placed on glass, only the high wavenumber region was assessed due to the opacity of glass in the fingerprint region. Second derivatives were calculated from spectra and both were employed as input for image segmentation using HCA and KMCA algorithms.

4.2.2 Tissue specimen

FTIR hyperspectral images were acquired from histological sections obtained from the same FFPP block of tissue of papilloma shown in Figure 18.A. False-color maps obtained after segmentation were compared in terms of image quality and consistency with standard histopathology in order to evaluate the ability of each method to reproduce the histological structures of the specimen (D, dermis; EP, epithelium; K, keratin; NT, regions with no tissue).

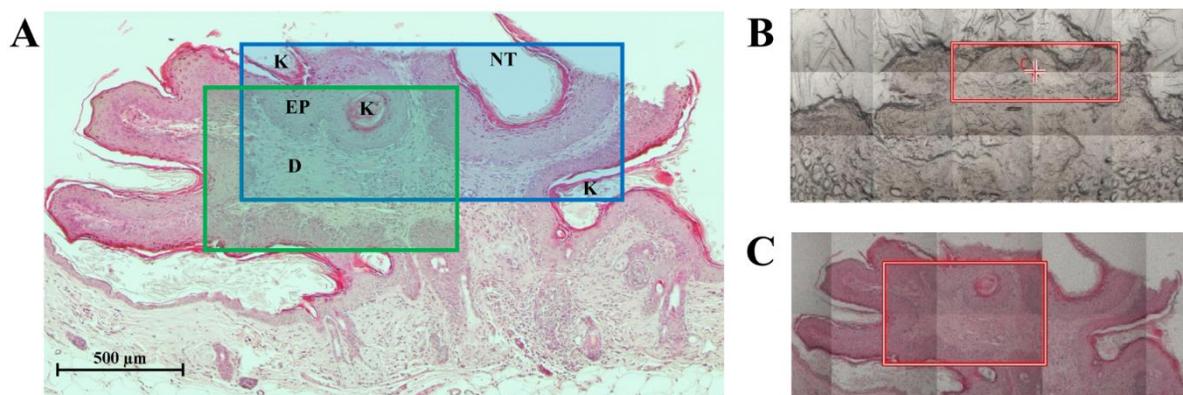


Figure 18. A: Photomicrograph of specimen used to collect FTIR images in order to evaluate the influence of different methodologies in the clustering task (K, keratin; EP, epithelium; NT, regions free of tissue; D, dermis). Green and red rectangles depict the ROI that FTIR images were acquired from tissue placed on CaF₂ (B) and glass (C).

Figure 18.B shows a screenshot of the software used to operate the spectrophotometer during the acquisition of FTIR image collected from tissue placed on CaF₂ (non-stained). The red rectangle represents the region of interest (ROI) in the specimen that the hyperspectral image was collected and it is depicted in Figure 18.A by the region delimited by blue rectangle. In Figure 18.C, the red rectangle depicts the ROI related to the hyperspectral image acquired from tissue placed on glass substrate and it is represented in Figure 18.A by the green rectangle. The comparison of blue and green rectangles in Figure 18.A indicate that FTIR images collected on both glass and CaF₂ substrates were not acquired from the exact same region.

4.3 Results

The basic idea of using clustering algorithms to obtain pseudocolor maps from FTIR spectral data collected on imaging mode is to partitionate all pixels of the image into subgroups (clusters), which are presented in distinct colors in the obtained image at the end of the process. Figure 19 shows a FTIR hyperspectral image segmented into 2 (Figure 19.B), 3 (Figure 19.C) and 4 (Figure 19.D) clusters in order to demonstrate how the segmentation process works. The FTIR image was collected from tissue placed on CaF₂ substrate and the pixels were clustered by KMCA using the fingerprint region of second derivatives.

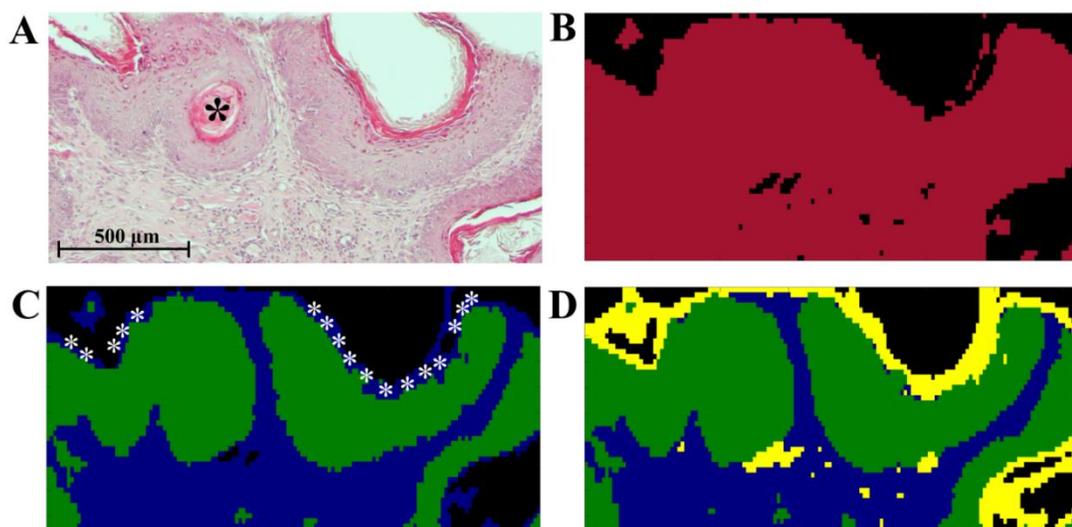


Figure 19. A: Photomicrograph of specimen depicting the ROI that FTIR image was acquired. FTIR image segmented into 2 (B), 3 (C), and 4 (D) clusters using KMCA

The comparison of the pseudocolor map depicted by Figure 19.B (2 clusters) to the photomicrograph of the specimen (Figure 19.A) indicate that 2 clusters are not sufficient to reproduce the histological structures of the skin. The only information provided by this false-color map is the identification of regions with and without tissue, which are represented by pixels colored in red and black respectively. In the image reconstructed by 3 clusters (Figure 19.C), the pixels previously associated to tissue are partitioned into 2 new subgroups representing epithelium (green) and dermis (blue). The white asterisks (*) highlight pixels that were assigned to regions free of tissue in the image partitioned into 2 clusters and that were updated for pixels representing tissue in the Figure 19.C. Changes in the assignment of pixels according to the number of subgroups used to partitionate the image may be expected due to the differences between the data within each cluster, especially for pixels located on the border of different regions whose spectra may retain mixed information from both sides. In the image reconstructed by 4 clusters (Figure 19.D), the pixels on the border are assigned to a new cluster related to the keratin layer (yellow). Thus, the identification of histological structures can be achieved by increasing the number of clusters used to reconstruct the FTIR image. However, despite the high correlation between the image segmented into 4 clusters with the photomicrograph, it is still not possible to identify the keratin structure depicted by the black asterisk in Figure 19.A. Therefore, it is necessary to modify the methodology used to obtain the

false-colors maps in order to precisely reproduce the histological structures of the specimen.

These alterations may include:

1. Increasing the number of sub-groups to partitionate the image;
2. Using other pattern recognition techniques (HCA, random forest, support vector machines, artificial neural networks, ...);
3. Modifying the substrate in which the sample was collected;
4. Changing the input data and/or the preprocessing methods (spectra, first derivatives, second derivatives, reference spectrum used in the RMieS-EMSC, ...)
5. Altering the spectral region (high wavenumber or fingerprint region or by using bands peaking in a specific range)

4.3.1 Reconstructing FTIR image collected from tissue placed on CaF₂ substrate (fingerprint spectral region)

Figure 20 show false-color maps obtained from a FTIR image segmented by HCA (panels A, B, E and F) and KMCA (panels C, D, G and H) using 4 (A, B, C and D) and 8 (E, F, G and H) clusters to partitionate the pixels. The FTIR image was collected from tissue placed on CaF₂ substrate and the fingerprint region of spectra (1000-1800 cm⁻¹) were used as input data for each clustering algorithm. Panels A, E, C and G relate to images submitted to RMieS-EMSC using Matrigel as reference spectrum, whereas images from panels B, F, D and H were corrected taking the average spectrum calculated from each image as reference.

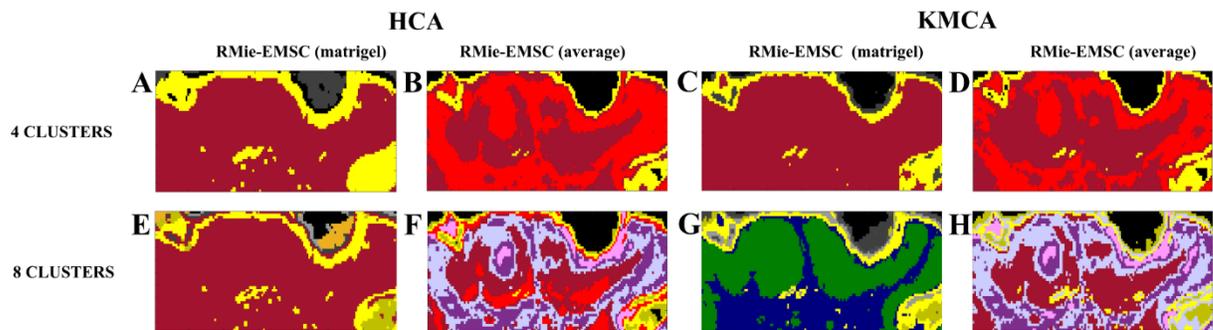


Figure 20. False-color maps achieved by using the fingerprint region of spectra from FTIR image that was segmented by HCA (panels A, B, E and F) and KMCA (panels C, D, G and H) using 4 (A, B, C and D) and 8 (E, F, G and H) clusters to partitionate the pixels. Panels A, E, C and G relate to images submitted to RMieS-EMSC using Matrigel as reference spectrum, whereas images from panels B, F, D and H were corrected taking the average spectrum calculated from each image as reference.

At a first glance, the false-color map depicted in panel G is the most similar to the photomicrograph of the specimen, in which it is possible to identify pixels related to epithelium (green), dermis (blue), keratin layer (different shades of yellow) and regions with thin layers of tissue (grey) and no tissue (black). Despite the poor similarity between the images depicted by panels A, B, C, D, E, F and H to the photomicrography, both HCA and KMCA algorithms presented similar results regarding the segmentation process, i.e., the maps $A \times C$, $B \times D$ and $F \times H$ are similar to each other. Only the pairwise comparison $E \times G$ presented poor correlation.

Figure 21 show maps obtained by the same parameters used to reconstruct the images depicted in Figure 20, but using second derivatives instead of spectra as input data.

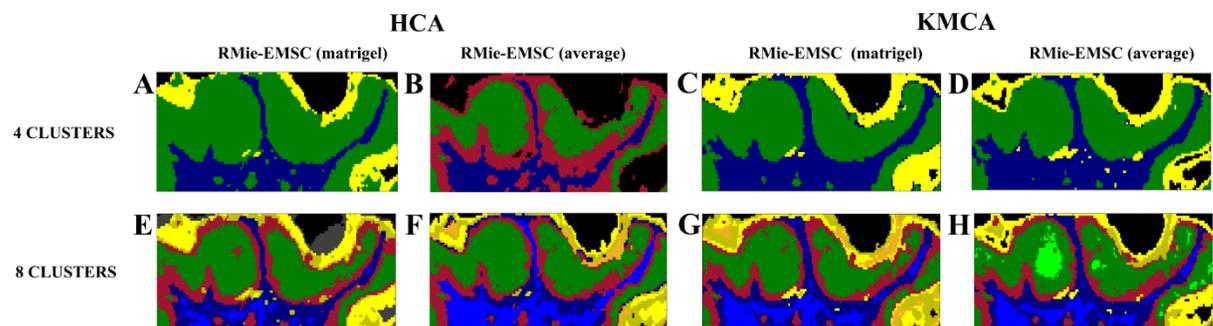


Figure 21. False-color maps obtained by using second derivatives of fingerprint region as input data to partitionate FTIR image collected from tissue placed on CaF_2 .

Most of the false-color maps showed high correlation with the photomicrograph. In panels A-H, the different shades of yellow encode keratin, green pixels depict epithelium, regions free of tissue are represented in black, blue relate to dermis and red indicate pixels on the border of distinct histological regions that could not be unequivocally assigned to epithelium, dermis or keratin. The image represented in panel H achieved the highest degree of correspondence in terms of reproducing the structures of the specimen and was the only one able to identify the keratin structure highlighted in Figure 19.A by the black asterisk. On the other hand, the image shown in panel B showed the lowest degree of similarity, since not even the keratin layer was evidenced.

4.3.2 Reconstructing FTIR image collected from tissue placed on CaF₂ substrate (high wavenumber region)

The previous section evaluated the false-color maps obtained by distinct image segmentation methods using the fingerprint region (1000-1800 cm⁻¹) of spectra as input. In this section, the high wavenumber region was used to reconstruct the FTIR hyperspectral images collected from tissue placed on CaF₂ substrates in order to evaluate the influence of different spectral ranges to reproduce the histological structures of skin. Unlike the fingerprint region that is composed by many vibrational modes associated to different molecular units (proteins, carbohydrates, lipids, and nucleic acids), the high wavenumber region presents only the vibration associated to Amide A (proteins). Figure 22 show the false-color maps achieved after image reconstruction using spectra as input.

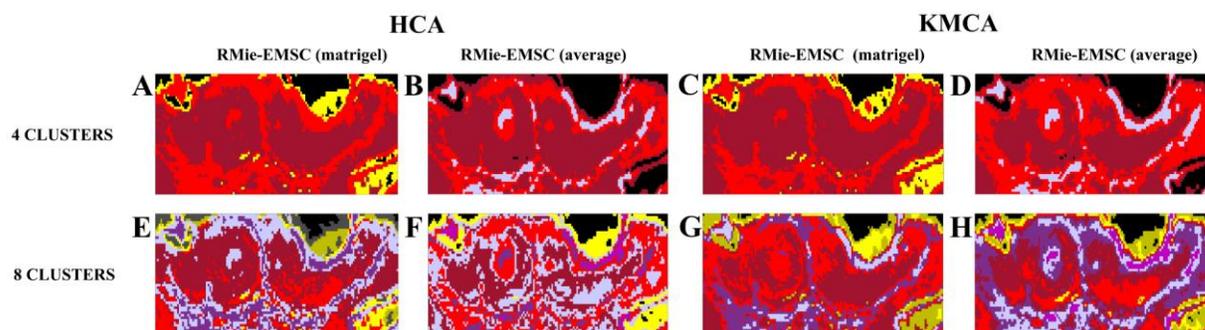


Figure 22. False-color maps obtained by using the high wavenumber region of spectra as input to partitionate the FTIR image.

All segmented images presented poor quality and consistency compared to standard histopathology. The pairwise comparisons of the results obtained by HCA and KMCA (A×C, B×D, E×G, and F×H) indicate similar outcomes using both clustering algorithms.

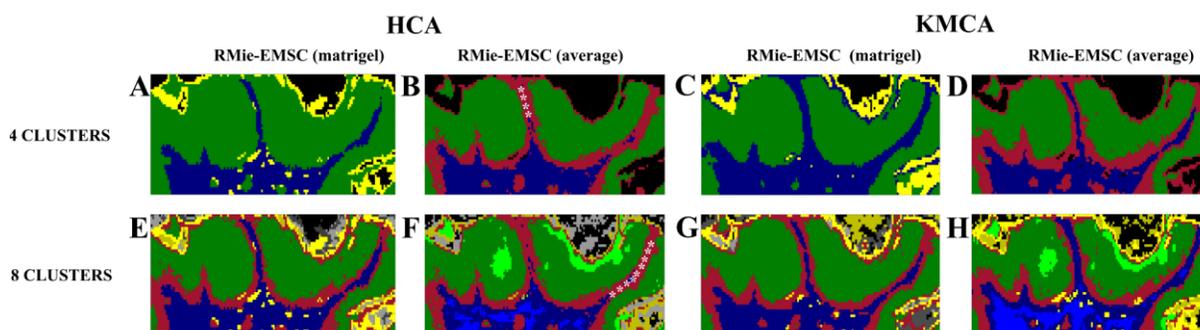


Figure 23. False-color maps obtained by using the high wavenumber region of second derivatives calculated from spectra as input to partitionate the FTIR image.

Figure 23 present the results obtained by using the second derivatives as input data. All images presented high resemblance to the photomicrography (green = epithelium, blue = dermis, yellow = keratin layer, black = regions free of tissue, red = pixels on the border of distinct histological regions), in which the best results were achieved by using the segmentation approaches that were used to reconstruct the morphochemical maps depicted by panels F and H. In both cases, the keratin structure highlighted by the black asterisk in Figure 19.A was identified and it is represented by pixels colored in light-green. Despite the high resemblance between panels F and H, there is a subtle variation that is highlighted in panel F by white asterisks (*). This region is associated to dermis and it was correctly assigned to it in panel H (blue pixels). On the other hand, the map reconstructed by HCA failed in correctly identifying such regions, and thus recognized them as regions on the border of distinct histological regions (red pixels). Similar findings were also observed in the pairwise comparison B×D. For the comparisons A×C and E×G, HCA and KMCA achieved similar results

4.3.3 Reconstructing FTIR image collected from tissue placed on glass slide (high wavenumber region)

Figure 24 and 26 show the false-color maps obtained by using spectra and second derivatives as input data to segment the FTIR image collected from tissue placed on glass

substrate. As mentioned before, the fingerprint region of spectra is not available due to the opacity of glass. Thus, only bands peaking at the high wavenumber range were used for image reconstruction.

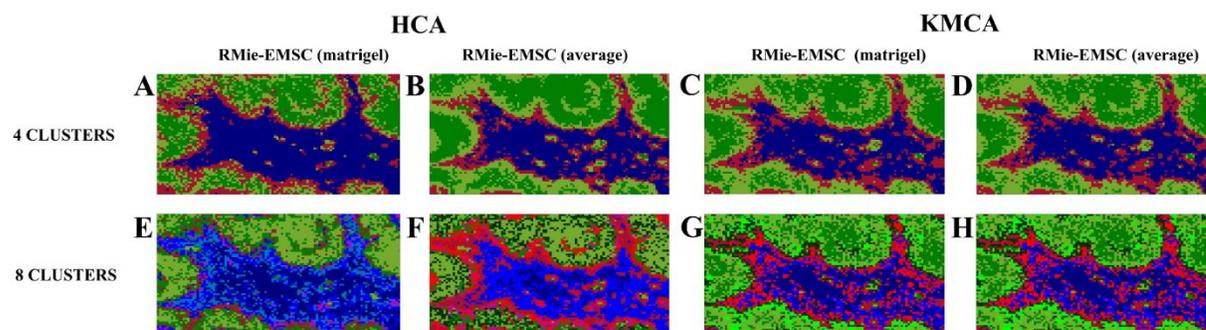


Figure 24. False-color maps obtained by using the high wavenumber region of spectra as input to partitionate the FTIR image acquired from tissue-samples placed on glass.

All morphochemical maps obtained by using spectra (Figure 24) presented poor quality and noisy aspect compared to the maps reconstructed from images collected on CaF_2 substrates. However, it is possible to identify pixels associated to dermis (blue) and epithelium (green) but with no success in evidencing the keratin structure highlighted by black asterisk in Figure 19.A.

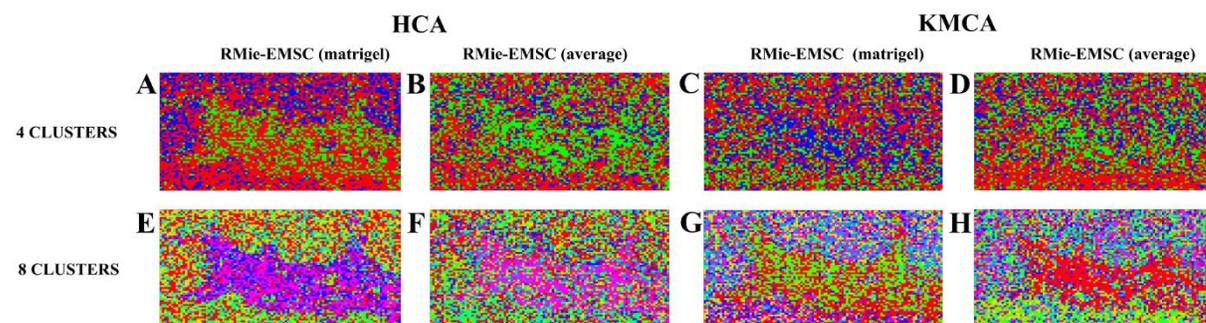


Figure 25. False-color maps obtained by using the high wavenumber region of second derivatives as input to partitionate the FTIR image acquired from tissue-samples placed on glass.

None of the false-color maps reconstructed using second derivatives (Figure 25) were able to reproduce the histological structures of the specimen. The images presented a considerable increase in the noisy aspect compared to the images segmented taking the spectra as input data.

4.4 Discussion

False-color maps obtained from FTIR images collected from tissue fixed onto CaF₂ presented higher quality and similarity to standard histopathology than samples placed on conventional glass. Second derivatives achieved better results than spectra, indicating that overlapped vibrational modes increase the information of FTIR images. Noisy aspect was observed in the images acquired from glass, suggesting that HCA and KMCA were not able to precisely assign the pixels to the histological structures of the specimen. The satisfactory results obtained using the high wavenumber region of images acquired on CaF₂ suggest that the recognition of histological features in images collected on glass is being adversely affected by the additional procedures used in sample preparation instead of the spectral range used as input in the clustering task. It is known that the removal of paraffin using chemicals contribute to spectral distortions related to Mie scattering²³, nevertheless, such effects are satisfactory removed by RMieS-EMSC algorithm. Thus, the dewaxing and staining procedures may have resulted in contributions that negatively minimize the spectral differences between the data within each cluster, and thus affecting the quality and consistency of the reconstructed images. Measurements should be performed on glass using paraffin-embedded and unstained samples in order to confirm this hypothesis. To the best of our knowledge there are only two studies that used conventional glass as substrate to collect FTIR chemical images from tissue-samples^{23; 28}. Satisfactory segmentation and high image quality were achieved in both cases, in which the authors were able to correctly identify pixels related to epithelium, stroma, blood, necrosis, and concretion (luminal secretions commonly found in benign prostate acini). However, it is important to notice that Random forest was employed in both cases to reconstruct the FTIR images instead of HCA or KMCA. Random forest is a robust pattern recognition method that operates using training and testing sets in the prediction task (supervised algorithm), which may be more sensitive in detecting small differences presented by the datasets and, consequently, achieve better results in correctly recognizing the histological features of samples. Random forest should be performed in our data to confirm this hypothesis.

Although there are several studies evaluating the influence of different algorithms to segment hyperspectral images, there is still no consensus about which method is suitable for the clustering task. According to our findings, the images segmented by KMCA presented slightly

superior ability in correctly assigning the pixels of false-color maps to the histological structures of the specimen. Lasch and coworkers showed that HCA proved to be better than fuzzy C-means and KMCA in differentiating the structures of colorectal adenocarcinoma section⁷⁵. Thus, the image segmentation seems to be a try and error process whose ability to correctly reproduce the morphological features of the specimen vary with the type of sample. Furthermore, the fact that distinct written in-house algorithms have been employed in different studies by research groups worldwide is also a key factor influencing the variability of the obtained results.

The large majority of proof-of-concept studies that evaluated the potential biomedical applications of FTIR spectroscopy have considered the vibrational modes of fingerprint region as most useful for diagnostic purposes due to the higher amount of bands associated to the molecular units that compose the cells (lipids, carbohydrates, nucleic acids, and proteins). However, in our study, the second derivatives of datasets obtained from tissue placed on CaF₂ achieved similar outcomes using the fingerprint and high wavenumber regions, suggesting that the vibrational modes in both spectral range provide sufficient information to correctly reconstruct the images. Even the false-color maps obtained by taking the spectra as input data, which failed in revealing the histological structures of the specimen, demonstrated similar outcomes comparing the results achieved by both regions (except the images corrected by the RMieS-EMSC using the Matrigel as reference spectrum). Matrigel spectrum is the most used as reference in several studies, however it does not take into account the paraffin contributions. Thus, we used both Matrigel and average spectrum as reference for RMieS-EMSC in order to evaluate their influence in the final result. Our findings indicate that the choice of reference spectrum has very little difference on the outcome of image analysis. In Figure 22 and 24, the keratin layer is not evidenced in the false-color maps segmented into 4 clusters using the average spectrum as reference, and thus suggest better outcomes by using Matrigel. However, this issue was solved by increasing the number of subgroups to partitionate the images, as can be observed in maps segmented into 8 clusters. In an investigation about the role of the initial spectrum used as reference for RMieS-EMSC, Bassan and coworkers investigated the effects of three different spectra used as reference in correcting the same data. According to their findings, similar outcomes are achieved regardless the reference spectrum when using higher number of iterations⁵³. Thus, the number of iterations used to preprocess the images in our study (10

iterations) may not have resulted in identical false-color maps, but the differences between the images reconstructed by Matrigel and average spectrum are relatively small.

4.5 Conclusions

In this chapter, HCA and KMCA statistical techniques were used as segmentation algorithms to reconstruct FTIR hyperspectral images in order to evaluate the ability of false-color maps in reproducing the histological structures of a primary papilloma. The images were segmented using several different approaches varying the substrate to place the sample (CaF₂ or conventional glass), the spectrum used as reference for RMieS-EMSC algorithm as well as the input data (spectra or second derivatives) for each clustering technique. Fingerprint (1000-1800 cm⁻¹) and high wavenumber (3100-4000 cm⁻¹) regions from images collected on CaF₂ were separately used as input for image reconstruction and only the high wavenumber range was employed in the case of samples placed on glass. All pseudocolor maps were compared to standard histopathology in order to evaluate the quality and consistency of images after segmentation. Satisfactory results were obtained by using the second derivatives instead of spectra, in which the false-color maps from FTIR images collected from tissue placed on CaF₂ presented higher resemblance to standard histopathology than tissue mounted onto glass. Our findings indicate that vibrations over the range of high wavenumber region provide sufficient information in reconstructing images similar to the obtained using fingerprint range. We demonstrated that the number of iterations used in RMieS-EMSC is more important than the spectrum used as reference to correct datasets. Thus, despite the slightly superior ability of KMCA in correctly assigning the pixels of morphochemical maps to the histological structures of the specimen, we concluded that the choice of substrate, input data and preprocessing methods, as well as the sample preparation have more influence in the final results than the clustering algorithm used to reconstruct the images.

5 Staging nonmelanoma skin cancer using Infrared Spectral Histopathology

5.1. Background

Over the past decades, FTIR spectroscopy has emerged as a potential candidate to complement histopathology in the study and diagnosis of tissue diseases^{47; 78; 79}. Contrary to the histological examination, which relies on the morphological tissue alterations assessed by visual inspection of stained samples⁷³, FTIR chemical imaging is a rapid and label-free tool that probe the state of biological samples based on their overall biochemical status¹¹. Despite the many proof-of-concept studies demonstrating the effectiveness of FTIR spectroscopy in detecting biological disorders with high levels of sensitivity and specificity, the technique still remains in the research environment and translation into clinical practice has been relatively slow⁸⁰.

The substantial cost of infrared transparent substrates required to collect chemical images is one of the main challenges that have hindered the transition of FTIR spectroscopy from fundamental research to clinical arena^{23; 28}. Typically, the price of infrared substrates ranges from \$40-150 per slide (according to the quality and purity), while 1 box containing 50 conventional glass slides costs only \$5. Thus, alternative substrates have been investigated by spectroscopic community in order to reduce the costs demanded by FTIR technology.

An additional factor that will facilitate the implementation of the technique into the clinical environment is the development of a methodology that do not disrupt the pathologist's workflow, i.e., the measurements must be performed using the same specimen and sample preparation currently used during the diagnostic process. In this context, glass slides are the most obvious alternative for acquiring infrared spectra. However, placing the samples into conventional histology slides implies in not using the fingerprint region of spectra due to the opacity of glass. Although it has already been demonstrated that glass presents a transmission window in the high wavenumber spectral region ($2500-4000\text{ cm}^{-1}$)²⁸, few studies have investigated its diagnostic potential. In light of this, the present chapter is focused in demonstrate the ability of FTIR spectroscopy in discriminating healthy skin from different stages of skin cancer using tissue samples placed on glass and CaF₂ substrates.

5.2. Methodology

5.2.1 Tissue samples

Tissue specimens were prepared and placed on CaF₂ and glass substrates following the methodology described in the section 2.2.2. The procedures used in sample preparation may induce artifacts in the histological sections, so that not all sample units were submitted to FTIR imaging. Table 4 shows the number of images acquired from each mice group.

Table 4. Number of FTIR hyperspectral images acquired from glass and CaF₂ substrates from different groups.

| Group | Number of images collected from samples placed on CaF ₂ substrates | Number of images acquired from samples mounted on glass |
|-------|---|---|
| G1 | 9 | 8 |
| G2 | 9 | 7 |
| G3 | 7 | 7 |
| G4 | 7 | 6 |

For G4 group, the hyperspectral images were collected only from specimens classified as invasive squamous cell carcinoma.

5.2.2 FTIR data acquisition and spectral analysis

Infrared chemical images were acquired from samples placed on each substrate as described in the section 3.2.1. Not all spectra in a hyperspectral image present good quality, and thus must be removed prior any analysis. K-means clustering algorithm was applied to the images in order to identify poor quality data (Figure 26). The images collected from tissue placed on CaF₂ were submitted to RMieS-EMSC algorithm using the averaged spectrum calculated from each image as reference to correct distortions induced by scattering as well as to normalize the datasets. Subsequently, the spectral contributions of paraffin were removed from spectra using the procedure reported in the section 3.3.2, and the fingerprint and high wavenumber region of second derivatives were separately used as input data in KMCA. For

samples mounted on glass, KMCA the images were only processed for non-classical Mie scattering.

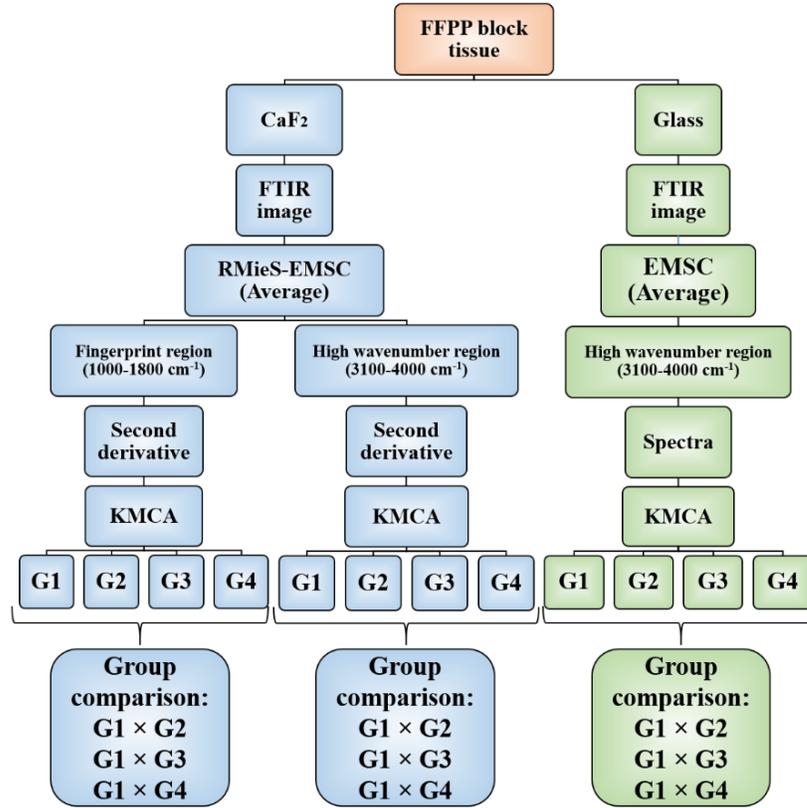


Figure 26. Workflow for preprocessing FTIR images

After removing poor quality data, healthy tissue was compared to other groups in order to evaluate the ability of FTIR spectroscopy to discriminate spectra from animals exposed to carcinogenic factors (G1×G2, G1×G3, G1×G4). For this, a matrix was constructed for each group (G1, G2, G3 and G4) containing non-excluded data from all images. Figure 27 shows the workflow used to compare G1 to G2, and similar methodology was used in other pairwise comparisons (G1×G3 and G1×G4). Firstly, the original matrix of each group was randomly split into 2 new matrices. A training set was constructed using 50% of data from G1 and 50% of G2, while the remaining data were employed to generate a testing set. Both training and testing sets were submitted to PCA and the first 3 principal components from training set were used as input to obtain a classification model via Linear Discriminant Analysis (LDA). The model was validated by hold-out cross-validation using 50% of data, and the LD function was subsequently

applied to the scores plot obtained using the 3 principal components calculated from testing set in order to evaluate the capability of the model in classifying new elements.

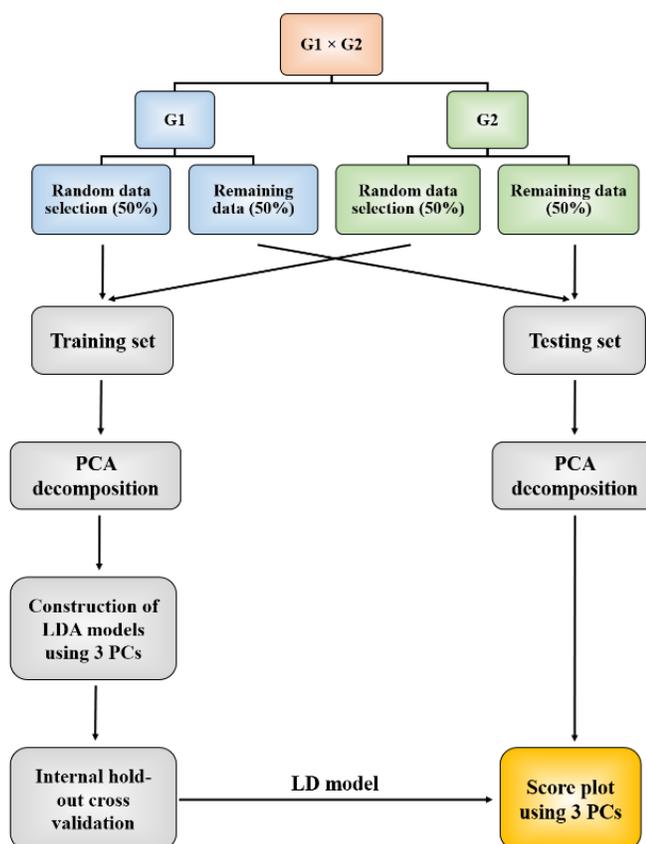


Figure 27. Workflow for the construction of Linear Discriminant model based on PCA that was used in spectral analysis.

At the end of the process, the accuracy, specificity and sensitivity obtained in each binary classification test were calculated considering true positive (TP) as PC-scores from animals exposed to chemicals that were correctly identified in the carcinogenesis groups; true negative (TN) as PC-scores from healthy animals correctly assigned to healthy group; false positive (FP) as PC-scores from healthy animals identified in groups that underwent to chemical; false negative (FN) as PC-scores from animals exposed to carcinogenic agents that were assigned to healthy group. Sensitivity measures the ability of the method to associate data from G2, G3, and G4 to the group of animals exposed to carcinogenic factors, whereas specificity measures the ability of the technique to correctly attribute spectral data acquired from healthy animals to healthy group.

5.3. Results

5.3.1 Tissue specimens placed on CaF₂ substrates (fingerprint spectral region)

Fingerprint region is by far the most used in studies evaluating the potential of FTIR spectroscopy as a diagnostic tool. Figure 28 shows the fingerprint region (1000-1800 cm⁻¹) of averaged second derivatives calculated from spectral data collected for each group, in which it is possible to identify the vibrational modes of distinct cellular components.

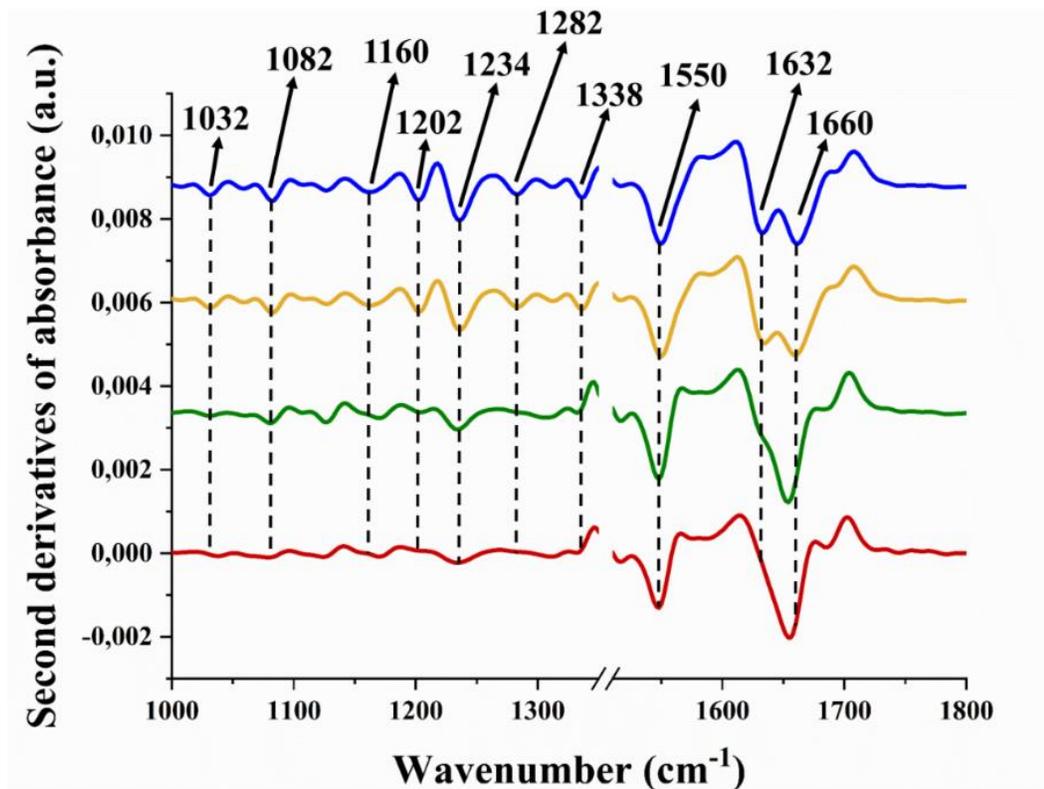


Figure 28. Averaged second derivatives calculated from fingerprint region of tissue placed on CaF₂. G1 (blue), G2 (yellow), G3 (green), and G4 (red).

Figure 29 shows the statistical measurements (accuracy, sensitivity and specificity) obtained by PC-LDA as a binary classification test to discriminate healthy tissue from groups submitted to carcinogenesis.

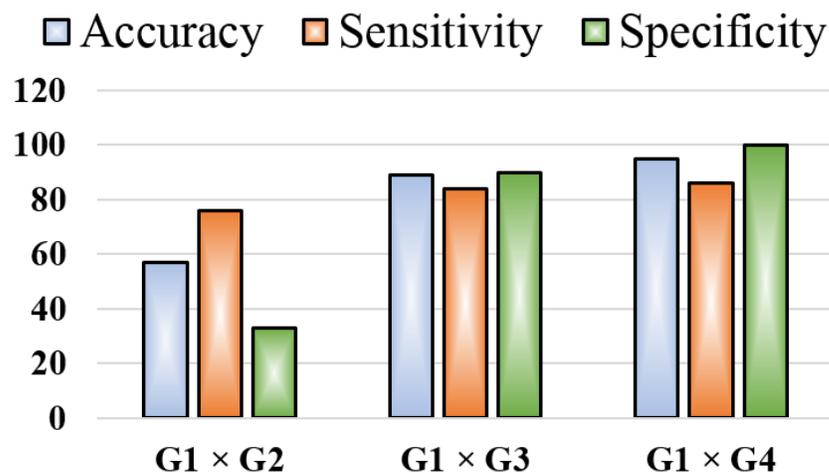


Figure 29. Statistical measurements obtained by PC-LDA as a binary classification test to discriminate healthy from diseased groups (G1×G2, G1×G3, and G1×G4).

Satisfactory discrimination (over 80%) were achieved comparing spectral data collected from healthy mice to animals exposed to 16 and 48 weeks of chemicals (G1×G3 and G1×G4). On the other hand, the method presented poor performance in classifying scores related to G1 and G2. Furthermore, it is possible to observe that the values of statistical measurements increase according to the period that animals underwent to carcinogenesis. Figure 30 show the scores and loadings plots from pairwise comparisons that achieved satisfactory discrimination by PC-LDA, i.e., G1×G4 and G1×G3.

According to the scores plots depicted in Figure 30.A and C, satisfactory data discrimination were achieved along the first principal component (PC-1) in both G1×G4 and G1×G3, in which most scores from animals exposed to carcinogenic factors are grouped on positive values of PC-1 axis and scores from healthy mice lie in the negative side. PC-1 loadings obtained in each pairwise comparison (Figure 30.B and D) indicate that similar variables are responsible by data grouping in both cases. Considering that PCA was applied on second derivatives, negative loadings are associated to positive scores and positive loadings relate to negative scores. Thus, the vibrational modes that presented positive loadings (1032, 1082, 1160, 1202, 1234, 1282, 1338 and 1562 cm^{-1}) represent bands with higher absorbance in the scores grouped on negative side of PC-1, i.e., healthy group. On the other hand, the band peaking at

1652 cm^{-1} presented higher absorbance in animals submitted to carcinogenesis. The vibrational mode at 1632 cm^{-1} in Figure 28 was not identified in the loadings, suggesting a significant decrease of β -sheet structures in mice submitted to 16 and 14 weeks of carcinogenesis. In Figure 28, the band related to Amide II is peaking at 1550 cm^{-1} , but presented loadings 12 cm^{-1} towards to higher wavenumbers (1562 cm^{-1}). Vibrations of α -helix also presented peak position shifted by 8 cm^{-1} , but towards to lower wavenumbers (1652 cm^{-1}).

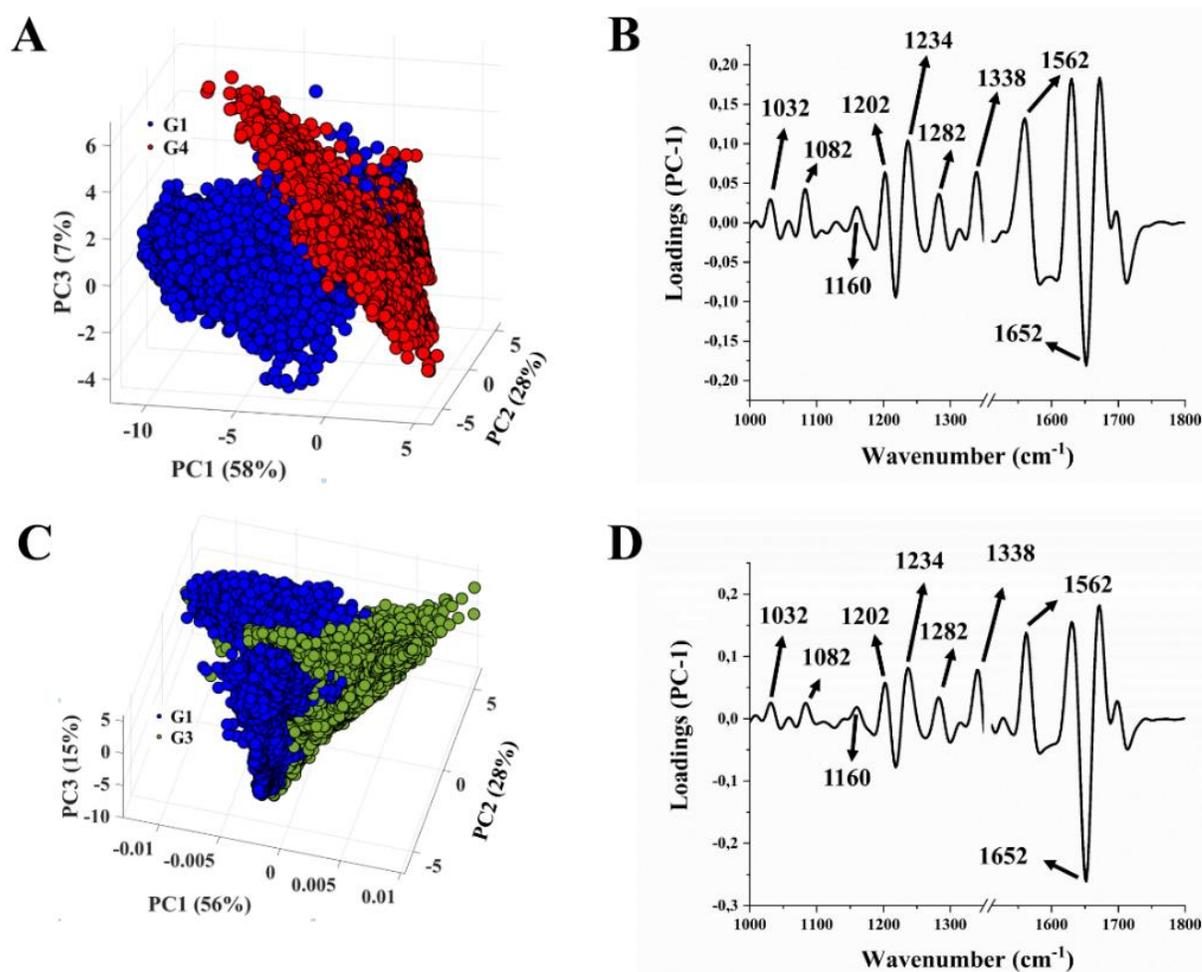


Figure 30. PCA results obtained comparing G1 to G2 and G3 using the fingerprint spectral region. A and C show scores plots of G1×G4 and G1×G3, respectively. B and D show loadings plots of PC-1 obtained in each pairwise comparison.

According to Figure 29, poor statistical measurements were obtained by PC-LDA comparing datasets from healthy group to animals exposed to carcinogenic agents for 8 weeks. The scores plot calculated for G1×G2 is presented in Figure 31.

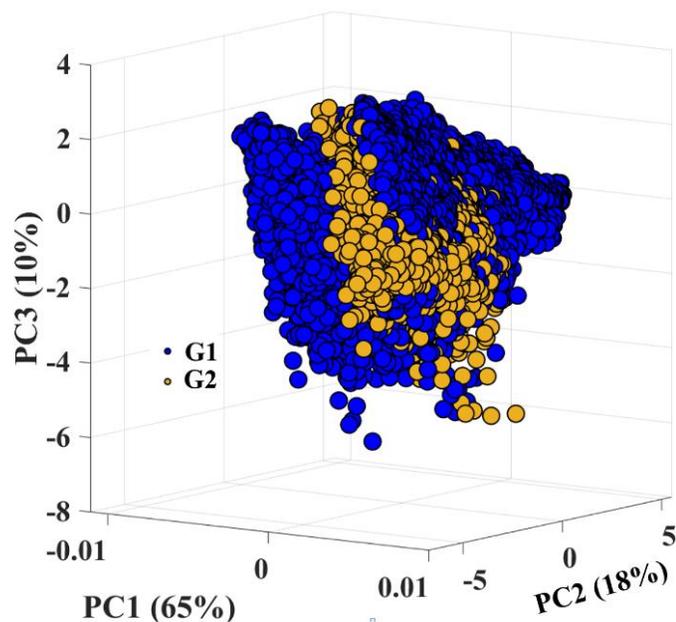


Figure 31. Scores plot obtained in the pairwise comparison of G1×G2

No marked differences were observed in the scores plot even plotting higher order principal components, suggesting that no clear biochemical changes are noticed between both groups. Thus, the loadings were not evaluated in this case.

5.3.2 Tissue specimens placed on CaF₂ substrates (high wavenumber spectral region)

Figure 32 shows the high wavenumber region (3100-4000 cm⁻¹) of averaged second derivatives calculated from spectral data collected from specimens placed on CaF₂, in which it is possible to identify the band associated to Amide A vibration.

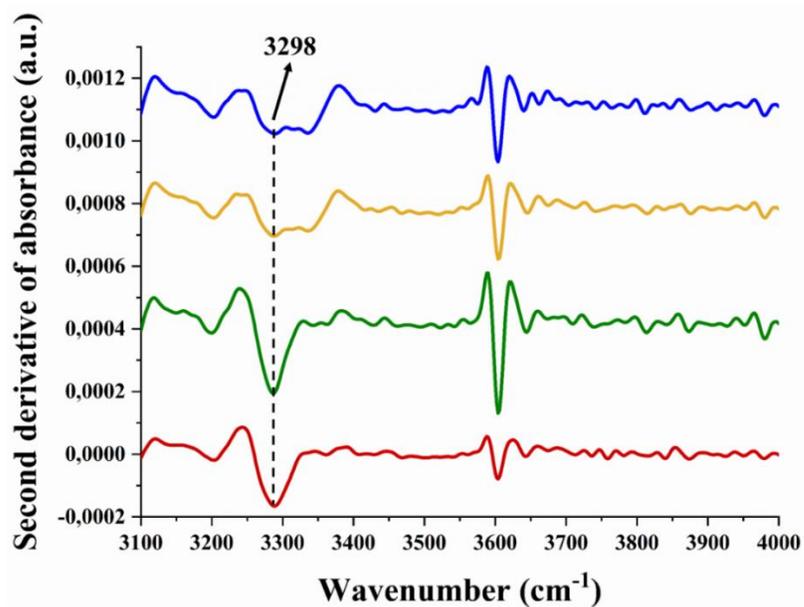


Figure 32. Averaged second derivatives calculated from high wavenumber region of spectral data from G1 (blue), G2 (yellow), G3 (green), and G4 (red).

Figure 31 shows the values of accuracy, sensitivity and specificity achieved by PC-LDA in discriminating data of healthy tissue from animals exposed to chemicals.

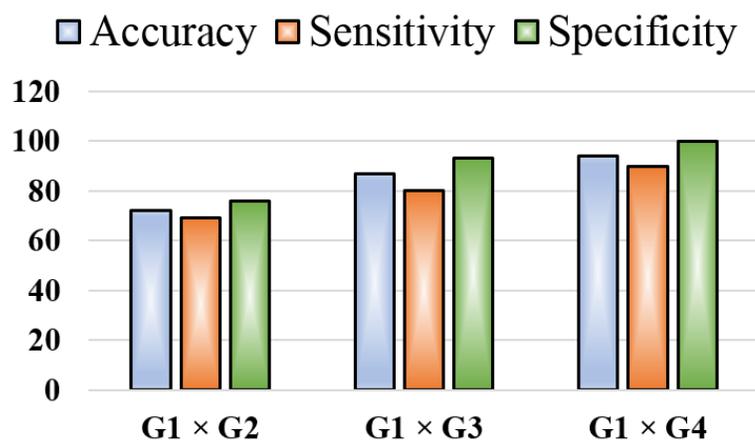


Figure 33. Statistical measurements obtained by PC-LDA as a binary classification test to discriminate healthy from groups exposed to carcinogenic agents (G1×G2, G1×G3, and G1×G4)

PC-LDA achieved performance above 80% for G1 × G3 and G1 × G4, and values over 70% for G1 × G2. The loadings and scores obtained for G1 × G3 and G1 × G4 are shown in Figure 34.

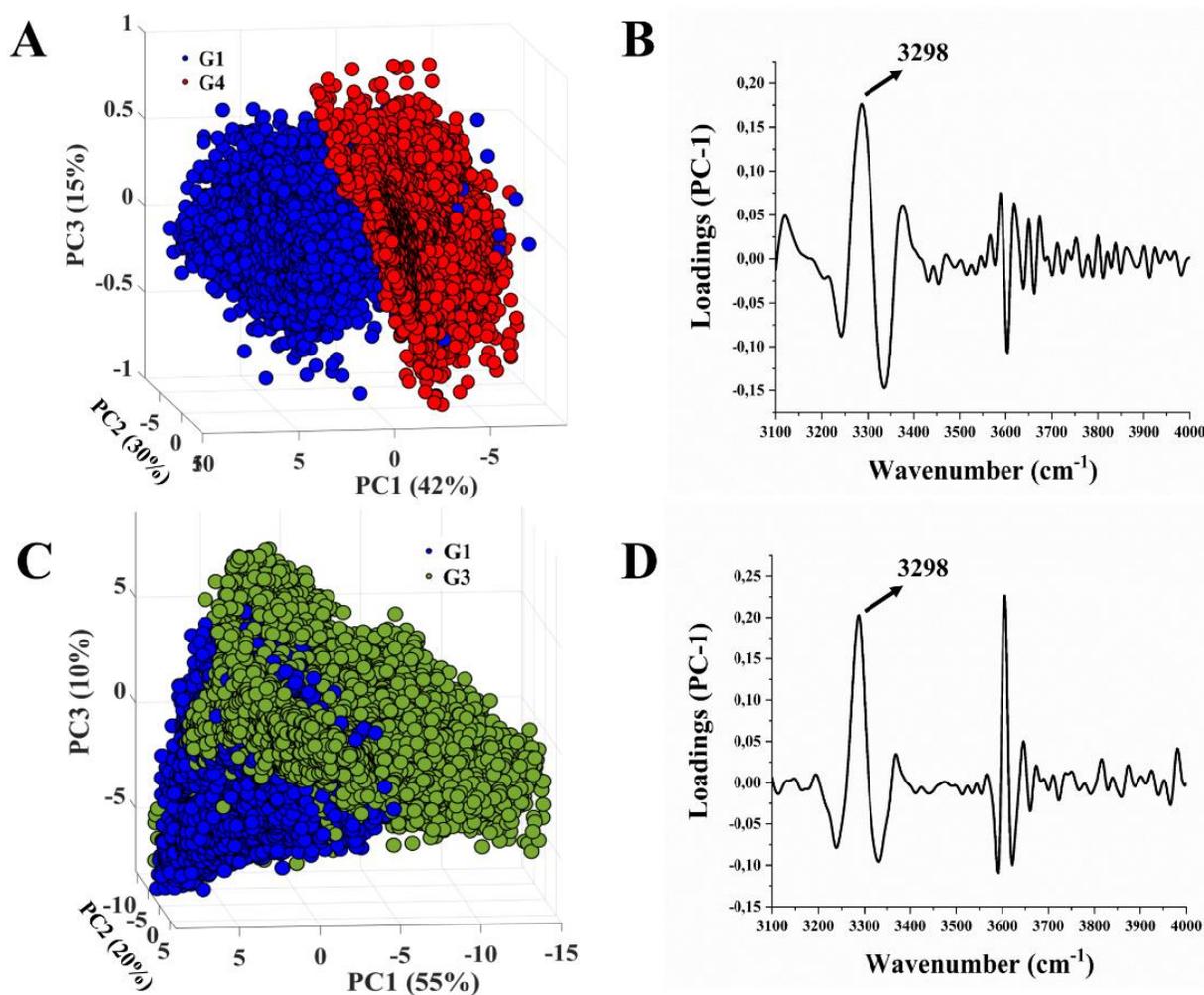


Figure 34. PCA results obtained comparing G1 to G2 and G3 using high wavenumber region. A and C show scores plots of G1×G4 and G1×G3, respectively. B and D show loadings plots of PC-1 obtained in each pairwise comparison.

PCA results indicated higher absorbance of Amide A in the groups submitted to carcinogenesis. Contrary to the results obtained in the fingerprint region, shifts on band peak position were not evidenced.

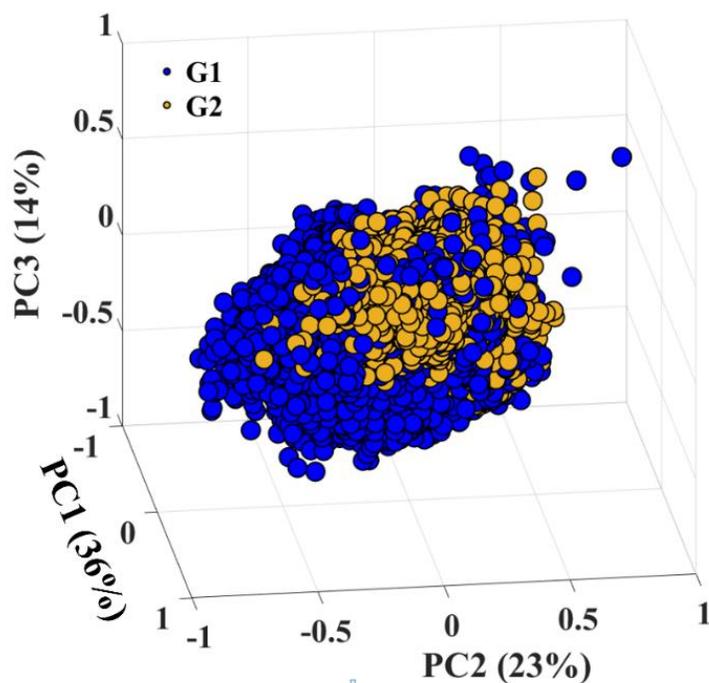


Figure 35. Scores plot obtained in the pairwise comparison of G1×G2 using the high wavenumber region of spectral data

As obtained for fingerprint region, the scores plot depicted in Figure 35 obtained poor discrimination comparing datasets from healthy group to animals exposed to carcinogenic factors for 8 weeks. Therefore, indicating high degree of biochemical similarity between animals from both groups.

5.3.3 Tissue specimens placed on glass substrates (high wavenumber spectral region)

Figure 36 depicts the high wavenumber region (3100-4000 cm^{-1}) of averaged spectra obtained from tissue placed on glass substrates.

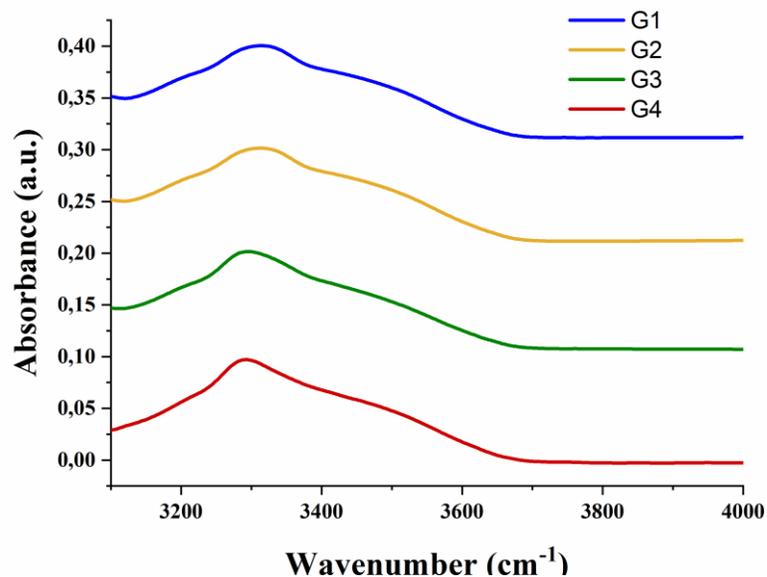


Figure 36. Averaged spectra calculated from high wavenumber region of spectra from G1 (blue), G2 (yellow), G3 (green), and G4 (red).

The performance obtained by PC-LDA in classifying data from healthy animals and groups that underwent to carcinogenesis is shown in Figure 37.

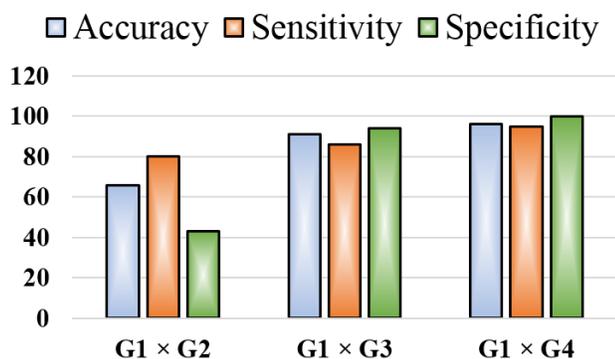


Figure 37. Statistical measurements obtained by PC-LDA as a binary classification test to discriminate healthy tissue from groups exposed to carcinogenic factors (G1×G2, G1×G3, and G1×G4).

Pairwise comparisons G1×G3 and G1×G4 presented performance over 80% and their scores and loadings plots are depicted by Figure 38.

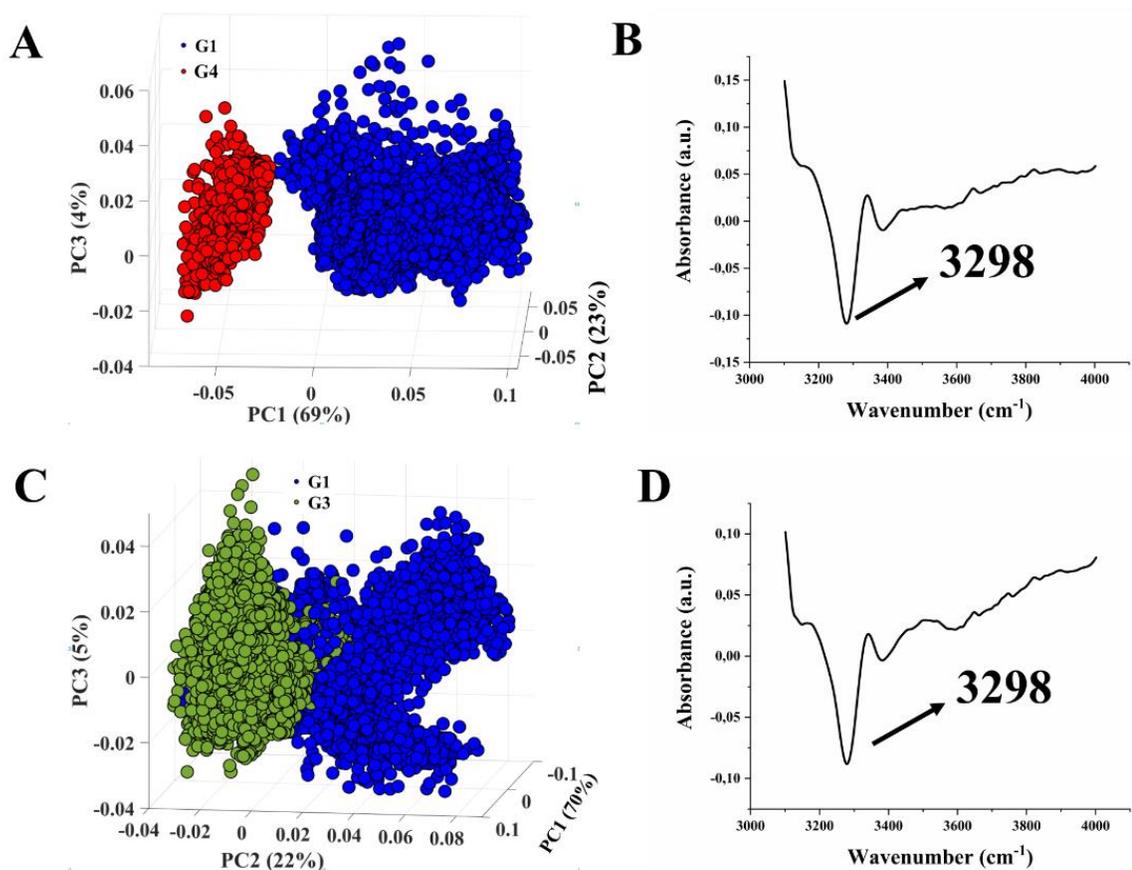


Figure 38. PCA results obtained comparing G1 to G2 and G3 using high wavenumber region collected from specimens placed on glass substrates. A and C show scores plots of G1×G4 and G1×G3, respectively. B and D show loadings plots of PC-1 obtained in each pairwise comparison.

The scores plots shown in Figure 38.A and C show that spectral data are grouped along different principal components in each pairwise comparison. For G1×G4, the scores are discriminated along PC-1, whereas in G1×G3 the second principal component achieved better results. Despite the distinct PCs in which the datasets were grouped, similar findings were evidenced assessing the loadings associated to each PC (Figure 38.B and D). Considering that spectra were employed as input data, positive loadings relate to positive scores and negative loadings are associated to negative scores. Thus, the Amide A vibration presented negative loadings in both plots, indicating higher absorbance in the groups submitted to carcinogenesis.

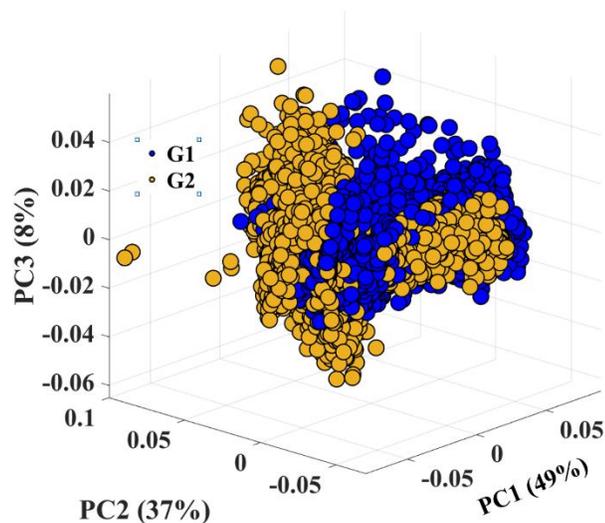


Figure 39. Scores plot obtained in the pairwise comparison of G1×G2 using the high wavenumber region of spectral data collected from samples placed on glass slides.

Figure 39 shows the scores plot obtained in the pairwise comparison G1×G2. Similar to the results obtained for tissue placed on CaF₂, healthy tissue compared to animals exposed to carcinogenic factors for 8 weeks presented poor discrimination, suggesting similar biochemical features in tissue collected from animals of both groups.

5.4. Discussion

Statistical measurements obtained by PC-LDA as a binary classification test in each pairwise comparison (G1×G2, G1×G3 and G1×G4) presented increasing values according to the period that animals were exposed to carcinogenic factors regardless the substrate used to place the specimens. The ability of PC-LDA in discriminating spectral data from different groups is associated to the level of biochemical similarity between the datasets compared. Thus, animals exposed to chemicals for 48 weeks presented the highest degree of biochemical changes compared to healthy tissue, while the animals that underwent to carcinogenesis protocol for only 8 weeks showed the lowest degree, i.e., higher resemblance to healthy features. These findings are associated to the biochemical changes that carcinogenic agents induce in tissue during the tumorigenesis process that lead to the appearance and progression of cancer. Histopathological examination classified the animals submitted to carcinogenesis for 16 and 48 weeks as

papilloma and invasive squamous cell carcinoma. On the other hand, histological findings indicating malignancy were not observed in tissue specimens from mice exposed to chemicals for 8 weeks. Thus, the ability of PC-LDA in correctly discriminate FTIR spectral data of animals exposed to carcinogenic agents from healthy skin is correlated to the emergence of cancer as well as to the stage of the disease.

Satisfactory data discrimination (over 80%) was achieved by PC-LDA comparing healthy skin to specimens of papilloma and iSCC (G1×G3 and G1×G4) regardless the substrate used to place the samples and spectral region. Thus, PCA loadings were evaluated in order to assess the biochemical changes responsible by data grouping. For samples placed on CaF₂, the vibrational modes of fingerprint region associated to collagen and elastin (1202, 1234, 1282 and 1338 cm⁻¹), as well as bands related to carbohydrates (1032, 1082 and 1164 cm⁻¹) showed higher absorption intensities in healthy tissue, therefore in reduced amounts in samples from papilloma and iSCC. Similar findings were obtained by Mantsch *et al*, who described such behavior as characteristic of malignant carcinomas⁸¹. Bodanese and coworkers associated the reduction of collagen vibrations to the degradation of collagen fibers in the extracellular matrix (ECM) of basal cell carcinoma⁸². The basic constituents of the ECM are fibrous proteins such collagen and elastin, and carbohydrates associated or not to other molecules (glycoproteins, glycosaminoglycans and proteoglycans)⁸³. Thus, spectral changes on vibrations of these molecules associated to the reduced connective tissue observed by histological examination in cancerous specimens suggest abnormalities and disruption of ECM. Several studies have demonstrated that deregulations in ECM play an important role in cancer by providing a tumorigenic microenvironment that lead to cancer progression^{84; 85}. Besides, ECM degradation is mediated by a family of enzymes (matrix metalloproteinases, MMPs, which have presented increased activity in several studies evaluating epithelial cancers induced by protocols similar to the used in our study (TPA/DMBA)^{83; 86; 87; 88}.

The abnormal features presented by cancer cells such uncontrolled proliferation, increased cell division, immortalization, and other characteristics are mediated by a complex system regulated by proteins. Thus, alterations in protein content may be expected due to the molecular events and signaling pathways that take place during carcinogenesis. PCA loadings revealed significant changes in chemical vibrations associated to proteins including shifts on band peak

position. The shifts identified in vibrational modes related to Amide II (1550 cm^{-1} in healthy tissue and 1562 cm^{-1} in groups exposed to carcinogenic factors) and α -helix structures (1660 cm^{-1} in healthy and 1652 cm^{-1} in animals that underwent to chemicals) indicate that cancerous tissue present proteins whose secondary structures are slightly different from proteins observed in tissue under normal conditions. These changes can be associated to the activation of new proteins required in the biological events related to cancer progression as well as mutations in the pre-existing proteins. In a study investigating the signatures of different histological structures of skin (epithelium, dermis, and stratum corneum), our group demonstrated that spectrum of epithelium is composed by bands associated to α -helix, but with no contributions of β -sheet vibrations⁸⁹. Therefore, the overexpression of α -helix and the significant decrease of β -sheet structures observed in specimens from papilloma and iSCC results from the intense proliferation of keratinocytes from epidermis invading surrounding tissue.

PC-LDA did not achieve satisfactory discrimination (over 80%) in classifying data from G1 and G2 regardless the substrate used to place the specimens. However, it is important to notice that the current gold standard methods classified all the specimens from mice exposed to chemicals for 8 weeks as healthy, and thus presented 100% of failure in identifying evidence of malignancy. On the other hand, statistical measurements obtained by PC-LDA ranged from 35-78%, which is superior to the obtained by gold standard techniques.

A key factor that may adversely affect the performance of PC-LDA is the inclusion of spectra extracted from healthy regions erroneously considered as altered by DMBA/TPA. Contrary to the animals from G3 and G4 groups, which developed visible cutaneous lesions, the skin of mice from G2 did not present clinical evidence of the disease. Thus, specimens from healthy regions may have been extracted instead of tissue with genetic/biochemical alterations. Although the chemicals have been applied throughout the dorsal part of the mice, the alterations induced by carcinogens arise in random regions. In addition, not all animals that are submitted to the carcinogenesis protocol develop the disease as described in Figure 5.B. Thus, if a specimen from G2 group was collected from an animal that was not affected by the chemicals, it will contribute with spectral signatures of healthy tissue and the performance of PC-LDA will be adversely affected.

5.5. Conclusions

In this chapter, FTIR spectral data acquired from healthy skin were compared to tissue specimens from animals submitted to different periods of exposure to carcinogenic factors in order to evaluate the ability of FTIR spectroscopy in staging skin cancer. Measurements were collected from samples placed on conventional glass and CaF₂ substrates aiming to evaluate the diagnostic potential of vibrational modes from high wavenumber region compared to fingerprint spectral range. Datasets from healthy skin were compared to each group using PCA and the performance of classification in each pairwise comparison was obtained via LDA model based on the PCA components (PC-LDA). Our findings indicate that the ability of PC-LDA in correctly classifying spectra of animals exposed to carcinogenic agents from healthy skin is associated to the stage of the disease regardless the substrate used to place the specimens. For both CaF₂ and glass substrates, the method presented poor results in differentiating healthy samples from specimens collected from mice submitted to carcinogenesis for 8 weeks, nevertheless, achieved satisfactory results in identifying animals exposed to chemicals for 16 (papilloma) and 48 (SCC) weeks. Thus, we demonstrate that spectra acquired from H&E stained samples on glass substrate provide valuable information in staging cancer, and thus the method can be used to complement histopathology in the study and diagnosis of cancerous diseases without requiring expensive substrates nor disrupting the pathologist's workflow.

Bibliography

- 1 ALBERTS, B. *Biologia Molecular da Célula*. 6. Artmed, 2017. 1464
- 2 COOPER, G. M. *Elements of Human Cancer*. Jones & Bartlett Learning, 1992. 368
- 3 INCA. Instituto Nacional do Câncer - INCA. 2019. Disponível em: < <http://www2.inca.gov.br/wps/wcm/connect/cancer/site/oquee> >. Acesso em: January, 3.
- 4 BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; SIEGEL, R. L.; TORRE, L. A.; JEMAL, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, v. 68, n. 6, p. 394-424, Nov 2018.
- 5 INCA. Estimativa de câncer no Brasil para 2018. **Instituto Nacional do Câncer**, <http://www.inca.gov.br/estimativa/2018/>, 2018. Disponível em: < <http://www.inca.gov.br/estimativa/2018/> >. Acesso em: 05/setembro.
- 6 KIM, R. H.; ARMSTRONG, A. W. Nonmelanoma skin cancer. *Dermatol Clin*, v. 30, n. 1, p. 125-39, ix, Jan 2012.
- 7 DUBAS, L. E.; INGRAFFEA, A. Nonmelanoma skin cancer. *Facial Plast Surg Clin North Am*, v. 21, n. 1, p. 43-53, Feb 2013.
- 8 WERNER, B. Biópsia de pele e seu estudo histológico: Por quê? Para quê? *Anais Brasileiros de Dermatologia*, v. 84, n. 4, p. 391-395, 2009.
- 9 YANOFSKY, V. R.; MERCER, S. E.; PHELPS, R. G. Histopathological variants of cutaneous squamous cell carcinoma: a review. *J Skin Cancer*, v. 2011, p. 210813, 2011.
- 10 KUMAR, V.; ABBAS, A. K.; ASTER, J. C. *Patologia - Bases patológicas das doenças*. Elsevier, 2016. 1440
- 11 DIEM, M.; MAZUR, A.; LENAU, K.; SCHUBERT, J.; BIRD, B.; MILJKOVIC, M.; KRAFFT, C.; POPP, J. Molecular pathology via IR and Raman spectral imaging. *J Biophotonics*, v. 6, n. 11-12, p. 855-86, Dec 2013.
- 12 FAQI, A. S. *A Comprehensive Guide to Toxicology in Preclinical Drug Development*. Academic Press, 2012. 1024

- 13 KALMODIA, S.; PARAMESWARAN, S.; YANG, W.; BARROW, C. J.; KRISHNAKUMAR, S. Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy: An analytical technique to understand therapeutic responses at the molecular level. *Sci Rep*, v. 5, p. 16649, Nov 16 2015.
- 14 PARASKEVAIDI, M.; MARTIN-HIRSCH, P. L.; MARTIN, F. L. ATR-FTIR Spectroscopy Tools for Medical Diagnosis and Disease Investigation. In: KUMAR, D. C. S. S. R. (Ed.). *Nanotechnology Characterization Tools for Biosensing and Medical Diagnosis*: Springer, 2018. p.163-211. ISBN 978-3-662-56333-5.
- 15 BONNIER, F.; BYRNE, H. J. Understanding the molecular information contained in principal component analysis of vibrational spectra of biological systems. *Analyst*, v. 137, n. 2, p. 322-32, Jan 21 2012.
- 16 LIMA, C. A.; GOULART, V. P.; CORREA, L.; PEREIRA, T. M.; ZECELL, D. M. ATR-FTIR spectroscopy for the assessment of biochemical changes in skin due to cutaneous squamous cell carcinoma. *Int J Mol Sci*, v. 16, n. 4, p. 6621-30, 2015.
- 17 SHAIKH, R.; PRABITHA, V. G.; DORA, T. K.; CHOPRA, S.; MAHESHWARI, A.; DEODHAR, K.; REKHI, B.; SUKUMAR, N.; KRISHNA, C. M.; SUBHASH, N. A comparative evaluation of diffuse reflectance and Raman spectroscopy in the detection of cervical cancer. *J Biophotonics*, v. 10, n. 2, p. 242-252, Feb 2017.
- 18 SUN, X.; XU, Y.; WU, J.; ZHANG, Y.; SUN, K. Detection of lung cancer tissue by attenuated total reflection-Fourier transform infrared spectroscopy-a pilot study of 60 samples. *J Surg Res*, v. 179, n. 1, p. 33-8, Jan 2013.
- 19 BIRD, B.; MILJKOVIC, M. S.; REMISZEWSKI, S.; AKALIN, A.; KON, M.; DIEM, M. Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer. *Lab Invest*, v. 92, n. 9, p. 1358-73, Sep 2012.
- 20 AKALIN, A.; ERGIN, A.; REMISZEWSKI, S.; MU, X.; RAZ, D.; DIEM, M. Resolving Interobserver Discrepancies in Lung Cancer Diagnoses by Spectral Histopathology. *Arch Pathol Lab Med*, Aug 24 2018.
- 21 ECKELA, R.; HUOA, H.; GUANB, H.; HUB, X.; CHEA, X.; HUANGA, W. Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue. *Vib Spectrosc*, v. 27, n. 2, p. 165-173, 2001.

- 22 BHATTACHARJEE, T.; KHAN, A.; KUMAR, P.; INGLE, A.; MARU, G.; KRISHNA, C. M. Raman spectroscopy of serum: A study on 'pre' and 'post' breast adenocarcinoma resection in rat models. *J Biophotonics*, v. 8, n. 7, p. 575-83, Jul 2015.
- 23 BASSAN, P.; MELLOR, J.; SHAPIRO, J.; WILLIAMS, K. J.; LISANTI, M. P.; GARDNER, P. Transmission FT-IR chemical imaging on glass substrates: applications in infrared spectral histopathology. *Anal Chem*, v. 86, n. 3, p. 1648-53, Feb 04 2014.
- 24 TOSI, G.; CONTI, C.; GIORGINI, E.; FERRARIS, P.; GARAVAGLIA, M. G.; SABBATINI, S.; STAIBANO, S.; RUBINI, C. FTIR microspectroscopy of melanocytic skin lesions: a preliminary study. *Analyst*, v. 135, n. 12, p. 3213-9, Dec 2010.
- 25 LIMA, C. A.; GOULART, V. P.; CORREA, L.; ZECELL, D. M. Using Fourier transform infrared spectroscopy to evaluate biological effects induced by photodynamic therapy. *Lasers Surg Med*, v. 48, n. 5, p. 538-45, Jul 2016.
- 26 HANDS, J. R.; DORLING, K. M.; ABEL, P.; ASHTON, K. M.; BRODBELT, A.; DAVIS, C.; DAWSON, T.; JENKINSON, M. D.; LEA, R. W.; WALKER, C.; BAKER, M. J. Attenuated total reflection fourier transform infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples. *J Biophotonics*, v. 7, n. 3-4, p. 189-99, Apr 2014.
- 27 SANDT, C.; NADARADJANE, C.; RICHARDS, R.; DUMAS, P.; SEE, V. Use of infrared microspectroscopy to elucidate a specific chemical signature associated with hypoxia levels found in glioblastoma. *Analyst*, v. 141, n. 3, p. 870-83, Feb 07 2016.
- 28 PILLING, M. J.; HENDERSON, A.; SHANKS, J. H.; BROWN, M. D.; CLARKE, N. W.; GARDNER, P. Infrared spectral histopathology using haematoxylin and eosin (H&E) stained glass slides: a major step forward towards clinical translation. *Analyst*, v. 142, n. 8, p. 1258-1268, Apr 10 2017.
- 29 BAKER, M. J.; GAZI, E.; BROWN, M. D.; SHANKS, J. H.; CLARKE, N. W.; GARDNER, P. Investigating FTIR based histopathology for the diagnosis of prostate cancer. *J Biophotonics*, v. 2, n. 1-2, p. 104-13, Feb 2009.

- 30 HUGHES, C.; GAUNT, L.; BROWN, M.; CLARKE, N. W.; GARDNER, P. Assessment of paraffin removal from prostate FFPE sections using transmission mode FTIR-FPA imaging. *Anal. Methods*, v. 6, n. 4, p. 1028-1035, 2014.
- 31 DE LIMA, F. A.; GOBINET, C.; SOCKALINGUM, G.; GARCIA, S. B.; MANFAIT, M.; UNTEREINER, V.; PIOT, O.; BACHMANN, L. Digital de-waxing on FTIR images. *Analyst*, v. 142, n. 8, p. 1358-1370, Apr 10 2017.
- 32 DONG, L.; SUN, X.; CHAO, Z.; ZHANG, S.; ZHENG, J.; GURUNG, R.; DU, J.; SHI, J.; XU, Y.; ZHANG, Y.; WU, J. Evaluation of FTIR spectroscopy as diagnostic tool for colorectal cancer using spectral analysis. *Spectrochim Acta A Mol Biomol Spectrosc*, v. 122, p. 288-94, Mar 25 2014.
- 33 PEREIRA, T. M.; ZECELL, D. M.; BIRD, B.; MILJKOVIC, M.; DIEM, M. The characterization of normal thyroid tissue by micro-FTIR spectroscopy. *Analyst*, v. 138, n. 23, p. 7094-100, Dec 07 2013.
- 34 PARASKEVAIDIA, M.; MORAISB, 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.; MARTIN, F. L. Differential diagnosis of Alzheimer's disease using spectrochemical analysis of blood. *Proc Natl Acad Sci USA*, v. 114, n. 38, p. 10, 2017.
- 35 SCOTT, D. A.; RENAUD, D. E.; KRISHNASAMY, S.; MERIC, P.; BUDUNELI, N.; CETINKALP, S.; LIU, K. Z. Diabetes-related molecular signatures in infrared spectra of human saliva. *Diabetol Metab Syndr*, v. 2, p. 48, Jul 14 2010.
- 36 SEVERCAN, F.; BOZKURT, O.; GURBANOV, R.; GORGULU, G. FT-IR spectroscopy in diagnosis of diabetes in rat animal model. *J Biophotonics*, v. 3, n. 8-9, p. 621-31, Aug 2010.
- 37 DERENNE, A.; VAN HEMELRYCK, V.; LAMORAL-THEYYS, D.; KISS, R.; GOORMAGHTIGH, E. FTIR spectroscopy: a new valuable tool to classify the effects of polyphenolic compounds on cancer cells. *Biochim Biophys Acta, Mol Basis Dis*, v. 1832, n. 1, p. 46-56, Jan 2013.
- 38 FALE, P. L.; ALTHARAWI, A.; CHAN, K. L. In situ Fourier transform infrared analysis of live cells' response to doxorubicin. *Biochim Biophys Acta, Mol Cell Res*, v. 1853, n. 10 Pt A, p. 2640-8, Oct 2015.

- 39 BATISTA DE CARVALHO, A. L. M.; PILLING, M.; GARDNER, B.; DOHERTY, J.; CINQUE, G.; WEHBE, K.; KELLEY, C.; BATISTA DE CARVALHO, L. A. E.; MARQUES, M. P. M. Chemotherapeutic response to cisplatin-like drugs in human breast cancer cells probed by vibrational microspectroscopy. *Faraday Discuss*, v. 187, p. 273-289, 2016.
- 40 GIORGINI, E.; SABBATINI, S.; ROCCHETTI, R.; NOTARSTEFANO, V.; RUBINI, C.; CONTI, C.; ORILISI, G.; MITRI, E.; BEDOLLA, D. E.; VACCARI, L. In vitro FTIR microspectroscopy analysis of primary Oral Squamous Carcinoma cells treated with cisplatin and 5-fluorouracil: a new spectroscopic approach for studying drugcell interaction *Analyst*, v. 143, n. 14, p. 3317-3326, 2018.
- 41 NOWOTARSKI, S. L.; FEITH, D. J.; SHANTZ, L. M. Skin Carcinogenesis Studies Using Mouse Models with Altered Polyamines. *Cancer Growth Metastasis*, v. 8, n. Suppl 1, p. 17-27, 2015.
- 42 KHAVARI, P. A. Modelling cancer in human skin tissue. *Nat Rev Cancer*, v. 6, n. 4, p. 270-80, Apr 2006.
- 43 ABEL, E. L.; ANGEL, J. M.; KIGUCHI, K.; DIGIOVANNI, J. Multi-stage chemical carcinogenesis in mouse skin: fundamentals and applications. *Nat Protoc*, v. 4, n. 9, p. 1350-62, 2009.
- 44 SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. *Principles of Instrumental Analysis*. Thomson Brooks/Cole, 2016.
- 45 LIMA, F. A. D. **Infrared microspectroscopy of inflammatory process and colon tumors**. 2016. (Ph.D.). Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Brasil.
- 46 BASSAN, P. **Light scattering during infrared spectroscopic measurements of biomedical samples**. 2011. (Ph.D.). Faculty of Engineering and Physical Sciences, The University of Manchester, UK.
- 47 BAKER, M. J.; TREVISAN, J.; BASSAN, P.; BHARGAVA, R.; BUTLER, H. J.; DORLING, K. M.; FIELDEN, P. R.; FOGARTY, S. W.; FULLWOOD, N. J.; HEYS, K. A.; HUGHES, C.; LASCH, P.; MARTIN-HIRSCH, P. L.; OBINAJU, B.; SOCKALINGUM, G. D.; SULE-SUSO, J.; STRONG, R. J.; WALSH, M. J.; WOOD, B. R.; GARDNER, P.; MARTIN, F. L. Using Fourier transform IR spectroscopy to analyze biological materials. *Nat Protoc*, v. 9, n. 8, p. 1771-91, Aug 2014.

- 48 BUTLER, H. J.; ASHTON, L.; BIRD, B.; CINQUE, G.; CURTIS, K.; DORNEY, J.; ESMONDE-WHITE, K.; FULLWOOD, N. J.; GARDNER, B.; MARTIN-HIRSCH, P. L.; WALSH, M. J.; MCAINSH, M. R.; STONE, N.; MARTIN, F. L. Using Raman spectroscopy to characterize biological materials. *Nat Protoc*, v. 11, n. 4, p. 664-87, Apr 2016.
- 49 KAZARIAN, S. G.; CHAN, K. L. ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems. *Analyst*, v. 138, n. 7, p. 1940-51, Apr 07 2013.
- 50 ANDREW CHAN, K. L.; KAZARIAN, S. G. Attenuated total reflection Fourier-transform infrared (ATR-FTIR) imaging of tissues and live cells. *Chem Soc Rev*, v. 45, n. 7, p. 1850-64, Apr 7 2016.
- 51 DORLING, K. M.; BAKER, M. J. Highlighting attenuated total reflection Fourier transform infrared spectroscopy for rapid serum analysis. *Trends Biotechnol*, v. 31, n. 6, p. 327-8, Jun 2013.
- 52 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*, v. 8, n. 3, p. 481-487, 2016.
- 53 BASSAN, P.; SACHDEVA, A.; KOHLER, A.; HUGHES, C.; HENDERSON, A.; BOYLE, J.; SHANKS, J. H.; BROWN, M.; CLARKE, N. W.; GARDNER, P. FTIR microscopy of biological cells and tissue: data analysis using resonant Mie scattering (RMieS) EMSC algorithm. *Analyst*, v. 137, n. 6, p. 1370-7, Mar 21 2012.
- 54 BASSAN, P.; KOHLER, A.; MARTENS, H.; LEE, J.; BYRNE, H. J.; DUMAS, P.; GAZI, E.; BROWN, M.; CLARKE, N.; GARDNER, P. Resonant Mie scattering (RMieS) correction of infrared spectra from highly scattering biological samples. *Analyst*, v. 135, n. 2, p. 268-77, Feb 2010.
- 55 FILIK, J.; FROGLEY, M. D.; PIJANKA, J. K.; WEHBE, K.; CINQUE, G. Electric field standing wave artefacts in FTIR micro-spectroscopy of biological materials. *Analyst*, v. 137, n. 4, p. 853-61, Feb 21 2012.
- 56 WROBEL, T. P.; WAJNCHOLD, B.; BYRNE, H. J.; BARANSKA, M. Electric field standing wave effects in FT-IR transflection spectra of biological tissue sections: Simulated models of experimental variability. *Vibrational Spectroscopy*, v. 69, p. 84-92, 2013.

- 57 BYRNE, H. J.; KNIEF, P.; KEATING, M. E.; BONNIER, F. Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. *Chem Soc Rev*, v. 45, n. 7, p. 1865-78, Apr 07 2016.
- 58 STANISZEWSKA-SLEZAK, E.; RYGULA, A.; MALEK, K.; BARANSKA, M. Transmission versus transflection mode in FTIR analysis of blood plasma: is the electric field standing wave effect the only reason for observed spectral distortions? *Analyst*, v. 140, n. 7, p. 2412-21, Apr 07 2015.
- 59 BASSAN, P.; BYRNE, H. J.; BONNIER, F.; LEE, J.; DUMAS, P.; GARDNER, P. Resonant Mie scattering in infrared spectroscopy of biological materials--understanding the 'dispersion artefact'. *Analyst*, v. 134, n. 8, p. 1586-93, Aug 2009.
- 60 AFSETH, N. K.; KOHLER, A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. *Chemometrics and Intelligent Laboratory Systems*, v. 117, p. 92-99, 2012.
- 61 MIAN, S. A.; COLLEY, H. E.; THORNHILL, M. H.; REHMAN, I. U. Development of a Dewaxing Protocol for Tissue-Engineered Models of the Oral Mucosa Used for Raman Spectroscopic Analysis. *Applied Spectroscopy Reviews*, v. 49, n. 8, p. 614-617, 2014.
- 62 LASCH, P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometrics and Intelligent Laboratory Systems*, v. 117, p. 100-114, 2012.
- 63 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. *Applied Spectroscopy Reviews*, v. 52, n. 5, p. 456-506, 2016.
- 64 MOVASAGHI, Z.; REHMAN, S.; UR REHMAN, D. I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Appl Spectrosc Rev*, v. 43, n. 2, p. 134-179, 2008.
- 65 WIERCIGROCH, E.; SZAFRANIEC, E.; CZAMARA, K.; PACIA, M. Z.; MAJZNER, K.; KOCHAN, K.; KACZOR, A.; BARANSKA, M.; MALEK, K. Raman and infrared spectroscopy of carbohydrates: A review. *Spectrochim Acta A Mol Biomol Spectrosc*, v. 185, p. 317-335, Oct 5 2017.

- 66 BELBACHIR, K.; NOREEN, R.; GOUSPILLOU, G.; PETIBOIS, C. Collagen types analysis and differentiation by FTIR spectroscopy. *Anal Bioanal Chem*, v. 395, n. 3, p. 829-37, Oct 2009.
- 67 CHEHELTANI, R.; MCGOVERIN, C. M.; RAO, J.; VORP, D. A.; KIANI, M. F.; PLESHKO, N. Fourier transform infrared spectroscopy to quantify collagen and elastin in an in vitro model of extracellular matrix degradation in aorta. *Analyst*, v. 139, n. 12, p. 3039-47, Jun 21 2014.
- 68 BARTH, A. Infrared spectroscopy of proteins. *Biochim Biophys Acta, Bioenerg*, v. 1767, n. 9, p. 1073-101, Sep 2007.
- 69 BARTH, A.; ZSCHERP, C. What vibrations tell about proteins. *Quarterly Reviews of Biophysics*, v. 35, n. 4, p. 369-430, 2002.
- 70 ABDI, H.; WILLIAMS, L. J. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, v. 2, n. 4, p. 433-459, 2010.
- 71 FERREIRA, M. M. C. **Quimiometria: Conceitos, métodos e aplicações:** Editora Unicamp: 496 p. 2015.
- 72 ALTURKISTANI, H. A.; TASHKANDI, F. M.; MOHAMMEDSALEH, Z. M. Histological Stains: A Literature Review and Case Study. *Glob J Health Sci*, v. 8, n. 3, p. 72-9, Jun 25 2015.
- 73 HIMMEL, L. E.; HACKETT, T. A.; MOORE, J. L.; ADAMS, W. R.; THOMAS, G.; NOVITSKAYA, T.; CAPRIOLI, R. M.; ZIJLSTRA, A.; MAHADEVAN-JANSEN, A.; BOYD, K. L. Beyond the H&E: Advanced Technologies for in situ Tissue Biomarker Imaging. *ILAR Journal*, 2018.
- 74 FERNANDEZ, D. C.; BHARGAVA, R.; HEWITT, S. M.; LEVIN, I. W. Infrared spectroscopic imaging for histopathologic recognition. *Nat Biotechnol*, v. 23, n. 4, p. 469-74, Apr 2005.
- 75 LASCH, P.; HAENSCH, W.; NAUMANN, D.; DIEM, M. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim Biophys Acta*, v. 1688, n. 2, p. 176-86, Mar 2 2004.
- 76 KOBRINA, Y.; TURUNEN, M. J.; SAARAKKALA, S.; JURVELIN, J. S.; HAUTA-KASARI, M.; ISAKSSON, H. Cluster analysis of infrared spectra of rabbit cortical bone samples during maturation and growth. *Analyst*, v. 135, n. 12, p. 3147-55, Dec 2010.

- 77 ZHONG, Q.; CHENYANG; GROBERÜSCHKAMP, F.; KALLENBACH-THIELTGES, A.; SEROCKA, P.; GERWERTAND, K.; MOSI, A. Similarity maps and hierarchical clustering for annotating FT-IR spectral images. *BMC Bioinformatics*, v. 14, n. 333, p. 1-11, 2013.
- 78 PIJANKA, J. K.; STONE, N.; CINQUE, G.; YANG, Y.; KOHLER, A.; WEHBE, K.; FROGLEY, M.; PARKES, G.; PARKES, J.; DUMAS, P.; SANDT, C.; VAN PITTIUS, D. G.; DOUCE, G.; SOCKALINGUM, G. D.; SULÉ-SUSO, J. FTIR microspectroscopy of stained cells and tissues. Application in cancer diagnosis. *Spectroscopy*, v. 24, p. 73–78, 2010.
- 79 PILLING, M.; GARDNER, P. Fundamental developments in infrared spectroscopic imaging for biomedical applications. *Chem Soc Rev*, v. 45, n. 7, p. 1935-57, Apr 07 2016.
- 80 BYRNE, H. J.; BARANSKA, M.; PUPPELS, G. J.; STONE, N.; WOOD, B.; GOUGH, K. M.; LASCH, P.; HERAUD, P.; SULE-SUSO, J.; SOCKALINGUM, G. D. Spectropathology for the next generation: quo vadis? *Analyst*, v. 140, n. 7, p. 2066-73, Apr 07 2015.
- 81 JACKSON, M.; CHOO, L.; WATSON, P. H.; HALLIDAY, W. C.; MANTSCH, H. H. Beware of connective tissue proteins: Assignment and implications of collagen absorptions in infrared spectra of human tissues. *Biochim Biophys Acta, Mol Basis Dis*, v. 1270, n. 1, p. 1-6, 1995.
- 82 BODANESE, B.; SILVEIRA, F. L.; ZANGARO, R. A.; PACHECO, M. T.; PASQUALUCCI, C. A.; SILVEIRA, L., JR. Discrimination of basal cell carcinoma and melanoma from normal skin biopsies in vitro through Raman spectroscopy and principal component analysis. *Photomed Laser Surg*, v. 30, n. 7, p. 381-7, Jul 2012.
- 83 GIALELI, C.; THEOCHARIS, A. D.; KARAMANOS, N. K. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *FEBS J*, v. 278, n. 1, p. 16-27, Jan 2011.
- 84 AFRATIS, N.; GIALELI, C.; NIKITOVIC, D.; TSEGENIDIS, T.; KAROUSOU, E.; THEOCHARIS, A. D.; PAVAO, M. S.; TZANAKAKIS, G. N.; KARAMANOS, N. K. Glycosaminoglycans: key players in cancer cell biology and treatment. *FEBS J*, v. 279, n. 7, p. 1177-97, Apr 2012.
- 85 LU, P.; WEAVER, V. M.; WERB, Z. The extracellular matrix: a dynamic niche in cancer progression. *J Cell Biol*, v. 196, n. 4, p. 395-406, Feb 20 2012.

- ⁸⁶ LIEBEL, F.; KAUR, S.; RUVOLO, E.; KOLLIAS, N.; SOUTHALL, M. D. Irradiation of skin with visible light induces reactive oxygen species and matrix-degrading enzymes. *J Invest Dermatol*, v. 132, n. 7, p. 1901-7, Jul 2012.
- ⁸⁷ DECOCK, J.; THIRKETTLE, S.; WAGSTAFF, L.; EDWARDS, D. R. Matrix metalloproteinases: protective roles in cancer. *J Cell Mol Med*, v. 15, n. 6, p. 1254-65, Jun 2011.
- ⁸⁸ FUNDYLER, O.; KHANNA, M.; SMOLLER, B. R. Metalloproteinase-2 expression correlates with aggressiveness of cutaneous squamous cell carcinomas. *Mod Pathol*, v. 17, n. 5, p. 496-502, May 2004.
- ⁸⁹ LIMA, C. A.; CORREA, L.; BYRNE, H. J.; ZECELL, D. M. **Assessing the spectrochemical signatures of skin components using FTIR microspectroscopy.** SPIE/OSA European Conferences on Biomedical Optics. Munich, Germany: SPIE/OSA, 2019

APPROVAL FOR COMMITTEE ON ANIMAL RESEARCH AND ETHICS



Comissão de Ética no Uso de Animais

Certificamos que o projeto intitulado “**AVALIAÇÃO BIOQUÍMICA VIA ESPECTROSCOPIA VIBRACIONAL DE CARCINOMA ESPINOCELULAR SUBMETIDO A TERAPIA FOTODINÂMICA**” protocolo nº **164/15** sob a responsabilidade de **DENISE MARIA ZEZELL**– que envolve a produção, manutenção e/ou utilização de animais pertencentes ao filo Chordata, subfilo Vertebrata (exceto o homem) – encontra-se de acordo com os preceitos da Lei nº 11.794, de 8 de outubro de 2008, do Decreto nº 6.899, de 15 de julho de 2009, e com as normas editadas pelo Conselho Nacional de Controle da Experimentação Animal (CONCEA), e foi **APROVADO** pela COMISSÃO DE ÉTICA NO USO DE ANIMAIS (CEUAIPEN) do Instituto de Pesquisas Energéticas e Nucleares (IPEN – CNEN/SP), em reunião de 23/09/2015 .

| | |
|---------------------|--|
| Vigência do Projeto | Setembro/2015 a Setembro/2017 |
| Espécie/linhagem | Camundongo/Swiss |
| Nº de animais | 84 animais |
| Peso/Idade | Peso: 20g e Idade: 8 a 10 semanas |
| Sexo | Fêmeas |
| Origem | Biotério do IPEN – CNEN/SP |

São Paulo, 05 de outubro de 2015.



Prof. Dra. Nanci do Nascimento
Coordenadora da CEUA-IPEN

IPEN-CNEN/SP
COMITÊ DE ÉTICA NO USO DE ANIMAIS
Av Prof Lineu Prestes, 2242 - Cidade Universitária - CEP 05508-000 - São Paulo - SP
Telefone: (011) 3133-9696 - Fax (011) 3133-9709
E-mail: ceuaipen@ipen.br

LIST OF PUBLICATIONS

Articles published in Scientific Journals

1. Lima, C. A.; Correa, L.; Bechara, E. J. H.; Zezell, D. Optimization and therapeutic effects of PDT mediated by ALA and MAL in the treatment of cutaneous malignant lesions: A comparative study. **Journal of Biophotonics**, v. 9, p. 1355-1361, 2016.
2. Lima, C. A.; Goulart, V.; Correa, L.; Zezell, D. Using Fourier transform Infrared spectroscopy to evaluate biological effects induced by Photodynamic Therapy. **Lasers in Surgery and Medicine**, v. 48, p. 538-545, 2016.
3. Lima, C. A.; Goulart, V.; Correa, L.; Pereira, T.; Zezell, D. ATR-FTIR Spectroscopy for the Assessment of Biochemical Changes in Skin Due to Cutaneous Squamous Cell Carcinoma. **International Journal of Molecular Sciences**, v. 16, p. 6621-6630, 2015.

Articles being prepared to submit to Scientific Journals

4. Lima, C. A.; Corrêa, L.; Diem, M.; Zezell, D. M. Infrared spectroscopy revealed similar biochemical changes in cutaneous dysplastic lesions treated with photodynamic therapy (PDT) mediated by aminolaevulinic acid (ALA) or aminolevulinate (MAL).
5. Lima, C. A.; Byrne, H.; Correa, L.; Zezell, D. FTIR image segmentation for histopathologic recognition.
6. Lima, C. A.; Byrne, H.; Correa, L.; Zezell, D. Staging non-melanoma skin cancers via FTIR microspectroscopy using samples placed on conventional glass.

Conference proceedings

1. **Lima, C. A.**; Correa, L. ; Byrne, H. J. ; Zezell, D. M. Diagnosis of advanced skin cancer using Infrared spectral histopathology. In: *Latin America Optics and Photonics Conference-LAOP*, Lima, Peru, 2018.

2. **Lima, C. A.;** Correa, L. ; Byrne, H. J. ; Zezell, D. M. K-means and Hierarchical Cluster Analysis as segmentation algorithms of FTIR hyperspectral images collected from cutaneous tissue. In: *SBFoton Conference*, Campinas, Brazil, 2018.
3. Zezell, D.M.; **Lima, C. A.** Discrimination of Healthy Skin and Cutaneous Malignant Lesions using FTIR Spectra and their Second Derivatives: A Comparative Study. In: *OSA Biophotonics Congress: Biomedical Optics Congress*, Hollywood, USA, 2018.
4. **Lima, C. A.;** Goulart, V. P.; Zezell, D. M. Chemometric methods applied to FTIR spectra to discriminate treated and non-treated cutaneous malignant lesions from healthy skin. In: *Latin America Optics and Photonics Conference-LAOP*, Medellin, Colombia, 2016.
5. **Lima, C. A.;** Goulart, V. P.; Castro, P. A. A.; Côrrea, L.; Benetti, C.; Zezell, D. M. Biochemical changes in cutaneous squamous cell carcinoma submitted to PDT using ATR-FTIR spectroscopy. In: *SPIE Biophotonics South America*, Rio de Janeiro, Brazil, 2015.
6. **Lima, C. A.;** Correa, L. ; Byrne, H. J. ; Zezell, D. M. FTIR spectroscopy: a valuable tool to diagnose cutaneous tumors. In: *Next frontiers to cure cancer*, Sao Paulo, Brazil, 2018.
7. **Lima, C. A.;** Correa, L. ; Byrne, H. J. ; Zezell, D. M. FTIR microspectroscopy discriminating skin cancer using tissue sections on glass. In: *The International Conference on Clinical Vibrational Spectroscopy-SPEC*, Glasgow, United Kingdom, 2018.
8. **Lima, C. A.;** Correa, L.; Zezell, D. M. Infrared spectroscopy determining the biochemical changes in premalignant skin lesions submitted to photodynamic therapy. In: *FTIR Spectroscopy in Microbiological and Medical Diagnostics*, Berlin, Germany, 2017.
9. **Lima, C. A.;** goulart, V. P.; zezell, D. M. Laser-induced fluorescence spectroscopy assess the Protoporphyrin IX generated by 5-aminolevulinic acid and its methyl ester in cutaneous malignant lesions. In: *National Physics Meeting*, Natal, Brazil, 2016.
10. **Lima, C. A.;** Goulart, V. P.; Zezell, D. M. FTIR for the study of the biochemical changes promoted by PDT mediated by 5-ALA and its methyl-ester in neoplastic skin lesions. In: *The International Conference on Clinical Vibrational Spectroscopy-SPEC*, Montreal, Canada, 2016.

11. **Lima, C. A.;** Goulart, V. P.; Côrrea, L.; Zezell, D. M. Biochemical changes in normal skin caused by squamous cell carcinoma using FTIR spectroscopy. In: *Annual Conference of American Society for Lasers in Surgery and Medicine*, Kissimmee, USA, 2015.
12. **Lima, C. A.;** Goulart, V. P.; Castro, P. A. A.; Correa, L.; Zezell, D. M.. Can FTIR spectroscopy evaluate the biochemical changes in cutaneous squamous cell carcinoma caused by photodynamic effect? In: *16th European Conference on the Spectroscopy of Biological Molecules*, Bochum, Germany, 2015.

MATLAB SCRIPTS USED TO ANALYZE DATA

All pre-processing steps and spectral analysis performed in this study were carried out on Matlab® R2015 (MathWorks, Natick, MA, USA).

1. Importing spectral data collected on Perkin Elmer FTIR spectrophotometer

FTIR spectral data acquired on imaging mode using a Perkin Elmer spectrophotometer can be imported/converted to Matlab platform using the functions and scripts available online in the website below:

<https://la.mathworks.com/matlabcentral/fileexchange/22736-perkinelmer-ir-data-file-import-tools>

2. RMieS-EMSC algorithm

The RMieS-EMSC algorithm used to correct spectral distortions induced by resonant and non-resonant Mie scattering can be downloaded from the website of The International Society for Clinical Spectroscopy (CLIRSPEC):

<https://clirspec.org/>

3. k-means Cluster Analysis (KMCA)

The code below describes how to perform KMCA in FTIR datasets. As a condition of use we ask that you reference a the following paper:

“ALI, S. M.; BONNIER, F.; LAMBKIN, H.; FLYNN, K.; MCDONAGH, V.; HEALY, C.; LEE, T. C.; LYNG, F. M.; BYRNE, H. J. A comparison of Raman, FTIR and ATR-FTIR micro spectroscopy for imaging human skin tissue sections. *Analytical Methods*, v. 5, n. 9, 2013.”

Each row of the matrix **Absorbance** represents one spectrum, whereas **wavenumber** is a vector containing the wavenumbers related to each spectral point:

```
Absorbance = [FTIR_data];  
wavenumber = [wavenumber];  
  
% Savitzky-Golay smoothing  
Absorbance = (Absorbance)';
```

```

Absorbance_Smoothing_1 = sgolayfilt(Absorbance,2,11);
Absorbance_Smoothing_2 = sgolayfilt(Absorbance_Smoothing_1,2,11);
Absorbance_Smoothing_3 = sgolayfilt(Absorbance_Smoothing_2,2,11);
Absorbance_Final = (Absorbance_Smoothing_3)';

%%%%%%%%%%%% IR CHEMICAL IMAGE %%%%%%%%%%%%%%

for i = 1:length(Absorbance_Final(:,1))
B1(i,:) = mean(Absorbance_Final(i,1:end));
end
C = reshape(B1, length(yAxis), length(xAxis));

% yAxis and xAxis are variables obtained after importing FTIR spectral images
using the code described in "1. Importing spectral data collected on Perkin
Elmer FTIR spectrophotometer"

%%%%%%%%%%%% KMEANS CLUSTERING ANALYSIS %%%%%%%%%%%%%%

% 2 clusters image
IDX2 = kmeans(Absorbance_Final,2,'replicate',10,'MaxIter',1000);
D2 = reshape(IDX2, length(yAxis), length(xAxis), 1);

% 3 clusters image
IDX3 = kmeans(Absorbance_Final,3,'replicate',10,'MaxIter',1000);%
D3 = reshape(IDX3, length(yAxis), length(xAxis), 1);

% 4 clusters image
IDX4 = kmeans(Absorbance_Final,4,'replicate',10,'MaxIter',1000);
D4 = reshape(IDX4, length(yAxis), length(xAxis), 1);

% 5 clusters image
IDX5 = kmeans(Absorbance_Final,5,'replicate',10,'MaxIter',1000);
D5 = reshape(IDX5, length(yAxis), length(xAxis), 1);

% 6 clusters image
IDX6 = kmeans(Absorbance_Final,6,'replicate',10,'MaxIter',1000);
D6 = reshape(IDX6, length(yAxis), length(xAxis), 1);

% 7 clusters image
IDX7 = kmeans(Absorbance_Final,7,'replicate',10,'MaxIter',1000);
D7 = reshape(IDX7, length(yAxis), length(xAxis), 1);

% 8 clusters image
IDX8 = kmeans(Absorbance_Final,8,'replicate',10,'MaxIter',1000);
D8 = reshape(IDX8, length(yAxis), length(xAxis), 1);

% 9 clusters image
IDX9 = kmeans(Absorbance_Final,9,'replicate',10,'MaxIter',1000);
D9 = reshape(IDX9, length(yAxis), length(xAxis), 1);

% 10 clusters image
IDX10 = kmeans(Absorbance_Final,10,'replicate',10,'MaxIter',1000);
D10 = reshape(IDX10, length(yAxis), length(xAxis), 1);

```

```
%%%%%%%%%% KMCA RECONSTRUCTED IMAGES %%%%%%%%%%
```

```
F1 = figure;
```

```
subplot(3,3,1)  
imagesc(C);  
colorbar  
title('Chemical Map')
```

```
subplot(3,3,2)  
imagesc(D3(:, :, 1))  
colorbar  
title('3 clusters')
```

```
subplot(3,3,3)  
imagesc(D4(:, :, 1))  
colorbar  
title('4 clusters')
```

```
subplot(3,3,4)  
imagesc(D5(:, :, 1))  
colorbar  
title('5 clusters')
```

```
subplot(3,3,5)  
imagesc(D6(:, :, 1))  
colorbar  
title('6 clusters')
```

```
subplot(3,3,6)  
imagesc(D7(:, :, 1))  
colorbar  
title('7 clusters')
```

```
subplot(3,3,7)  
imagesc(D8(:, :, 1))  
colorbar  
title('8 clusters')
```

```
subplot(3,3,8)  
imagesc(D9(:, :, 1))  
colorbar  
title('9 clusters')
```

```
subplot(3,3,9)  
imagesc(D10(:, :, 1))  
colorbar  
title('10 clusters')
```

```
% Plotting all spectra from an image after segmentation into 6 clusters
```

```
aa = find(0<IDX6 & IDX6 < 2);  
bb = find(1<IDX6 & IDX6 < 3);  
cc = find(2<IDX6 & IDX6 < 4);
```

```

dd = find(3<IDX6 & IDX6 < 5);
ee = find(4<IDX6 & IDX6 < 6);
ff = find(5<IDX6 & IDX6 < 7);
gg = find(6<IDX6 & IDX6 < 8);
hh = find(7<IDX6 & IDX6 < 9);
ii = find(8<IDX6 & IDX6 < 10);
jj = find(9<IDX6 & IDX6 < 11);

Cluster1 = Absorbance_Final(aa,:);
Cluster2 = Absorbance_Final(bb,:);
Cluster3 = Absorbance_Final(cc,:);
Cluster4 = Absorbance_Final(dd,:);
Cluster5 = Absorbance_Final(ee,:);
Cluster6 = Absorbance_Final(ff,:);
Cluster7 = Absorbance_Final(gg,:);
Cluster8 = Absorbance_Final(hh,:);
Cluster9 = Absorbance_Final(ii,:);
Cluster10 = Absorbance_Final(jj,:);

F3 = figure;
imagesc(D6(:, :, 1))
colorbar
title('6 clusters')

F4 = figure;
subplot(2,5,1)
plot(wavenumber, Cluster1)

subplot(2,5,2)
plot(wavenumber, Cluster2)

subplot(2,5,3)
plot(wavenumber, Cluster3)

subplot(2,5,4)
plot(wavenumber, Cluster4)

subplot(2,5,5)
plot(wavenumber, Cluster5)

subplot(2,5,6)
plot(wavenumber, Cluster6)

subplot(2,5,7)
plot(wavenumber, Cluster7)

subplot(2,5,8)
plot(wavenumber, Cluster8)

subplot(2,5,9)
plot(wavenumber, Cluster9)

subplot(2,5,10)
plot(wavenumber, Cluster10)

```

4. Hierarchical Cluster Analysis (HCA)

The code below describes how to perform HCA in FTIR datasets:

```
Absorbance = [FTIR_data];
wavenumber = [Wavenumbers];

% Savitzky-Golay smoothing
Absorbance = (Absorbance)';
Absorbance_Smoothing_1 = sgolayfilt(Absorbance,2,11);
Absorbance_Smoothing_2 = sgolayfilt(Absorbance_Smoothing_1,2,11);
Absorbance_Smoothing_3 = sgolayfilt(Absorbance_Smoothing_2,2,11);
Absorbance_Final = (Absorbance_Smoothing_3)';

%%%%%%%%%%%% IR CHEMICAL IMAGE %%%%%%%%%%%%%

for i = 1:length(Absorbance_Final(:,1))
    B1(i,:) = mean(Absorbance_Final(i,1:end));
end
C = reshape(B1, length(yAxis), length(xAxis));

%%%%%%%%%%%% HIERARCHICAL CLUSTERING ANALYSIS %%%%%%%%%%%%%

Dist_Euclidiana = pdist(Absorbance_Final,'euclidean');
Cluster_Tree = linkage(Dist_Euclidiana,'Ward');

% 2 clusters image
HCA_2 = cluster(Cluster_Tree,'maxclust',2);
D22 = reshape(HCA_2, length(yAxis), length(xAxis), 1);

% 3 clusters image
HCA_3 = cluster(Cluster_Tree,'maxclust',3);
D33 = reshape(HCA_3, length(yAxis), length(xAxis), 1);

% 4 clusters image
HCA_4 = cluster(Cluster_Tree,'maxclust',4);
D44 = reshape(HCA_4, length(yAxis), length(xAxis), 1);

% 5 clusters image
HCA_5 = cluster(Cluster_Tree,'maxclust',5);
D55 = reshape(HCA_5, length(yAxis), length(xAxis), 1);

% 6 clusters image
HCA_6 = cluster(Cluster_Tree,'maxclust',6);
D66 = reshape(HCA_6, length(yAxis), length(xAxis), 1);

% 7 clusters image
HCA_7 = cluster(Cluster_Tree,'maxclust',7);
D77 = reshape(HCA_7, length(yAxis), length(xAxis), 1);
```

```

% 8 clusters image
HCA_8 = cluster(Cluster_Tree, 'maxclust', 8);
D88 = reshape(HCA_8, length(yAxis), length(xAxis), 1);

% 9 clusters image
HCA_9 = cluster(Cluster_Tree, 'maxclust', 9);
D99 = reshape(HCA_9, length(yAxis), length(xAxis), 1);

% 10 clusters image
HCA_10 = cluster(Cluster_Tree, 'maxclust', 10);
D1010 = reshape(HCA_10, length(yAxis), length(xAxis), 1);

%%%%%%%%%%%% KMCA RECONSTRUCTED IMAGES %%%%%%%%%%%%%

F1 = figure;

subplot(3,3,1)
imagesc(C);
colorbar
title('Chemical Map')

subplot(3,3,2)
imagesc(D33(:, :, 1))
colorbar
title('3 clusters')

subplot(3,3,3)
imagesc(D44(:, :, 1))
colorbar
title('4 clusters')

subplot(3,3,4)
imagesc(D55(:, :, 1))
colorbar
title('5 clusters')

subplot(3,3,5)
imagesc(D66(:, :, 1))
colorbar
title('6 clusters')

subplot(3,3,6)
imagesc(D77(:, :, 1))
colorbar
title('7 clusters')

subplot(3,3,7)
imagesc(D88(:, :, 1))
colorbar
title('8 clusters')

subplot(3,3,8)
imagesc(D99(:, :, 1))

```

```

colorbar
title('9 clusters')

subplot(3,3,9)
imagesc(D1010(:,:,1))
colorbar
title('10 clusters')

% Plotting all spectra from an image after segmentation into 6 clusters

aa = find(0<HCA_6 & HCA_6 < 2);
bb = find(1<HCA_6 & HCA_6 < 3);
cc = find(2<HCA_6 & HCA_6 < 4);
dd = find(3<HCA_6 & HCA_6 < 5);
ee = find(4<HCA_6 & HCA_6 < 6);
ff = find(5<HCA_6 & HCA_6 < 7);
gg = find(6<HCA_6 & HCA_6 < 8);
hh = find(7<HCA_6 & HCA_6 < 9);
ii = find(8<HCA_6 & HCA_6 < 10);
jj = find(9<HCA_6 & HCA_6 < 11);

Cluster1 = Absorbance_Final(aa,:);
Cluster2 = Absorbance_Final(bb,:);
Cluster3 = Absorbance_Final(cc,:);
Cluster4 = Absorbance_Final(dd,:);
Cluster5 = Absorbance_Final(ee,:);
Cluster6 = Absorbance_Final(ff,:);
Cluster7 = Absorbance_Final(gg,:);
Cluster8 = Absorbance_Final(hh,:);
Cluster9 = Absorbance_Final(ii,:);
Cluster10 = Absorbance_Final(jj,:);

F3 = figure;
imagesc(D66(:,:,1))
colorbar
title('6 clusters')

F4 = figure;
subplot(2,5,1)
plot(wavenumber, Cluster1)

subplot(2,5,2)
plot(wavenumber, Cluster2)

subplot(2,5,3)
plot(wavenumber, Cluster3)

subplot(2,5,4)
plot(wavenumber, Cluster4)

subplot(2,5,5)
plot(wavenumber, Cluster5)

subplot(2,5,6)

```

```
plot(wavenumber, Cluster6)
```

```
subplot(2,5,7)  
plot(wavenumber, Cluster7)
```

```
subplot(2,5,8)  
plot(wavenumber, Cluster8)
```

```
subplot(2,5,9)  
plot(wavenumber, Cluster9)
```

```
subplot(2,5,10)  
plot(wavenumber, Cluster10)
```



INSTITUTO DE PESQUISAS ENERGÉTICAS E NUCLEARES
Diretoria de Pesquisa, Desenvolvimento e Ensino
Av. Prof. Lineu Prestes, 2242 – Cidade Universitária CEP: 05508-000
Fone/Fax(0XX11) 3133-8908
SÃO PAULO – São Paulo – Brasil
<http://www.ipen.br>

**O IPEN é uma Autarquia vinculada à Secretaria de Desenvolvimento, associada
à Universidade de São Paulo e gerida técnica e administrativamente pela
Comissão Nacional de Energia Nuclear, órgão do
Ministério da Ciência, Tecnologia e Inovação.**