Université de Reims Champagne-Ardenne

UFR Pharmacie

Année 2012 N°

THESE

Présentée pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE

Mention : Ingénierie de la santé

Soutenue publiquement le 14/09/2012

Par

Jayakrupakar NALLALA

Né le 23 mars 1984 à Metpally, Inde

Caractérisation moléculaire de lésions tumorales par imagerie spectrale infrarouge : implémentation d'un nouveau concept basé sur l'histopathologie spectrale pour le diagnostic du cancer du côlon

Unité MEDyC, CNRS FRE 3481, SFR Cap Santé,
UFR de Pharmacie

JURY

Rapporteur:	Dr. Dominique GUENOT (Strasbourg)
Rapporteur:	Dr. François LE-NAOUR (Paris)
Examinateur:	Dr. Nicholas STONE (Exeter, GB)
Examinateur:	Dr. Jacques KLOSSA (Paris)
Examinateur:	Pr. Marie-Danièle DIEBOLD (Reims)
President:	Pr. Michel MANFAIT (Reims)
Directeur de thèse:	Pr. Ganesh SOCKALINGUM (Reims)
Directeur de thèse:	Dr. Olivier PIOT (Reims)

Université de Reims Champagne-Ardenne

UFR Pharmacie

Année 2012 N°

THESE

Présentée pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE REIMS CHAMPAGNE-ARDENNE

Mention : Ingénierie de la santé

Soutenue publiquement le 14/09/2012

Par

Jayakrupakar NALLALA

Né le 23 mars 1984 à Metpally, Inde

Molecular characterization of tumoral lesions by infrared spectral imaging: implementation of a new concept based on spectral histopathology for colon cancer diagnosis

Unité MEDyC, CNRS FRE 3481, SFR Cap Santé,

UFR de Pharmacie

JURY

Rapporteur:	Dr. Dominique GUENOT (Strasbourg)
Rapporteur:	Dr. François LE-NAOUR (Paris)
Examinateur:	Dr. Nicholas STONE (Exeter, GB)
Examinateur:	Dr. Jacques KLOSSA (Paris)
Examinateur:	Pr. Marie-Danièle DIEBOLD (Reims)
President:	Pr. Michel MANFAIT (Reims)
Directeur de thèse:	Pr. Ganesh SOCKALINGUM (Reims)
Directeur de thèse:	Dr. Olivier PIOT (Reims)

Remerciements

A Monsieur le Professeur Michel Manfait,

Je vous remercie très sincèrement pour m'avoir accueilli au sein de votre unité, pour m'avoir donné l'opportunité de réaliser ce travail et de m'avoir fait profiter de votre expérience en spectroscopie vibrationnelle. Je voudrais également vous remercier pour m'avoir donné l'occasion de présenter mes travaux dans des congrès internationaux et nationaux. Veuillez trouvez ici l'expression de ma reconnaissance.

A Madame le Docteur Dominique Guenot,

Je vous remercie de me faire l'honneur d'être Rapporteur de ce travail. Veuillez accepter mes sincères remerciements pour votre présence dans ce jury.

A Monsieur le Docteur François Le-Naour,

Je vous remercie de me faire l'honneur d'être Rapporteur de ce travail. Veuillez accepter mes sincères remerciements pour votre présence dans ce jury.

To Dr. Nicholas Stone,

I would like to convey my sincere gratitude for having accepting to be a Jury member for my thesis. Your presence in the jury members is an honour to me.

A Madame le Professeur Marie-Danièle Diebold,

Je vous adresse mes remerciements les plus respectueux pour avoir accepté d'être membre du jury. Je vous remercie pour l'intérêt que vous avez apporté à ce travail, pour m'avoir fait partager votre expertise en anatomopathologie et ainsi de m'avoir permis d'améliorer mes connaissances.

Remerciements

A Monsieur le Professeur Ganesh D. Sockalingum,

Je te serai toujours reconnaissant pour m'avoir encadré et guidé dans ce travail. Je te remercie sincèrement pour tes conseils et tes remarques pertinentes, pour ta disponibilité et surtout pour la grande liberté que tu m'as accordée. J'ai également apprécié ta bonne humeur et ta sympathie tout au long de mon doctorat.

A Monsieur le Docteur Olivier Piot,

Je te serai toujours reconnaissant pour m'avoir encadré et guidé dans ce travail. Je tiens à t'adresser toute ma gratitude pour l'appui scientifique que tu m'as apporté, pour le temps que tu m'as consacré, ainsi que pour ton soutien amical et pour la confiance que tu as su m'accorder.

A Monsieur le Docteur Jacques Klossa,

Je vous adresse mes remerciements les plus respectueux pour avoir accepté d'être membre du jury.

A Cyril,

Un très grand merci pour ta participation dans ce projet. Sans le temps que tu as accordé pour le traitement des données, la fin de ma thèse aurait été difficile à imaginer. Je te remercie également pour ta patience et pour ta disponibilité toutes les fois où Matlab indiqué des messages ROUGES.

A Monsieur le Professeur Olivier Bouché,

Un grand merci pour votre participation et votre collaboration à ce travail.

A Madame le Docteur Eva Brabencova,

Un grand merci pour votre participation et votre collaboration à ce travail.

Remerciements

Un grand merci :

Au Dr. C. Murali Krishna pour m'avoir donné l'opportunité de découvrir le monde de la spectroscopie vibrationnelle et de m'avoir appris les bases.

A Valérie pour toute l'aide qu'elle m'a apporté pour les manipulations, et à Saviz, Nicole et Caroline pour leur aide précieuse pour la préparation des coupes de tissus.

A Aurélie en tant que 'bureau-mate' pour m'avoir soutenu et pour avoir maintenu une très bonne ambiance de travail toujours dans la joie et la bonne humeur. Je te remercie pour toute l'aide que tu m'as apporté et en particulier pour les traductions en français.

A tous les doctorants : Teddy, Hadrien, Marie, The Thuong, Georges, Mathilde, Nathalie et Caroline pour leur soutien et leurs encouragements, je n'oublierai jamais tous les bons moments passés ensemble au laboratoire ainsi que ceux passés en dehors.

A David, Mohammed, Elodie, Adeline, Julien, Monsieur Angiboust, Hamid, Abdel, Jennifer, Irène et Madame Pisani pour leur gentillesse, leur amabilité et leur hospitalité.

Je voudrais également saluer tous mes amis et mes proches, en particulier Zakir bhai, Neah, Bébé, Deepak-ji, Bapu-ji, Surendra, Vijeta, Arnaud, Jean Claude, Lucie, Mai anh, Didier, Catherine et Aurore pour tous les bons moments que l'on a partagé tout au long de ma thèse, je ne les oublierais jamais.

Enfin, je dédie ce travail à toute ma famille, les mots ne sont pas suffisants pour leur exprimer ma profonde gratitude. Je ne saurais jamais les remercier assez pour leur compréhension et pour m'avoir soutenu et encourager de manière inconditionnelle même dans les moments de doute. Leur présence tout au long de mon doctorat a été, pour moi, un soutien et un réconfort sans faille.

Je ne saurais terminer cette page sans avoir une pensée pour Alexandre Mazine, avec lequel j'ai passé d'agréables moments durant ma thèse.

Table of contents

Avant-propos

Ce présent travail d'histopathologie spectrale infrarouge (IR) est composé de deux objectifs principaux. Le premier d'entre eux était de développer une méthodologie pour l'application de l'imagerie spectrale IR pour les tissus de côlon.

Pour ce faire, l'imagerie IR spectrale a été réalisée sur des coupes de tissus congelées en vu d'établir la méthodologie. L'analyse des images spectrales en utilisant des analyses multivariées ont permis le développement d'un nouveau concept de code-barres spectraux qui constitue un outil facile pour la représentation et l'interprétation des marqueurs spectraux discriminants entre les signatures normales et tumorales des tissus du côlon.

Cette approche d'imagerie a ensuite été effectuée sur des tissu arrays paraffinés constitués d'échantillons de plus grande taille provenant de banque de tumeurs. Initialement, l'imagerie spectrale IR a été mise en œuvre sur un petit nombre de tissu arrays afin de standardiser la méthodologie. Étant donné que ces tissu arrays sont paraffinés et stabilisés au sein d'une matrice d'agarose, les interférences spectrales dues à ces deux constituants ont été neutralisé à l'aide d'une correction mathématique appellée « Extended Multiplicative Signal Correction ». Comme ces coupes de tissu sont relativement grandes par rapport aux tissu microarray, la méthode de clustering par k-means, connue pour l'obtention d'une classification des données de manière rapide et robuste, a été appliquée pour classer les différents spectres des images IR en fonction des constituants caractéristiques histologiques des coupes de tissu. La comparaison de ces images spectrales aux images colorées de manière conventionnelle en histologie, a permis l'identification et l'attribution de classes caractéristiques des différentes structures tissulaires à l'aide de l'expertise d'une pathologiste. Cette méthodologie constitue l'histopathologie spectrale d'une manière non-destructive et ne nécessitant aucun marquage.

Jusqu'à ce point, la méthodologie d'imagerie spectrale IR adaptée à des tissu arrays paraffinés a été créée constituant l'histopathologie spectrale. Une fois que cette méthodologie a été disponible, le deuxième objectif était d'appliquer cette nouvelle approche à un plus grand nombre d'échantillon à des fins diagnostiques. En utilisant un petit nombre d'échantillon, un modèle de prédiction a été développé décrivant l'histologie des tissus normaux et des tissus tumoraux de côlon. Ce modèle a ensuite été appliqué sur des échantillons de tissu colique inconnu afin d'identifier leurs différents constituants caractéristiques histologiques, et ainsi de prédire si les échantillons sont cancéreux.

Ce travail est présenté en cinq chapitres. Le chapitre introduit brièvement les principaux aspects du cancer du côlon, les aspects cliniques de diagnostic, et les aspects techniques relatifs à l'imagerie spectrale IR.

Le deuxième chapitre décrit l'instrumentation et les aspects méthodologiques, depuis la préparation des échantillons au sein du laboratoire d'anatomopathologie, l'acquisition par imagerie spectrale IR, jusqu'à l'analyse statistique multivariée des données spectrales.

Les résultats obtenus au cours de ce travail interdisciplinaire sont présentés dans le troisième chapitre. Ce chapitre comprend trois articles qui ont été soumis pour publication dans différents journaux internationaux.

Le quatrième chapitre porte sur un travail supplémentaire qui a été mené au cours de ce projet. Il décrit essentiellement l'utilisation d'autres approches de spectroscopie vibrationnelle comme l'imagerie IR en mode transmission, l'imagerie IR en mode réflexion totale atténuée (ATR), et l'imagerie Raman. Les avantages et les limitations de chaque méthode ont été élucidés en réalisant une comparaison de ces différentes approches. Puis, un autre travail d'imagerie IR appliquée aux tissus mammaires est également décrit.

Enfin, le cinquième chapitre décrit les conclusions importantes observées dans cette étude, ainsi que les perspectives en recherche et dans le domaine clinique, basées sur les potentiels de ce travail interdisciplinaire.

Summary

The current work of infrared (IR) spectral histopathology principally consisted of two main objectives. The first of these was to develop a methodology for the application of IR spectral imaging to colonic tissues.

For this, IR spectral imaging was performed on frozen tissue sections in order to establish the methodology. The analysis of the spectral images using multivariate analyses enabled the development of a new concept of spectral barcodes which constituted an easy-to-interpret representation of the discriminant spectral markers between normal and tumoral signatures of the colonic tissues.

This imaging approach was then carried on onto paraffinized tissue arrays that constituted a larger sample size accessible from the tumor bank. Initially, the IR spectral imaging was implemented on a small number of tissue array cores in order to standardize the methodology. Since these tissue arrays were paraffinized and stabilized in an agarose matrix, their interference with the tissue spectra was neutralized using a modified version of the Extended Multiplicative Signal Correction (EMSC). Due to the relatively large size of the tissue arrays compared to tissues microarrays, k-means clustering method, known for its rapid and robust data classification, was applied to cluster the IR spectral images into their constituent histological features. Comparison of these clustered images to the conventionally stained histological images allowed identification and class attribution of various tissue structural features using a pathologist's expertise. This methodology constituted spectral histopathology in a non-destructive and label-free manner.

Up to this point, IR spectral imaging methodology adapted to paraffinized tissue arrays was established constituting a spectral histopathology. Once this was available, the second objective was to apply it to large scale sample set for diagnostic purposes. Multivariate analyses were then employed in order to develop a prediction model describing the normal and the tumoral histology of the colonic tissues. This model was then applied on unknown colonic tissue sample to identify their different constituent histological features, and also to predict if the samples are cancerous.

The current work is categorized into five chapters. The introductory chapter briefly covers the major aspects of colon cancer, the clinical aspects of diagnosis, and the technical aspects pertaining to infrared spectral imaging.

The second chapter describes the instrumentation and the methodological aspects, starting from sample preparation in the pathological laboratory, data acquisition in the imaging systems, to the multivariate analysis of the spectral data.

This highly interdisciplinary work is reflected in third chapter which assimilates the important results obtained during the course of this work. This chapter includes three articles that have been submitted for publication in different international journals.

The fourth chapter describes other supplementary work that has been carried out during this project. It basically describes the approaches tested using conventional IR-transmission, IR-attenuated total reflection, and Raman imaging and a comparison among them in order to elucidate various advantages and limitations. An alternative work of IR imaging applied to breast tissues is also described.

Finally, the concluding chapter 5 describes the important conclusions observed in this study, and also the foreseen perspectives in research and clinical aspects based on the potentials of this inter-disciplinary work.

List of figures and tables

Chapter I: Introduction

Chapter II: Materials and methods

Chapter III: Results and discussion - Infrared spectral histopathology: Concept and application to colon cancer

Article 1: Infrared imaging as a cancer diagnostic tool: introducing a new concept of spectral barcodes for identifying molecular changes in colon cancers

Article 2: Infrared spectral imaging as a novel approach for histopathological recognition in colon cancer diagnosis

Article 3: Infrared spectral histopathology for cancer diagnosis; a novel approach for automated pattern recognition of colon adenocarcinoma

Chapter 4

Abbreviations

APC: Adenosis polyposis Coli

ATR: Attenuated total reflection

CaF_2: Calcium Fluoride

CCD: Charge-coupled device

CRC: Colorectal cancers

DNA: Deoxyribose nucleic acid

EMSC: Extended Multiplicative Signal Correction

FOBT: Fecal occult blood test

FTIR: Fourier-transform infrared

HE: Hematoxyline-Eosine

HPS: Hematoxyline Phloxine Saffron

IHC: Immunohistochemistry

IR: Infrared

LDA: Linear discriminant analysis

MCT: mercury cadmium telluride

PCFS: Peri-cryptal fibroblastic sheath

PCA: Principal component analysis

PC: principal component

RNA: Ribose nucleic acid

TNM: Tumor Node Metastasis

CHAPTER I

Introduction

I.1: Résumé:

Les cancers colorectaux, tous sexes confondus, présentent des taux de morbidité et de mortalité très important. Les principaux facteurs de risques associés à ces cancers sont le vieillissement, les habitudes alimentaires, la consommation d'alcool, l'obésité, le tabagisme, les maladies inflammatoires de l'intestin, certains facteurs génétiques et les antécédents familiaux. La détection de polypes adénomateux est considérée l'un des facteurs de risque importants pour les cancers colorectaux.

Les facteurs génétiques impliqués dans les cancers colorectaux entrainent une évolution de la maladie de l'adénome vers le carcinome via une séquence de mutations génétiques bien connues. Ces mutations concernent les gènes suppresseurs de tumeurs, les oncogènes et les gènes impliqués dans la réparation de l'ADN. La mutation du gène suppresseur de tumeur Apc (adenosis polyposis coli) est considérée comme initiatrice de cette progression adénome-carcinome. L'initiation de cette séquence est appuyée par d'autres mutations, comme celles de l'oncogène Ras qui est impliqué dans la progression et la différenciation cellulaire. Par la suite, les mutations touchant les gènes DCC, SMAD 4 et p53 (mutations des deux allèles) aboutissent à la formation d'une tumeur carcinomateuse. Par ailleurs, les tumeurs présentant une instabilité dans les séquences microsatellites sont connues pour avoir une transition plus rapide adénome-carcinome.

La majorité des cancers colorectaux sont des adénocarcinomes. L'adénocarcinome est une tumeur maligne qui se caractérise histologiquement par une modification de l'architecture de l'épithélium glandulaire au niveau de la muqueuse colorectale. Le diagnostic précoce et précis des cancers augmente considérablement les chances de survies et peut permettre une meilleure compréhension des mécanismes biomoléculaires responsables des modifications morphologiques et pathologiques. Actuellement, les méthodes de dépistages les plus utilisées comprennent le test de fecal occult blood (FOBT), la coloscopie et la sigmoïdoscopie. L'examen anatomopathologique est la méthode de référence pour identifier des modifications morphologiques et déterminer la malignité d'une tumeur. Cet examen est basé sur la visualisation au microscope d'un échantillon tissulaire.

De nouvelles méthodes d'analyses, complémentaires de l'examen anatomopathologique, sont en cours de développement. La spectroscopie infrarouge (IR) se place comme étant l'une des techniques les plus prometteuses. Elle permet de fournir des informations sur la composition

biomoléculaire des tissus sans colorations ou marquages au préalable. De plus, l'association d'un imageur et d'un spectromètre IR permet l'acquisition rapide d'images spectrales IR apportant simultanément des informations sur la morphologie et sur la composition biomoléculaire de l'échantillon.

Dans ce contexte, la technique d'imagerie IR couplée avec des analyses multivariées (k-means, Mann-Whitney *U* test, et l'analyse discriminante linéaire) a été appliquée sur des échantillons tissulaires de côlon. Les principaux objectifs de cette étude étaient d'identifier les structures histologiques présentent dans le côlon et de mettre en évidence de marqueurs spectraux caractéristiques de l'état histopathologique des tissus. Dans un premier temps, cette étude visait à exploiter ces markers spectroscopiques pour créer un code-barres spectral spécifique de l'adénocarcinome modérément différencié. Dans un deuxième temps, ces markers spectroscopiques ont été utilisés pour développer un modèle de prédiction afin de détecter et d'identifier numériquement la malignité d'une tumeur au sein des échantillons analysés en « aveugle ». Cette méthodologie vise à proposer un diagnostic automatisé de l'adénocarcinome du côlon.

I.2: Summary:

Colorectal cancer (CRC) is one of the leading cancers in terms of both morbidity and mortality, and which is common to both sexes. Several risk factors are associated with CRCs that include old age, diet habits, alcohol, obesity, smoking, inflammatory bowel disease, genetic factors and family history. Presence of adenomatous polyps is considered as one of the important risk factors.

The genetic factors involved in CRCs are known to follow a sequence of adenoma-carcinoma progression that is associated with defined genetic events. This sequence consists of characteristic mutations categorized into mutations of tumor suppressor genes, oncogenes and DNA repair genes. The mutation of the adenosis polyposis coli (APC) regulatory pathway is considered to be the first, early stage step of this process. This initial step is sustained by other genetic mutations like RAS that promote progression of the genetic events. Furthermore, mutations in genes such as DCC, SMAD 4, and finally loss of both alleles of p53 drive progression to carcinoma. Additionally, microsatellite instable tumors are considered to have faster rate of adenoma-carcinoma progression.

Adenocarcinoma, which is a malignant tumor originating from glandular epithelium of the colorectal mucosa accounts for the majority of the CRC types. Early and accurate diagnosis of cancers which enhance biomolecular level understanding of the morphological and pathological changes occurring in the host tissue, and which can improve the chances of survival are two of the most important factors. At present, different detection and screening methods such as fecal occult blood test (FOBT), sigmoidoscopy, colonoscopy, etc are utilized for colorectal cancers. However, the final diagnosis is based on the microscopic examination of the symptomatic tissue with the ‘gold standard’ histopathology using which various tissue morphological aberrations are visualized.

At the same time, techniques that could provide complementary information of the diseased condition to histopathology are being tried and tested. Of these, a biophotonic approach of infrared (IR) spectroscopy is being considered as one of the promising candidates due to its ability to provide biomolecular fingerprint of cells and tissues. Combined with an imaging device, spectral imaging can be performed to obtain IR spectral images in a rapid and in a label-free manner, the information from which can be exploited to gain insights into the histopathological aspects of a diseased state.

In this context, IR imaging in conjunction with multivariate analyses (k-means clustering, Mann-Whitney *U* test, and linear discriminant analysis) was carried out on colonic tissue samples. The main objectives of the study were to identify spectral markers representative of various histological structures and histopathological aspects of the colonic tissues. Further, it was aimed to use these markers to construct "spectral-barcodes" specific to moderately differentiated colon adenocarcinoma; and also to construct a prediction model to digitally detect and identify malignancy and its associated features in unknown tissues, thereby constituting an automated diagnostic approach for colon adenocarcinoma.

I.3: Introduction to colorectal cancers: Basic features of colorectal histology, carcinogenesis and current diagnostic methods:

I.3.1: Colorectal cancers and incidence

Colorectal cancer (CRC) is one of the most common cancer types affecting both sexes. It is the third and second most common cancer in men and women respectively among all cancers worldwide (figure 1). It is estimated that about 8 % of all cancer deaths would be from CRCs making it the fourth most common cause of death from cancers. About 60 % of the cases occur in developed regions of the world (Ferlay, 2010). Although CRC manifests in several types, a majority of them are the adenocarcinomas which accounts for about 90-95% of CRCs. The other less prevalent types include soft tissue sarcomas (leiomyosarcoma), lymphomas, squamous cell cancers, and carcinoid tumors.

I.3.2: Causes and risk factors (non-genetic and genetic)

There are several risk factors associated with CRCs and most of them affect subjects with little or no genetic risk. Risk factors include old age, diet habits with high fat intake, alcohol and red meat, obesity, smoking, inflammatory bowel diseases, genetic factors, and family history. Below the age of 40, CRC without a genetic predisposition are rare and the risk increases with increasing age (IARC, 2004).

High fat diet

High fat diet and low fiber diet are considered risk factors for CRCs although there are exceptions (Rose 1986). The breakdown products of fat metabolism are believed to produce carcinogens and predispose individuals to CRCs, while high fiber diet is believed to have a protective effect. Other dietary components like vitamin B6, calcium, and folate have been proposed as protective factors (Giovannucci 1998).

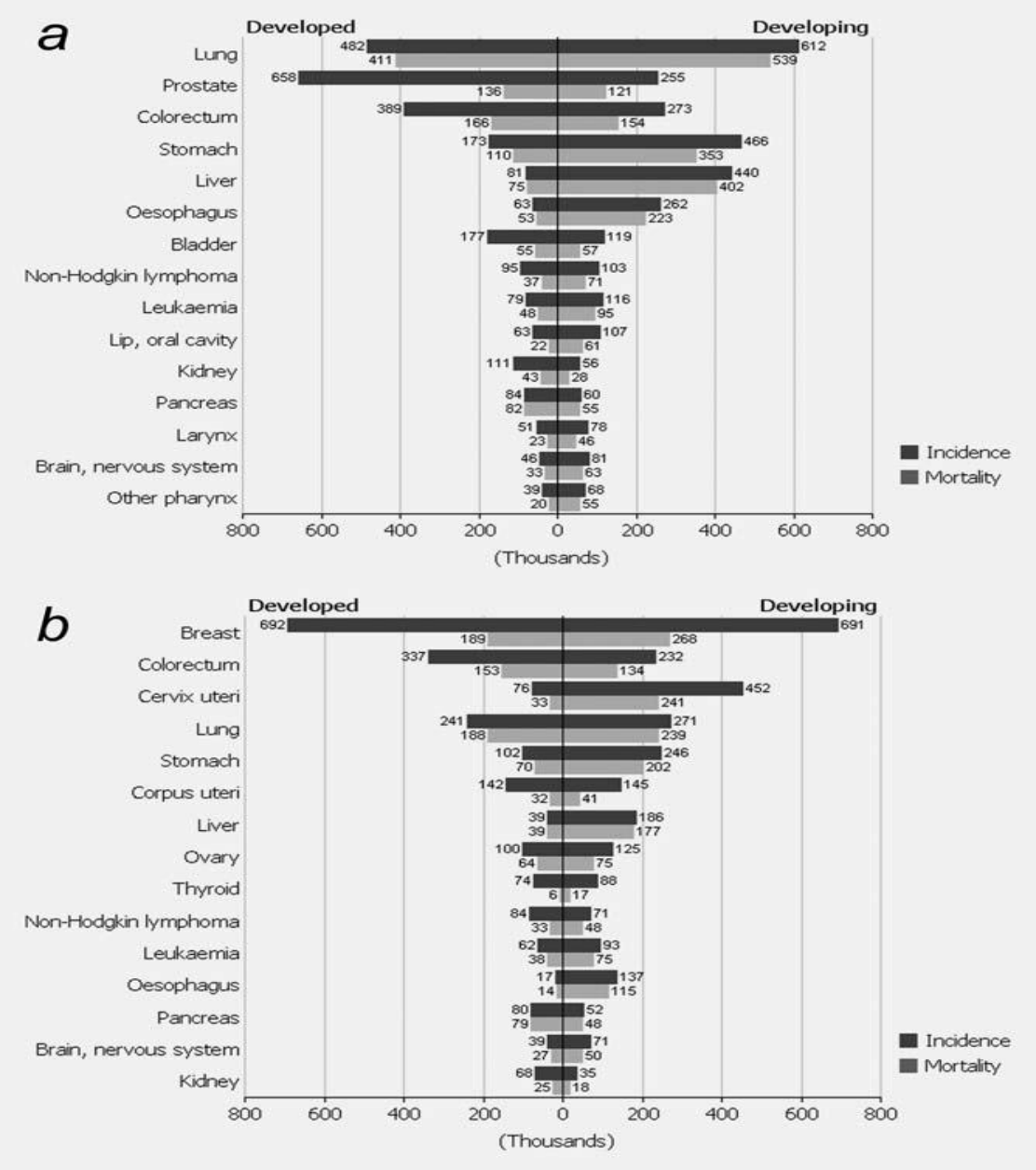


Figure 1: Cancer incidence and mortality.

Estimated numbers (in thousands) of new cancer cases (incidence) and deaths (mortality) in developed and developing regions of the world in 2008: (a) in men and (b) in women. (Source: Int J Cancer 2010; 127(12), 2893-2917).

Inflammatory bowel disease

Studies indicate that individuals with inflammatory bowel disease (ulcerative colitis and Crohn's disease) are at an increased risk of CRCs. The risk is greater the longer the patient has had the disease, along with the severity of the inflammation (Triantafillidis, 2009). Although both these disorders are inflammatory conditions, they are differentiated based on the location, and on the nature of the inflammatory changes. The Crohn's disease can affect any part of the gastrointestinal tract, from mouth to anus, although a majority of the cases start in the terminal ileum. In contrast, ulcerative colitis is restricted to the colon and the rectum. Microscopically, ulcerative colitis is restricted to the mucosa, while Crohn's disease affects the whole bowel wall ("transmural lesions"). Crohn's disease, also known as regional enteritis, is caused by interactions between environmental, immunological, and bacterial factors in genetically susceptible individuals. This result in a chronic inflammatory disorder, in which the body's immune system attacks the gastrointestinal tract possibly directed at microbial antigens. Genetic susceptibility has been associated to the Crohn's disease, primarily with variations of the *NOD2* gene and its protein (Hugot 2001). Ulcerative colitis on the other hand is a form of colitis that includes characteristic ulcers, or open sores. Although no known cause exists, genetic susceptibility is presumed. Ulcerative colitis is treated as an autoimmune disease. The disease may be triggered in a susceptible person by environmental factors.

Adenomatous polyps

Presence of precancerous polyps is one of the important risk factors for CRC (Winawer, 1993). Polyps are fleshy growths that occur on the inside of the colon or rectum (figure 2) which are increasingly observed with increasing age. Although polyps are associated with CRC they are often benign. Polyps are generally classified as hyperplastic, neoplastic (adenomatous & malignant), hamartomatous, and inflammatory. Neoplastic polyps that constitute adenoma or the adenomatous polyps are the most significant polyp types that are associated with CRC. Several subtypes of adenoma exist that differ primarily in the way the cells of the polyp are assembled when examined under the microscope, like tubular, villous, or tubulo-villous adenomas. Villous adenomas are associated with the highest malignant potential as they generally have the largest surface area and are most likely to become cancerous, while the tubular adenomas are the least likely.

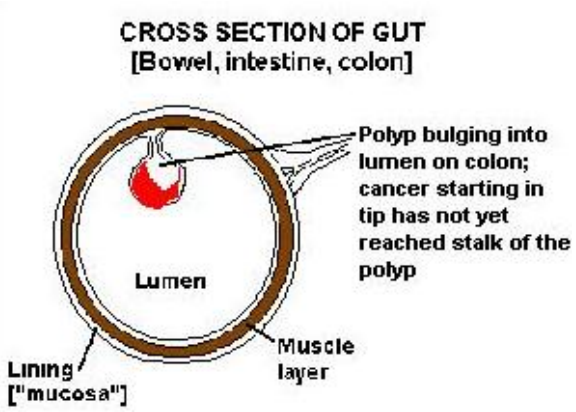
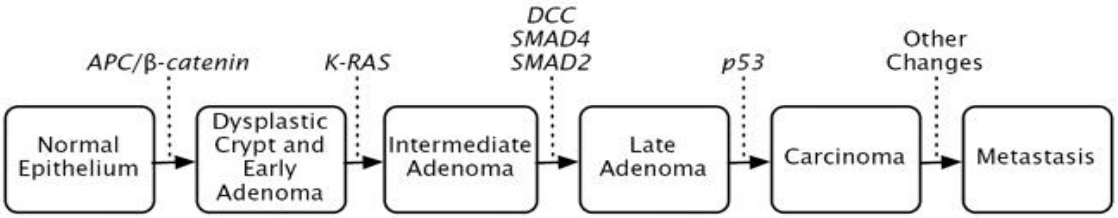


Figure 2: A cross section representation of a polyp lining the mucosa of the large intestine.

Genetics

A majority of CRCs are sporadic. However, among individuals with a family history of CRC, genetic contribution is indicated, that follow a pattern of autosomal dominant inheritance of cancer susceptibility (each child has a 50% chance of inheriting the predisposition with both sexes carrying the same risk, and 3 times more with CRCs occurring in first-degree relative). In around 5% to 6% of CRCs, genetic mutations have been identified as the cause of inherited cancer risk in few colon cancer prone families (Burt, 2004). Over 70% of CRCs, regardless of etiology, arise from adenomatous polyps. Hereditary and somatic mutations have been identified in adenomatous polyps, and are thought to follow a multistep process beginning with early adenoma before transforming into invasive carcinoma (Fearon, 1990) (see figure 3). This adenoma-carcinoma sequence includes characteristic mutations that are categorized into three types: mutations of tumor suppressor genes (autosomal recessive trait where both alleles need to be damaged to lose function), oncogenes (mutated proto-oncogenes where only one allele need to be mutated to cause dysfunction), or DNA repair genes. These alterations result in mucosal proliferation forming a polyp and finally carcinoma. In the sequence of events, mutation of the adenosis polyposis coli (APC) regulatory pathway is believed to be the first, early stage step of this process (Laken, 1999).

A.



B.

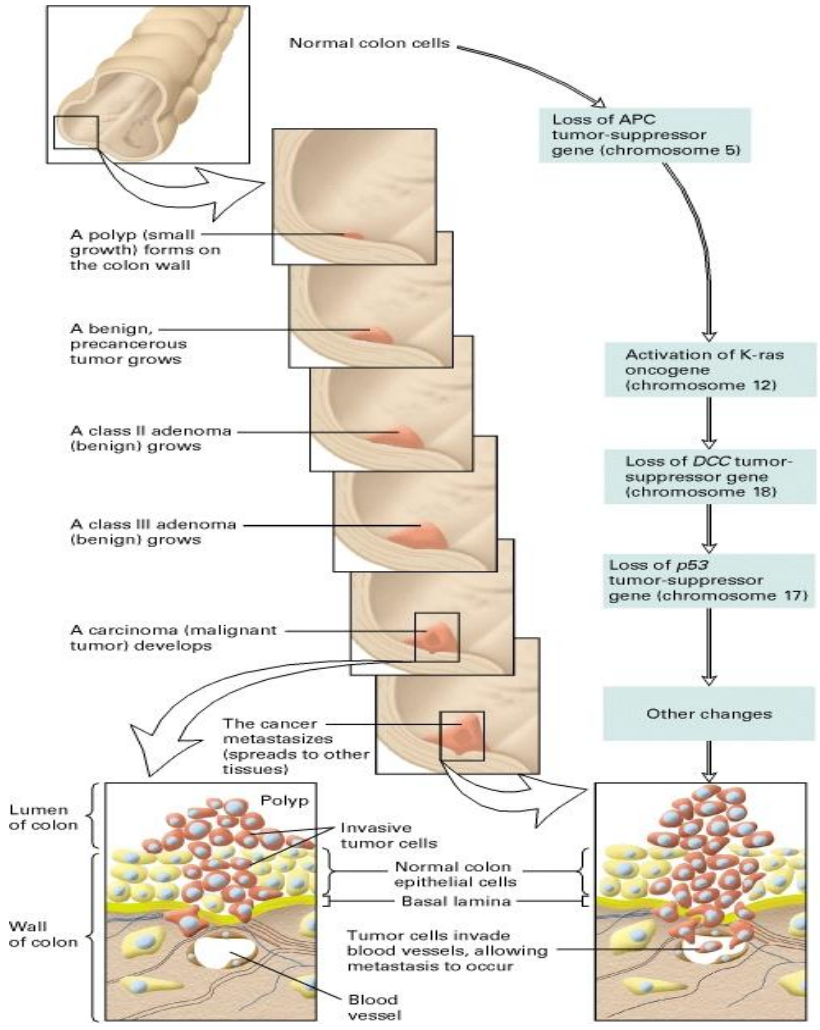


Figure 3: The adenoma-carcinoma progression.

The multistep process of colorectal adenoma - carcinoma progression (A) together with the morphology of colorectal cancer progression (B). This classical pathway is characterized by traditional adenoma morphology, slow progression, and high adenoma: carcinoma ratio, frequent chromosomal instability and aneuploidy, and rare microsatellite instability (Original source: Fearon and Vogelstein, 1990).

The APC represses β-catenin, which is known to promote cell growth (via enhancing c-Myc expression that promotes cellular division), thereby reducing abnormal tissue expansion. As cells migrate from the base of the crypts toward the epithelial surface, a rise in APC expression which represses β-catenin, is associated with increased apoptosis, necessary to balance production from the base of crypt. The tumors lacking APC mutations frequently harbor β-catenin mutations that resist repression by APC. Disruption of the APC pathway may be sufficient to start a small adenomatous growth (Alberici 2006). The APC and the β-catenin signaling form part of the WNT signaling pathway that has been shown to be associated with gastrointestinal tumors.

This crucial initiation step is sustained by mutation of other genes like RAS that often occur among the next genetic events of progression (Fearon, 1990; Janssen, 2006). The strong tendency for APC mutations to appear in the early morphological stages, and RAS mutations to occur only in later morphological stages suggests the important role of the order of mutational steps in colorectal carcinogenesis. Additional genetic events are associated with continuing morphological progression through late adenoma and early carcinoma stages, with the genes DCC, SMAD4, and SMAD2 in 18q21. Finally, the loss of both alleles of the functional p53 drives progression to carcinomas (Fearon 1990).

Chromosomal instability

About 85 percent of colorectal tumors have major chromosomal aberrations (Alberici, 2007). Often, part of a chromosome or a whole chromosome is lost. A lost chromosome is usually replaced by duplication of the remaining chromosome from the original pair. Duplication creates two copies of the same allele at a locus, with loss of one of the original parental alleles called loss of heterozygosity (LOH) and is known to accelerate the genetic changes that drive carcinogenesis (Thiagalingam 2001). Chromosomal instability (CIN) arises from mutations and other genomic changes that abrogate the normal controls on chromosome duplication and segregation in mitosis. Because CIN increases the rate at which genetic changes occur, it can accelerate the sequence of genetic events that drive carcinogenesis. Most solid tissue tumors have CIN, but it remains controversial whether CIN arises early in carcinogenesis and thus plays a key role in driving genetic change, or CIN develops late in tumorogenesis as the genome becomes increasingly disrupted by the later stages of carcinogenesis (Kinzler, 1998).

Other genetic/molecular causes

While the majority of CRCs are due to events that result in chromosomal instability, 20% to 30% of CRCs display characteristic patterns of gene hypermethylation, termed CpG island methylator phenotype (CIMP), of which a portion display microsatellite instability (15% of CRCs) (Lengauer, 1998; Kinzler, 1998; Weisenberger, 2006). The chromosomal instability cancers include alterations in chromosome number (aneuploidy) and detectable losses at the molecular level of portions of chromosome 5q, chromosome 18q, and chromosome 17p; and mutation of the KRAS oncogene (Vogelstein, 1993, Vogelstein, 2002).

Microsatellite instability

Approximately 15 percent of colorectal tumors do not have CIN or widespread chromosomal abnormalities. Instead, these tumors usually have mutations in their mismatch repair (MMR) system that is a component of DNA repair. Loss of mismatch repair function increases mutations in repeated DNA sequences, such as in the microsatellite regions and alters the length of repetitive microsatellites at a higher rate than normal during DNA replication resulting in microsatellite instability (MSI). Genes with repetitive sequences seem to be at greater risk for mutation in microsatellite instable tumors. Most colorectal tumors have either MSI or CIN, but not both. Compared with microsatellite-stable tumors, microsatellite-instable tumors appear to have faster rate of adenoma-to-carcinoma progression. In such tumors, characteristic histological changes such as increased mucin production are also observed, while some mucin types are decreased, suggesting that some molecular events contribute to the histological features of the tumors (Ionov 1993; Thibodeau, 1993).

I.3.3: Anatomy and histology of the colon

Anatomy of the colon

The colon constitutes a part of the digestive system, followed by the rectum which is the end of the colon, adjacent to the anus. The colon and the rectum, together with the cecum make up the large intestine. The human colon as a whole consists of four sections: the ascending colon, the transverse colon, the descending colon, and the sigmoid colon (figure 4). The ascending colon and transverse colon together are usually referred to as the proximal colon.

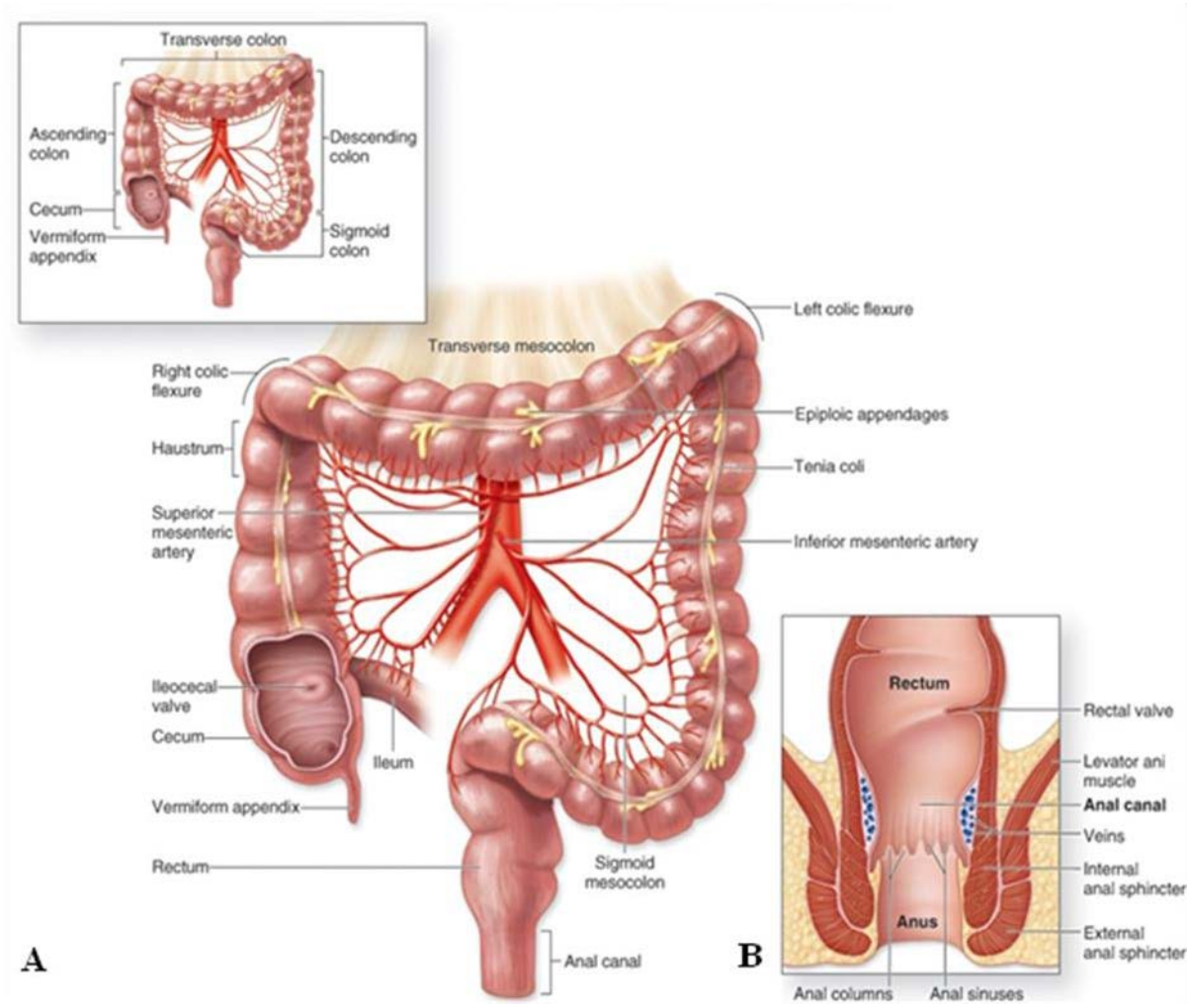


Figure 4: Anatomy of a normal human colon.

A general view (top left), detailed view (A), and (B) represent the anal canal. (Copyright © The McGraw-Hill Companies, Inc).

The ascending colon is present on the right side of the abdomen measuring about 13 cm long that starts from the cecum to the hepatic flexure. The colonic part from the hepatic flexure to the splenic flexure constitutes the transverse colon. This part is supported to the the stomach, attached by a wide band of tissue called the greater omentum. On the posterior side, the transverse colon is connected to the posterior abdominal wall by a mesentery known as the transverse mesocolon. The transverse colon is encased in peritoneum, and is therefore mobile (unlike the parts of the colon immediately before and after it). Cancers are formed more frequently further along the large intestine as the contents become more solid to form feces. The colonic part from the splenic flexure to the beginning of the sigmoid colon constitutes the descending colon. Finally, the sigmoid (S-shaped) colon follows the descending colon before the rectum. The walls of the sigmoid colon are muscular, and contract to increase the pressure inside the colon, causing the stool to move into the rectum.

Histology of the colon

Microscopic observation of a normal colonic tissue biopsy section via hematoxylin and eosin (H&E) staining consists of defined histological structures namely mucosa, areolar (submucosa), muscular and serous as shown in figure 5.

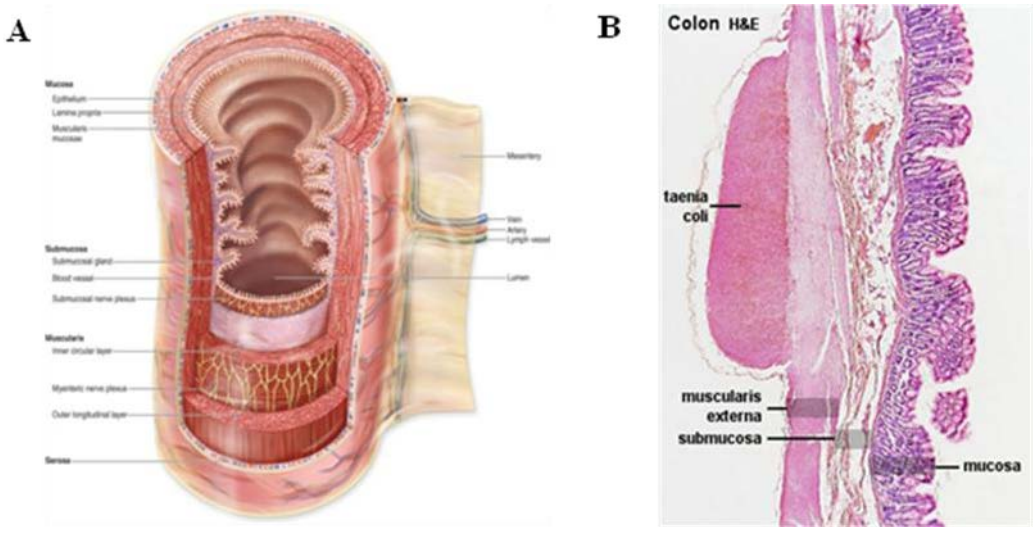


Figure 5: Histology of a normal human colon (A: schematic and, B: HE stained tissue section). Microscopic observation shows the main histological features namely mucosa, areolar tissue (submucosa), muscularis mucosa and serosa.
(A. Copyright © The McGraw-Hill Companies, Inc., B. Blue Histology, The University of Western Ausralia).

The mucous membrane (*tunica mucosa*): The mucous membrane in the colon is smooth, without villi, raised into numerous crescentic folds which correspond to the intervals between the sacculi. In the rectum it is thicker, more vascular, and connected loosely to the muscular coat, as in the esophagus. Similar to the small intestine, the mucous membrane consists of a muscular layer called the muscularis mucosæ; a retiform tissue in which the vessels ramify; a basement membrane and an epithelium which is of the columnar variety. The mucous membrane of the large intestine consists of glands and solitary lymphatic nodules. The glands are minute tubular prolongations of the mucous membrane arranged perpendicularly, side by side, over its entire surface. They are longer and more numerous than those of the small intestine, and open by minute rounded orifices upon the surface, giving it a cribriform appearance. Each gland is lined by short columnar epithelium and contains numerous goblet cells. The functional glands of the colon are constituted by crypts. These are shaped into straight tubular glands by a simple columnar epithelium. There are no villi. In cellular composition, the epithelium of the large intestine resembles that of the small intestine, but with a higher proportion of goblet cells interspersed among the absorptive cells. Although the absorptive cells remain more numerous throughout, goblet cells in the colon are so numerous and so large (bulging against the adjacent absorptive cells) that the colon epithelium sometimes appears to consist mostly of goblet cells. The crypt epithelium also includes stem cells which replenish the epithelium every few days, enteroendocrine cells, and paneth cells. The crypts are separated by the lamina propria, a loose connective tissue infiltrated by many white blood cells, with capillaries and thin strands of smooth muscle. Occasional neutrophils are present in the lamina propria of normal colonic biopsies. Lymphoid follicles of B-lymphocytes are present in colonic mucosa and may extend through the muscularis mucosae into the submucosa.

The areolar coat (*tela submucosa; submucous coat*): The areolar coat connects the muscular and mucous layers closely together. The submucosa is a loose connective tissue supporting the mucosa. It allows the mucosa to move flexibly during peristalsis. The submucosa contains a vascular plexus, relatively large veins and arteries which give rise to the capillary bed of the mucosa.

The muscular coat (*tunica muscularis*): The muscular coat consists of an external longitudinal, and an internal circular, layer of non-striped muscular fibers. The longitudinal fibers form a discontinuous layer over the whole surface of the large intestine. In the cecum and colon they are especially collected into three flat longitudinal bands (*tænæi coli*), each of about 12 mm wide. These bands serve to produce the sacculi which are characteristic of the cecum and colon; accordingly, when they are dissected off, the tube can be lengthened, and its sacculated character disappears. In the sigmoid colon the longitudinal fibers become more scattered; and around the rectum they spread out and form a layer, which completely encircles this portion of the gut. The circular fibers form a thin layer over the cecum and colon, being especially accumulated in the intervals between the sacculi; in the rectum they form a thick layer. The Muscularis mucosa of the lower tract forms a thin layer (only a few muscle fibres in thickness) beneath the deep ends of the crypts.

Serous coat (*tunica serosa*): The serous coat is derived from the peritoneum, and invests different portions of the large intestine to a variable extent. The serous membrane almost completely covers the cecum. The ascending, descending, and iliac parts of the colon are usually covered only in front and at the sides, and a variable amount of the posterior surface is uncovered. The transverse colon is almost completely invested, except the parts corresponding to the attachment of the greater omentum and transverse mesocolon. The sigmoid colon is entirely surrounded. The rectum is covered above on its anterior surface and its sides, and below on its anterior surface.

I.3.4: Patho-physiology of colorectal cancer

CRC originates in the epithelial cells lining the colon or rectum of the gastrointestinal tract. Adenocarcinoma accounts for the most common colorectal cancers types. A conventional histopathological comparison between a normal and a moderately differentiated colon adenocarcinoma is presented in figure 6.

A. **B.**

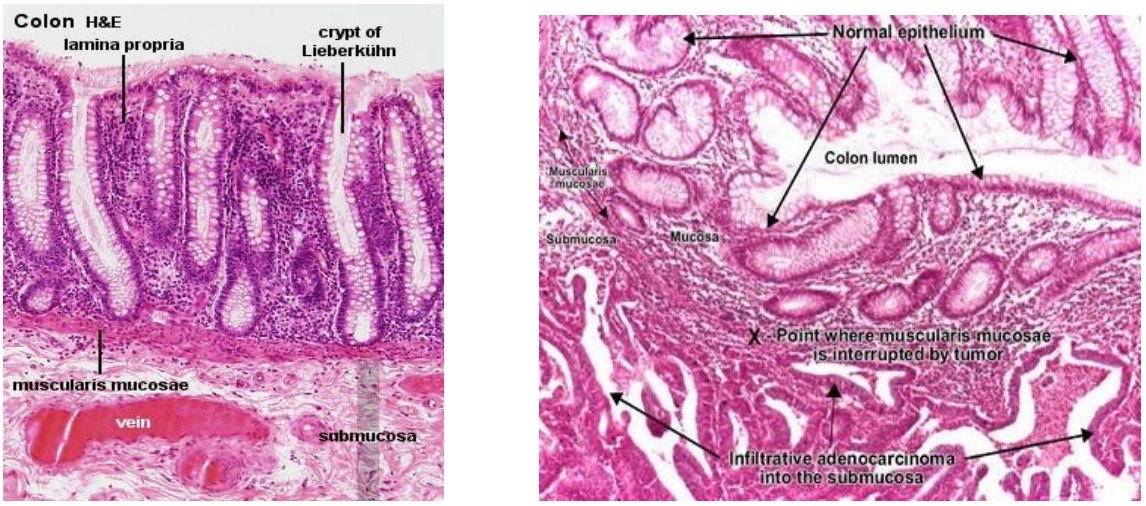


Figure 6: Histological comparison of normal and cancerous colon tissue section.

The normal tissue (A) shows well-differentiated crypts and the surrounding connective tissue, while the differentiation is reduced in the cancerous regions of the tissue (B). The tissue sections are stained with HE. (A. Copyright © The McGraw-Hill Companies, Inc., B. Source: Histology Atlas).

An adenocarcinoma is a malignant tumor, originating from glandular epithelium of the colorectal mucosa. The well-differentiated normal architecture of the colonic epithelium (as described in the sub-section 'Histology of the colon') is lost as the carcinoma advances. The crypts show less open lumen, darkly stained nucleus, decreased cytoplasm which are the signs of malignancy. When the tumor is restricted to the glandular epithelium and has not yet begun to invade the wall of the colon or rectum it is called carcinoma *in situ.* However, once cancer forms in the large intestine, in time it can invade the lining of the colonic or rectal wall. Such cancers can also penetrate blood vessels or lymph vessels. Cancer cells typically

spread first into nearby lymph nodes and can also be carried in blood vessels to the liver or lungs, or can spread in the abdominal cavity to other areas, such as the ovary. The process is called as metastasis. Tumor cells describe irregular tubular structures, harboring pluristratification, multiple lumens and reduced stroma. Sometimes, tumor cells are discohesive and secrete mucus, which invades the interstitium producing large pools of mucus called mucinous (colloid) adenocarcinoma which is poorly differentiated. If the mucus remains inside the tumor cell, it pushes the nucleus at the periphery called as signet-ring cell. Depending on glandular architecture, cellular pleomorphism, and mucosecretion of the predominant pattern, adenocarcinoma may present three degrees of differentiation: well, moderately, and poorly differentiated.

Most adenocarcinomas are moderately differentiated and lack specific histological features, although colorectal tumors tend to show cribriform patterns with central necrosis; a feature that is useful if a metastatic tumor is encountered when no colorectal primary has been diagnosed. Dysplasia in adjacent mucosa may be seen, but frequently the invasive tumor obliterates any pre-existing polyp from which it may have arisen. Mucinous adenocarcinoma is a subtype of adenocarcinoma which secretes extracellular mucin and is known to be associated with MSI. At least 50% of the tumor must be mucinous in order to make this diagnosis.

I.3.5: Colorectal cancer screening and diagnostic tests

Most individuals with early CRC do not have symptoms of the disease. Symptoms usually only appear with more advanced stage of the disease, and hence, screening tests play an important role in identifying the early suspicions indicative of the disease. A positive indication would be a sign to perform the diagnostic tests to find out the cause of the disease. Screening tests involves looking for cancer in individuals who do not have symptoms of the disease. The common screening tests employed for CRC are:

Fecal occult blood test

The fecal occult blood test (FOBT) is used to find occult blood in fecal material (Miyoshi, 2000). The presence of blood in the feces indicates fragile and easily damaged blood vessels

due to the passage of stool at the surface of larger colorectal polyps or cancers. There are two types of FOBTs available. One type, called the guaiac FOBT, uses the chemical guaiac to detect heme, the iron-containing component of the blood protein hemoglobin in samples of stool. The other type of FOBT, called immunochemical (or immunohistochemical) FOBT, uses antibodies to detect human hemoglobin protein in samples of stool. A positive test requires a colonoscopy to find the cause of bleeding and for further investigations.

Double-contrast barium enema

The double-contrast barium enema (DCBE) also called an air-contrast barium enema is basically a type of X-ray examination of the large intestine. The intestine is made visible on an X-ray picture by outlining the inner part of the intestine with the contrast material barium sulphate and air poured in through a tube inserted into the anus. The barium blocks the X-rays, causing the barium-filled colon to show up clearly on the X-ray picture. A colonoscopy will be needed for further examination if any areas are suspected for abnormalities.

Flexible sigmoidoscopy

Flexible sigmoidoscopy enables examining the inside of the large intestine from the rectum through the last part of the sigmoid colon, with a sigmoidoscope (Zuber, 2001). A small video camera fixed at the end of this flexible lighted tube aids to view the images on a display monitor. Since the sigmoidoscope is only 60 cm long, abnormalities in only less than half of the colon that includes the entire rectum can be detected with this procedure. With flexible sigmoidoscopy, intestinal bleeding, inflammation, abnormal growths, ulcers, benign and malignant polyps, as well as early signs of cancer in the descending colon and rectum can also be viewed. The detected abnormalities (e.g. polyp) can be possibly removed and the biopsy sent for further examination.

CT colonography (virtual colonoscopy)

This test is an advanced type of computed tomography (CT or CAT) scan of the colon and rectum. A CT scan is an X-ray examination that produces detailed cross-sectional images of the body. Instead of obtaining a single picture, as in a regular X-ray, a CT scanner takes many

pictures as it rotates around the individual. A computer then combines these pictures into images of slices of the part of the body being studied. For CT colonography, special computer programs create both 2-dimensional X-ray pictures and a 3-dimensional view of the inside of the colon and rectum, which allows one to look for polyps or cancer. In case of detection of an abnormality, a colonoscopy will still likely be needed to remove them or to explore them further.

Colonoscopy

Colonoscopy is basically longer version of a sigmoidoscope that enables examination of the entire length of the colon and rectum with a colonoscope (Rex, 2000). This tube, inserted through the rectum all the way to the beginning of the colon (cecum) via a video camera on the end connected to a display monitor, helps in visualization and closer examination of the inside of the colon. Special instruments can be passed through the colonoscope to remove (biopsy) any suspicious looking areas such as polyps, and if needed are sent for further laboratory examinations.

I.3.6: Histopathology for cancer diagnosis

The cure for cancer relies to a large extent on its diagnosis. The two important factors concerned with diagnosis are, the early detection, which improves the chances of survival, and the biomolecular level understanding of the morphological and pathological changes occurring in the host tissue (Kendall, 2009). The different screening methods in use including the fecal occult blood test (FOBT) (Miyoshi, 2000), colonoscopy (Rex, 2000), sigmoidoscopy (Zuber, 2001), etc., provide firsthand information on the commencement of the disease and show different grades of sensitivity. In case of a positive identification, the tissue is subjected to histopathological analysis. As of now, the diagnosis of cancers is always confirmed and settled upon by microscopic examination of the excised tissue biopsy using histopathology.

The microscopic visualization of the histological components is enhanced by staining the microtome sectioned tissues biopsies. The most widely used staining technique is H&E, in which the nuclei of cells are stained blue by the hematoxylin, while the cytoplasm and the

extracellular connective tissue is stained pink by the eosin. This enables visualization of specific cell types and morphological changes (size, shape, coloration of the nucleus and other tissue features) indicative of disease as shown in figure 6. In certain cases, presence, localization and abundance of specific proteins is detected by antibody based techniques like immunohistochemistry (IHC) that enhances further understanding of the disease. Using such microscopic examination, cancer is staged into different categories which are based on several aspects. As an example, in case of CRC, the staging describes the severity of the cancer in an individual based on the extent of the primary tumor and its penetration to adjacent tissues and organs in the body. Thus, the staging system assesses the extent of local invasion, the degree of lymph node involvement and whether there is distant metastasis. It is performed for diagnostic and research purposes, and to determine the best method of treatment and estimate the individual's prognosis. The TNM staging system [from the American Joint Committee on Cancer (AJCC)] is one of the most commonly used staging systems which is based on three categories, "T" denotes the degree of invasion of the intestinal wall, "N" the degree of lymphatic node involvement, and "M" the degree of metastasis. The broader stage of a cancer is usually quoted by a number I, II, III, IV derived from the TNM value grouped by prognosis; the higher the number, the more advanced is the cancer and more likely a worse outcome. Details of this classification system are presented in table I.

At present histopathology is the gold standard method for cancer diagnosis. At the same time, several techniques are being tried and tested that could provide complementary information of the diseased condition to histopathology (Kendall, 2003). Of these techniques, biophotonic approaches such as infrared (IR) and Raman spectroscopies are being seen as promising candidates due to their ability to provide biomolecular information from cells and tissues.

Table I: TNM staging criteria for colorectal cancers

AJCC stage	**TNM stage**	**2002 6th edition TNM stage criteria for colorectal cancer (superceded by 2010 7th edition)**
Stage 0	Tis N0 M0	Tis: Tumor confined to mucosa; cancer-*in-situ*
Stage I	T1 N0 M0	T1: Tumor invades submucosa
Stage I	T2 N0 M0	T2: Tumor invades muscularis propria
Stage II-A	T3 N0 M0	T3: Tumor invades subserosa or beyond (without other organs involved)
Stage II-B	T4 N0 M0	T4: Tumor invades adjacent organs or perforates the visceral peritoneum
Stage III-A	T1-2 N1 M0	N1: Metastasis to 1 to 3 regional lymph nodes. T1 or T2.
Stage III-B	T3-4 N1 M0	N1: Metastasis to 1 to 3 regional lymph nodes. T3 or T4.
Stage III-C	any T, N2 M0	N2: Metastasis to 4 or more regional lymph nodes. Any T.
Stage IV	any T, any N, M1	M1: Distant metastases present. Any T, any N.

I.4: Introduction to infrared spectroscopy and rationale of the work:

The diagnosis using histopathology is based to a large extent on the microscopic examination of the symptomatic tissue in which preferential stains are used to enhance visualization of the tissue morphological aberrations (Kendall, 2009). Such pre-cancerous or cancerous aberrations are the manifestations of the biomolecular changes that have already undergone the provocative changes for malignancy. However, the ongoing state of the tissue molecular changes during the onset or progression of malignancy, without any visible morphological signatures, poses a challenge for identification. In certain cases, immunohistochemistry (IHC) is used to identify specific proteins of interest, which can give a molecular level understanding of the malignant condition. Histopathology requires precise human expertise which is a limit for high-throughput diagnosis. Therefore, if it can be combined with approaches that could provide complementary biochemical information in a rapid, cost effective manner and reducing human involvement, the efficacy of the histopathological diagnosis could be completed.

In this regard, potential diagnostic methods based on vibrational spectroscopic approaches are foreseen as one of the contenders (Kendall, 2009). Vibrational spectroscopy enables one to understand the structural organization and functional properties of simple molecules and, complex systems such as cells and tissues based on the interaction of light with vibrational states of biomolecules (Martin, 2010). The two most important modalities in vibrational spectroscopies are IR absorption and Raman scattering. Although, both these techniques have different physical origins, they involve the vibrational modes of molecules.

I.4.1: Infrared spectroscopy

In the electromagnetic spectrum, IR radiation covers the region from 14000 cm^{-1} to 10 cm^{-1} (0.8 μm to 1000 μm), which are at longer wavelengths and lower frequencies than the visible light (figure 7). This IR portion is further divided into three regions; the near-, mid- and far-IR. As per ISO 20473 scheme, the higher energy near-IR, approximately from 14000 cm^{-1} to 4000 cm^{-1} (0.8 μm to 3 μm) can excite overtones or harmonic vibrational modes. The mid-IR,

approximately from 4000 cm^{-1} to 400 cm^{-1} (3 µm to 50 µm), is used to study the fundamental vibrations and associated rotational vibrational structure.

The far-IR, approximately from 400 cm^{-1} to 10 cm^{-1} (50 µm to 1000 µm), lying adjacent to the microwave region, has low energy and may be used for rotational spectroscopy.

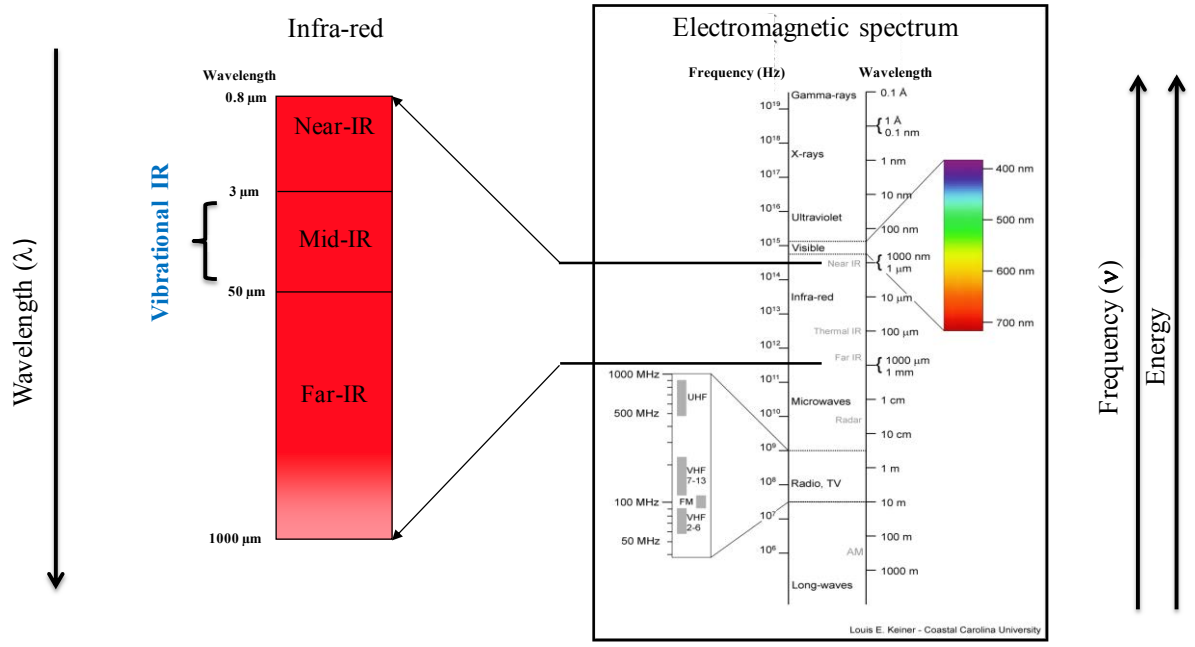


Figure 7: The electromagnetic spectrum showing the location of the IR spectral range

IR spectroscopy is based on the principle of absorption of light by molecules. At the atomic level, IR energy provokes vibrational modes in a molecule through a change in the dipole moment, making it a useful frequency range for study of the energy states of the proper symmetry. Each molecule has its natural vibrational modes with characteristic vibrational frequencies and energies. The energy of molecular vibration is measured by its amplitude (the distance moved by the atoms during the vibration); the higher the vibrational energy, the larger is the amplitude of the motion.

A molecular vibration is excited when the molecule absorbs a quantum of energy corresponding to the vibrational frequency according to the relation (equation 1):

$E = h\nu$ (1)

where,

E = energy

ν = vibrational frequency (Hz)

h = Planck's constant (6.626×10^{-34} J.s)

A fundamental vibration is excited when one such quantum of energy is absorbed by the molecule in its ground state. However, for the molecule to be IR active, this absorption should cause a change in dipole moment of the molecule. A molecule with N atoms, if is linear has 3N-5 degrees of freedom whereas if is a non-linear molecule, it has 3N-6 degrees of freedom where, six corresponds to translations and rotations of the molecule itself. The vibrational modes are named as stretching (a change in the length of a bond), bending or scissoring (a change in the angle between two bonds), wagging (a change in angle between the plane of a group of atoms), twisting (a change in the angle between the planes of two groups of atoms), and rocking (a change in angle between a group of atoms). These different fundamental modes are shown in figure 8.

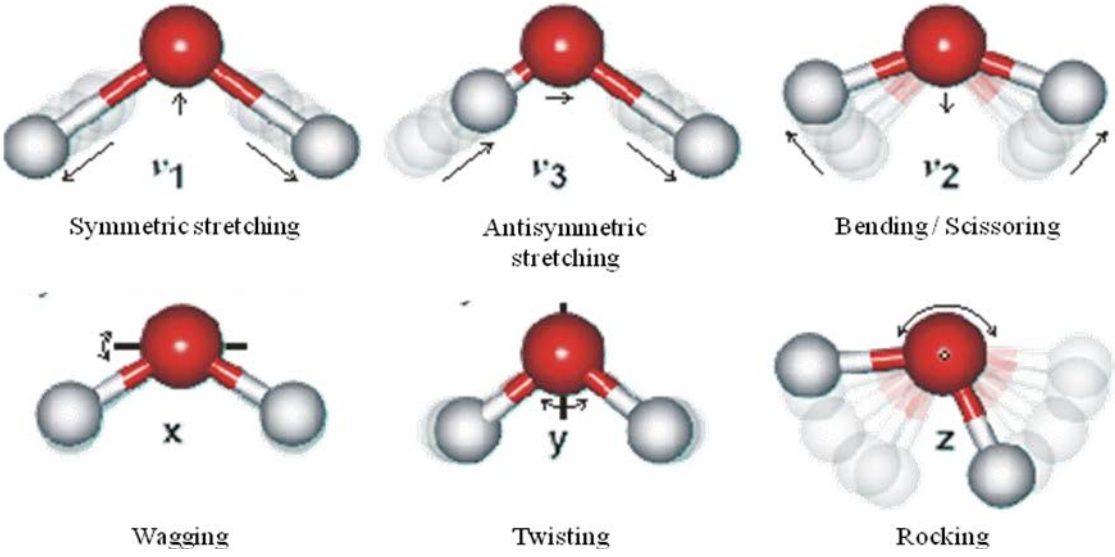


Figure 8: Different vibrational modes of molecules.

The covalent bonding could be compared to a stiff spring that can be stretched and bent. The energy required to stretch (or compress) a bond is more than to bend it. The energy or frequency that characterizes the stretching vibration of a given bond is proportional to the bond dissociation energy. The major factors that influence the stretching frequency of a covalent bond between two atoms of mass m_1 and m_2 can be represented by the equation (2) where the force constant (f) is proportional to the strength of the covalent bond between m_1 and m_2.

$$\tilde{\nu} = \frac{1}{2\pi c}\sqrt{\frac{f(m_1+m_2)}{m_1 \cdot m_2}} \qquad (2)$$

$\tilde{\nu}$ = frequency in cm^{-1}
f = the force constant
c = the velocity of light

Hence, a C=N double bond is about twice stronger than a C-N single bond and hence the energy required to stretch would also increase based on the bond strength.

When IR light is passed through an IR active sample, some of the wavelengths are absorbed by the sample and some are transmitted through. Any absorption band can be characterized by two parameters: the wavelength at which maximum absorption occurs and the intensity of absorption at this wavelength. Examination of the transmitted light reveals how much energy was absorbed at each frequency (or wavelength). It can be represented by the equation 3:

Transmittance $T = I / Io$

Absorbance $A = \log (1 / T) = \log (Io / I) = \Sigma cl$ (3)

This is also called the Beer-Lambert Law.
where,
Io = Intensity of incident radiation
I = Intensity of transmitted radiation
Σ = molar extinction coefficient ($L\ mol^{-1}\ cm^{-1}$)
c = concentration (mole / L)
l = sample path length (cm)

The absorbance or transmittance over the whole mid-IR range is represented in the form of an IR spectrum that can be exploited both qualitatively and quantitatively. This information gives the IR spectral fingerprint of the biomolecules in cells and tissues which gives an insight into the molecular composition, structural and metabolic changes occurring in these tissues.

I.4.2: Infrared spectroscopy in biomedical research

Since the mid- 20th century, IR spectroscopy coupled to microscopy was recognized with several potential applications in the field of biomedical research importantly cancer (Barer, 1949; Blout 1949; Woernley, 1952). The IR spectra originating from tissue sections and blood smears were determined in some of these studies.

The micro-level analysis became possible in 1980s with the advent of commercially available FTIR micro-spectroscopy. Further advances in the development of IR spectroscopy for biomedical studies were started by the coupling of interferometer, a microscope and an automated stage in the recent decades (Kwiatkoski, 1987; Wetzel, 1999; Gremlich, 2000). In the following years, several other biomedical studies of IR spectroscopy were carried out on cells (Wood, 1998), blood cells (Liu, 2007) cell lines (Mourant, 2003; White, 2006), etc.

Combined with an imaging set up, it was possible to acquire images of tissue sections which led to large scale biomedical applications especially as a potential cancer diagnostic tool. Several studies were undertaken to exploit these capabilities on different tissue types for discrimination between normal and the malignant features of the tissues, classification of different grades of tumors etc.

The cancerous tissues were analyzed for the diagnosis of benign and malignant lesions of breast tissue using IR imaging (Fabian, 2006). Classification of malignant glioma, including different malignancy grades, was reported (Krafft, 2007). Discrimination was achieved between basal layer, dysplastic lesions and squamous cell carcinoma of cervical tissues (Steller, 2006), and between malignant and benign nodules of thyroid (Zhang, 2010). IR imaging was also feasible in the quantification of endogenous biomolecules from the tissues originating from esophagus (Wang, 2007), and segregation of different types of prostate cells (German, 2006). These studies were carried out on fresh or frozen tissue samples. More recently, with the employment of effective spectral pre-treatment algorithms that can

neutralize the contributions of paraffin, a common tissue embedding medium, IR imaging has been applied directly to paraffinized tissues section without any chemical pretreatments. (Ly, 2008; Travo, 2010).

As far as colonic tissue are concerned, such tissues displaying abnormal spectra were detected by IR spectroscopy (Rigas, 1990; Rigas, 1992). In these studies it has been shown that the tumors display abnormal signatures compared to normal tissues involving vibrational modes such as phosphates and C-O vibrations. Following studies reported characterization of IR spectra of colon adenocarcinoma (Salman, 2001; Lasch 2002; Conti 2008). These studies presented various multivariate analyses to characterize the spectral images, and to develop topological images for histopathological verification. Also changes corresponding to RNA/DNA ratio, phosphate and carbohydrate were highlighted. Few studies on the inflammatory conditions of the colonic tissues in relation to the tumoral tissues were also reported (Argov, 2004, Katukuri, 2010). The inflammatory conditions that reflect themselves as intermediate stages between normal and malignant conditions were shown by IR spectroscopy. The importance of mucin as a diagnostic marker has been elucidated in colon cancers by IR imaging methodology (Travo, 2010). The significant IR spectral variations to discriminate between the normal and adenocarcinomatous tissues were associated with the secondary structure of mucin. In recent years Synchrotron based FTIR studies have also been implemented to cells in view of its high brightness, which permits to record high-quality spectra at diffraction-limited spot sizes (Pijanka, 2010; Pijanka, 2010).

I.4.3: Implementation of the work

Based on the capabilities of the IR spectroscopy, it was envisaged to exploit them as a potential diagnostic tool for histopathology. IR spectroscopy probes intrinsic chemical bond vibrations of biomolecules and thus provides a biochemical fingerprint of cells and tissues (Martin, 2010). Combined with an imaging device with an array detector, spectral images can be obtained rapidly in a label-free manner, in which each pixel element harbors an IR spectrum containing biochemical information at each wavenumber. Such IR images can be exploited using computer based multivariate cluster analysis to generate digital false-color maps of the tissue histology. Since the constituent IR spectra of each digitally stained histological class represent its biochemical signature, such as collagen features in the connective tissue, specific spectral signatures can be identified from different histological classes. Such signatures can

be used to train predictive algorithms for identification of unknown tissues in a rapid and user friendly manner. One of the key points in using this methodology is the automation of the protocol, which can reduce human burden and provide a biochemical based diagnostic approach.

In this regard, IR imaging in conjunction with multivariate analyses (k-means clustering and Mann-Whitney *U* test) was carried out on colonic tissue samples. Initially, a feasibility test was carried out on frozen colonic tissues. This pilot scale study constituted the basis for the development of a new concept of spectral barcodes, which enabled easy representation and interpretation of the discriminant spectral signatures between normal and tumoral colonic epithelial features.

In order to extend IR spectral imaging to larger sample size, paraffinized colonic tissue arrays from the tumor bank were recruited for the study. The tissue arrays were paraffinized and embedded in an agarose matrix which necessitated employment of pre-processing methods. Once again, a small sample set of the paraffinized tissue arrays were initially analyzed prior to extending the methodology to the entire sample set. Primarily, a modified Extended Multiplicative Signal Correction (EMSC) method was employed in order to neutralize the spectral interferences arising from paraffin and agarose. The samples were then analyzed using multivariate statistical analyses (k-means clustering, Mann Whitney *U* test, and Principal component analysis) which formed the basis for the development of the concept of IR spectral histopathology for colon tissue arrays.

Once the IR imaging methodology was standardized and established for the tissue arrays, it was expanded to the remaining tissue arrays constituting a larger sample set. The main objectives of the study were to develop spectral markers representative of various histological structures and histopathological aspects of the colonic tissues. Further, it was aimed to use these markers to construct a prediction model (based on linear discriminant analysis) to digitally detect and identify malignancy and its associated features in unknown tissues without any chemical staining, and constituting an automated diagnostic approach for colon adenocarcinoma. These objectives were aimed at validating the concept of IR spectral histopathology of colon tissue arrays on a larger scale.

Alongside, in order to characterize complementary approaches to IR spectral histopathology, other imaging modalities employing IR- attenuated total reflection (ATR) and Raman micro-spectroscopy have been tested on frozen colonic tissue samples. An approach similar to IR

spectral imaging of colon tissue arrays has also been tested on paraffinized breast tissues. This approach constituted a preliminary approach to establish IR spectral imaging for breast tissues in order to characterize IR spectral markers associated with breast cancers.

CHAPTER II

Materials and methods

II.1: Sample preparation:

II.1.1: Choice of samples

The IR spectral imaging studies have been carried out on different types of tissue samples. They principally included human colon tissues that were formalin fixed, paraffin embedded and obtained in the form of tissue arrays, and secondarily obtained in the form of frozen tissues. Additionally, in a feasibility study, this approach was extended to formalin fixed paraffinized human breast tissues. In total, 84 paraffinized human colon tissues (33 non-tumoral, 47 tumoral, and 4 adenomatous) from 39 individuals were analyzed. In the case of frozen human colon tissues, 12 samples (7 non-tumoral and 5 tumoral) from 7 patients were analyzed. The number of paraffinized breast tissues analyzed was 32 (16 non-tumoral and 16 tumoral) from 16 patients. All the samples were obtained with the approval of the Institutional Review Board of CHU Reims. The sample details are presented in the table I.

Table I: Sample numbers utilized in the study

	Colon tissue array (paraffinized)	**Colon tissues (frozen)**	**Breast tissues (paraffinized)**
Number of samples	**84**	**12**	**32**
Non-tumoral	33	7	16
Tumoral	47	5	16
Adenomatous	4		
Number of patients	39	7	16

II.1.2: Tissue array

The principal tissue types analyzed in this study consisted of normal and moderately differentiated colon adenocarcinoma, which were obtained in the form tissue arrays. The tissue arrays consisted of an assembly of selected tissue spots originating from different paraffinized tissue blocks. Such arrays facilitated multiplexing the samples on a single optical substrate for measurements of the tissues without disturbing the sample compartment. The

tissue arrays were constructed manually in the pathology laboratory as shown in figure 1. The tissue arrays were embedded in paraffin and stabilized in an agarose matrix. For this, 4 % agarose was cast in a mold to obtain a gel of 2 to 3 mm thick. After polymerization, it was cut to the dimension of a standard laboratory embedding cassette (28.5 x 41 x 6.7 mm). The cassette with the solidified agarose was then passed into the vacuum infiltration processor (VIP) instrument for automated standard protocol of dehydration of agarose, and paraffin infiltration, as commonly used for tissues. Once chilled, this paraffin-agarose matrix then served as the recipient block for constructing the tissue arrays.

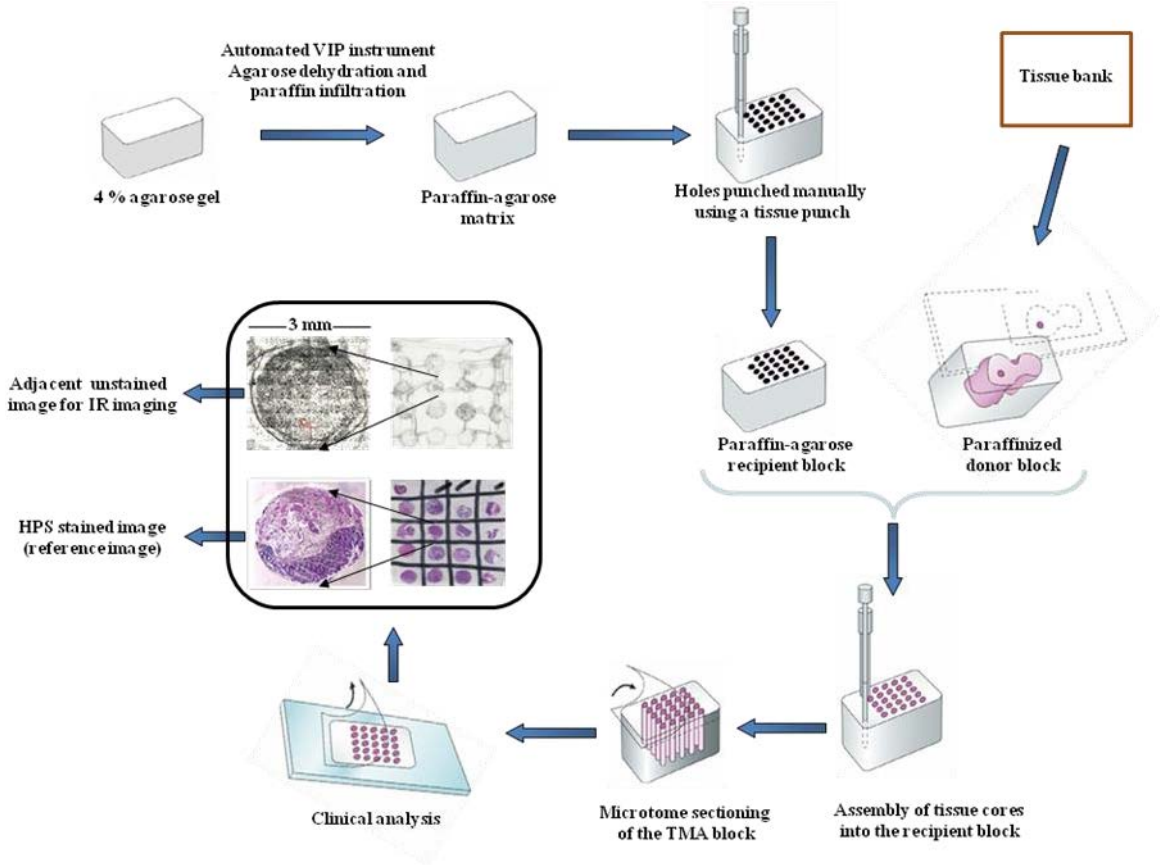


Figure 1: A schematic representations of a manual tissue array slide construction.

Holes were punched in the recipient block manually using a tissue punch of diameter 3 mm. Paraffin blocks containing non-tumoral and tumoral colon tissues served as the donor blocks. Cylindrical tissues cores were excised from the donor block using a tissue punch and

introduced into the punched holes, in the paraffin-agarose recipient block in an array of choice with defined coordinates. The most stable cylindrical tissue cores were placed on the top row of the tissue array block for maximum mechanical stability and to ensure uniform microtome sectioning. The tissue array block was then transferred into an inclusion mold and filled with molten paraffin to completely integrate the cylindrical tissue cores into the matrix. This paraffin-agarose block is now referred to as the tissue array block. Once cooled, the tissue array blocks were taken out of the molds and stored at normal room conditions until sectioning. Each tissue array block consisted of 12 to 16 tissue cores (3 mm in diameter) of selected colon tissue type. Before sectioning the paraffin-agarose tissue array block, the block was chilled to -20C for 30 min.

For the initial feasibility study of IR spectral imaging, and complementary approaches involving IR-ATR, and Raman micro-spectroscopies, the frozen human colorectal tissues were sectioned (10 μm) using cryo-microtome and an optimal cutting temperature (OCT) embedding medium. For the preliminary study on human breast tissues, the samples were obtained in the form of paraffinized microtome sections (10 μm).

II.2: Instrumentation:

II.2.1: FTIR spectral imaging system

The FTIR imaging system consisted of a microscope (Spectrum Spotlight 300, Perkin Elmer, France) coupled to a spectrometer (Spectrum One, Perkin Elmer, France) (figure 2). The microscope was equipped with liquid nitrogen-cooled 16-element MCT detector and a visible camera (resolution power 0.8 μm) that enabled to capture visible images of the sample via the microscope. Guided by a motorized stage, this permitted to select the regions of interest of the samples for IR spectroscopy. The visible images were obtained under a LED white light illumination source. The spectrometer contained a Globar ® source that was projected in such a way that it coincided with the selected zone of the white light image. The imaging system enabled collecting IR spectral information either in point mode where a single element MCT detector was used or in the image mode where the multi-element (16 pixels) detector was

used. Both MCT detectors are placed in the same Dewar. One detector is selected at a time using a mask. The system has been upgraded with a second Dewar placed on top of the first one that serves as a liquid nitrogen reservoir. In this way, autonomy of 23 hours was reached, allowing the imaging of large samples. In the image mode, two options were available: pixel size of 6.25x6.25 μm^2 or 25x25 μm^2 for rapid sampling.

Further, in an effort to limit the effects of atmospheric (CO_2 and water vapor) contribution on the spectra, the sample compartment was equipped with a purge box where a continuous flow of dry air was maintained. In this way, a stable environment was reached.

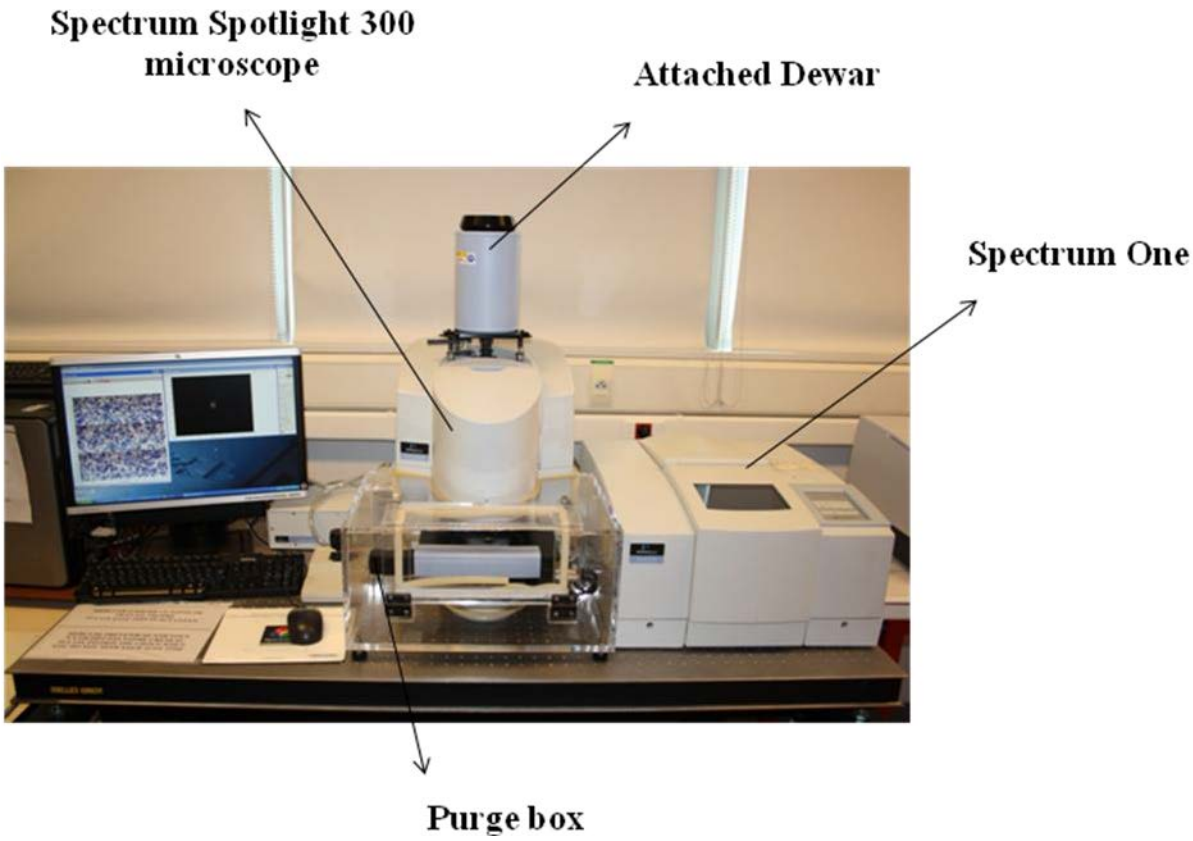


Figure 2: The infrared imaging system (Spotlight 300 from Perkin Elmer, Courtaboeuf, France).

An interferometer enabled to examine all frequencies in a wide spectral range simultaneously. A schematic representation of the FTIR instrumentation and a Michelson interferometer is shown in figure 3.

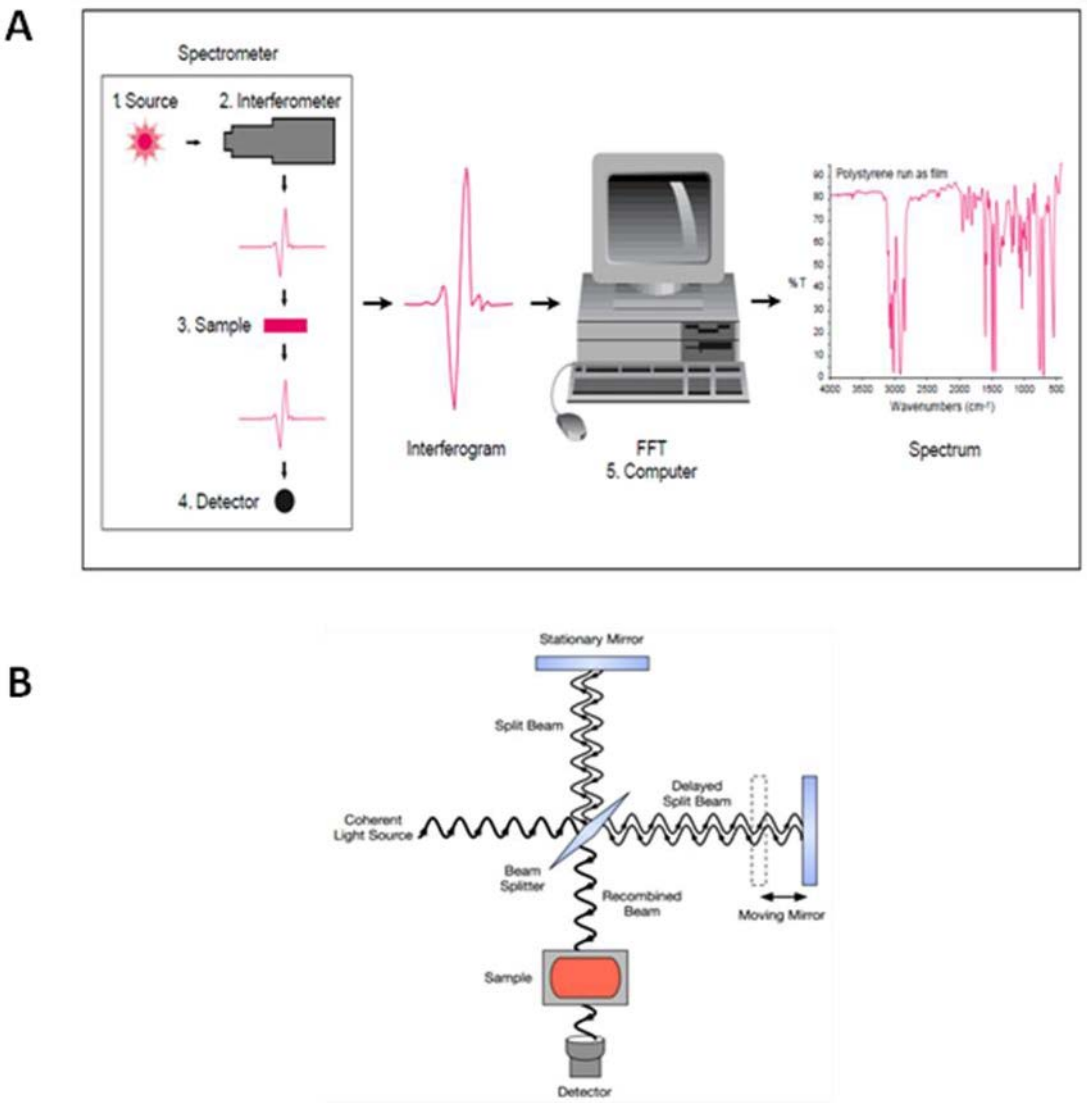


Figure 3: A typical representation of the instrumentation of a FTIR spectrometer (A), and, a schematic of a Michelson interferometer (B).

In this system, an interference pattern is produced by splitting the IR light into two paths by using a KBr beam splitter, and then recombining the light beams. In the recombined beam, the superposition of the signals from different wavelengths generates an interferogram. The interferogram contains all the vibrational information and is displayed as the light intensity as a function of the optical path difference. The interferogram is then converted into a spectrum using a Fourier transform algorithm, over the entire mid-IR spectral range. A typical IR spectrum consisted of wavenumbers presented on the X-axis and the intensity of absorption

(or % transmission) at each wavenumber on the Y-axis. The absorbance is proportional to the concentration of the analyte based on the Beer-Lambert's law (see equation 3).

II.2.2: FTIR spectral image acquisition methodology

A schematic representing the methodology for IR imaging of a spot from a tissue array is shown in Figure 4. The IR images were obtained from ten microns thick microtome sections cut from the tissue array block. The first 10 μm section was chemically deparaffinized for conventional histopathological analysis via HPS staining that served as a morphological reference. An adjacent 10 μm thick paraffinized tissue section was directly mounted onto an IR transparent calcium fluoride (CaF_2) (Crystran, UK) window for IR imaging without any chemical deparaffinization. This procedure was followed for all the tissue array blocks included in the study. Spots were selected for IR analysis by an expert pathologist using a 3 μm HPS stained image as the reference that provided finer tissue details.

The acquisition parameters used were 6.25x6.25 μm^2 pixel size, and 4 cm^{-1} spectral resolution averaged over 16 scans in the mid-IR range from 750 to 4000 cm^{-1}. These parameters permitted to obtain good quality IR images which after multivariate processing enabled high-degree of morphological correlation to the reference HPS image. Before settling on these parameters, IR imaging using a pixel size of 25x25 μm^2 was tested. However, the spectral images were not resolved enough in order to provide good enough correlation to the reference HPS images. Similarly, to reduce the acquisition time, spectral resolution of 4 cm^{-1} was used averaged to 8 scans per pixel. However, the images acquired using 16 scans per pixel at same spectral resolution gave marginally better correspondence to the reference images.

Each time, prior to an image acquisition, a background spectrum of the CaF_2 window was acquired which was automatically subtracted from each pixel spectrum. The same methodology and similar acquisition parameters of IR imaging were kept for other tissue types in this study (frozen colon tissue samples and paraffinized breast tissue samples).

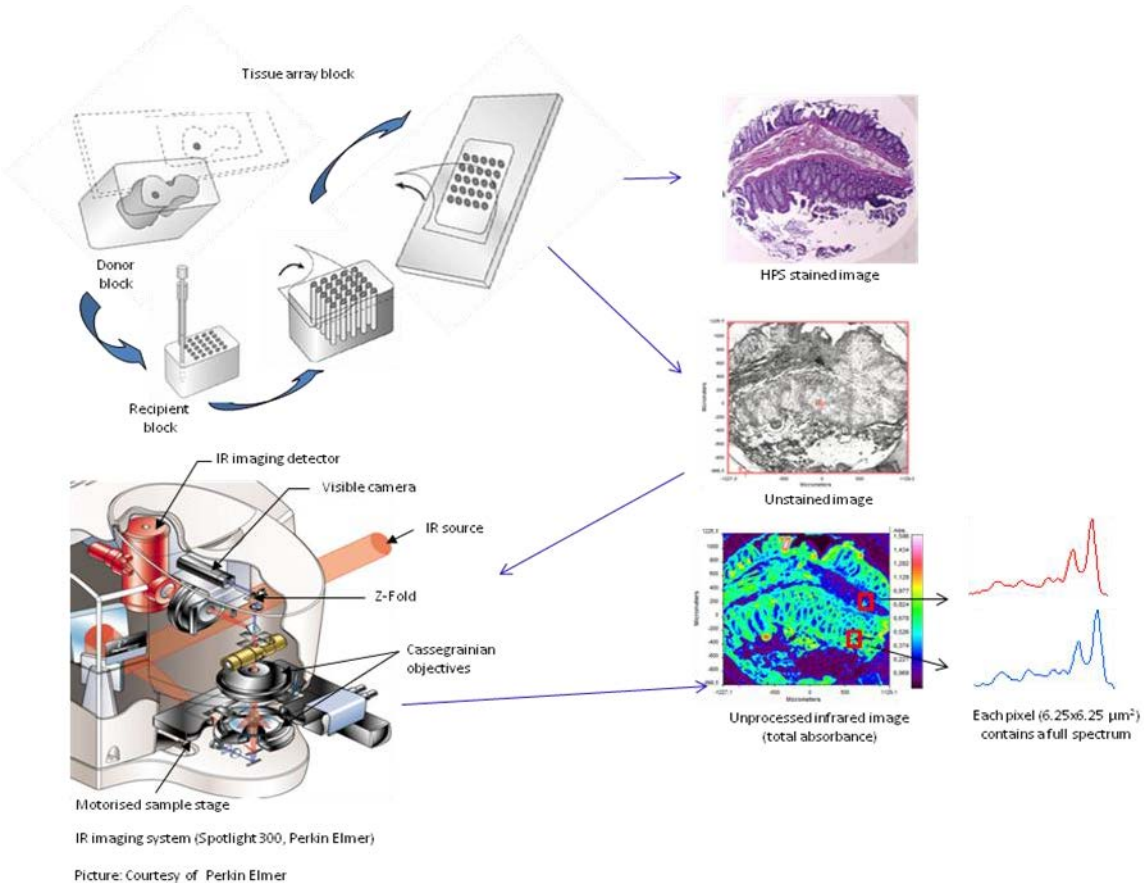


Figure 4: A schematic showing the adapted methodology for infrared spectral imaging of a tissue array spot.

II.3: Data pre-processing:

II.3.1: Pre-processing of IR spectra from paraffinized tissues arrays

It is known that the interaction of light with matter gives rise to physical effects that can contribute to the tissue spectra. The physical information needs to be identified and be separated in order to exploit the biochemical information. Therefore, a pre-processing step was essential after image acquisition, to eliminate or correct, different kinds of inherent and extrinsic spectral interferences originating during image acquisition. The raw tissue spectra from the IR images of the paraffinized tissue arrays are quite complex since they contain contributions from the physical effects (e.g. scattering), atmospheric absorptions of water

vapor and CO_2, chemical absorptions of paraffin and agarose, and biochemical absorptions from the tissue itself. With the purpose of preserving only the biochemical information, stringent pre-processing steps were employed, to neutralize the interfering, non-informative contributions. In order to achieve this, atmospheric correction of water vapour and CO_2 was performed on each pixel by the built-in software of Spectrum IMAGE (Perkin Elmer, Version R 1.6.4.0394). Further analyses were performed using in-house algorithms written in Matlab 7.2 (The Mathworks, Natick, MA). A modified EMSC model (Extended Multiplicative Signal Correction) (Ly, 2008) was used for correcting paraffin, agarose, and baseline interferences, followed by normalization. Pre-processing, processing and analysis of the IR spectra were carried out on spectral images in the IR absorption range of 900-1800 cm^{-1}, considered as the most informative IR spectral region (Khanmohammadi, 2009; Khanmohammadi, 2010) as far as the tissue features are concerned.

II.3.2: Construction of EMSC model

EMSC was developed to correct the spectra from the physical light scattering effects that are different from the chemical light absorbance effects (Martens, 2003; Kohler, 2005). Along with the biochemical information, FTIR spectra of paraffinized colon tissue array sections exhibited absorption bands of paraffin (1378 cm^{-1} and around 1467 cm^{-1}) and agarose (at 1072 cm^{-1} and minor peaks at 932 cm^{-1}, 1155 cm^{-1} and 1185 cm^{-1}) in the 900-1800 cm^{-1} spectral region (Figure 5).

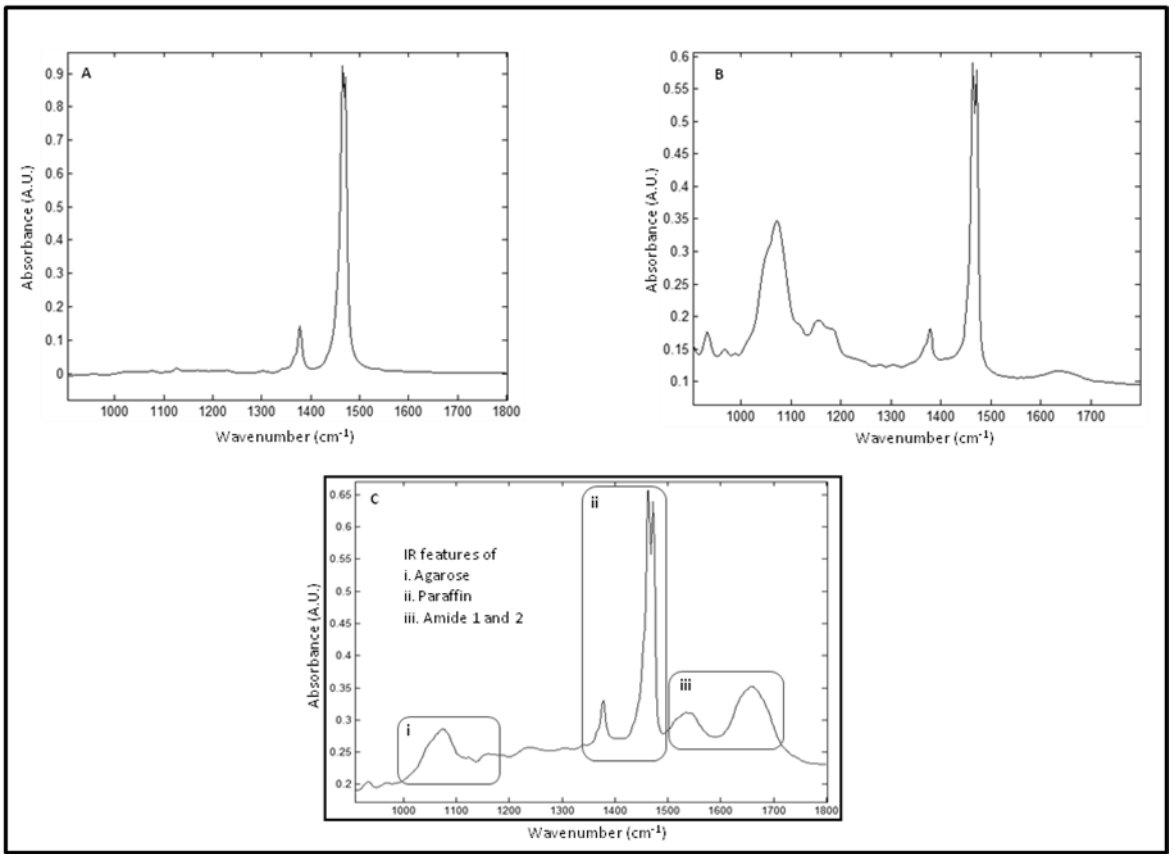
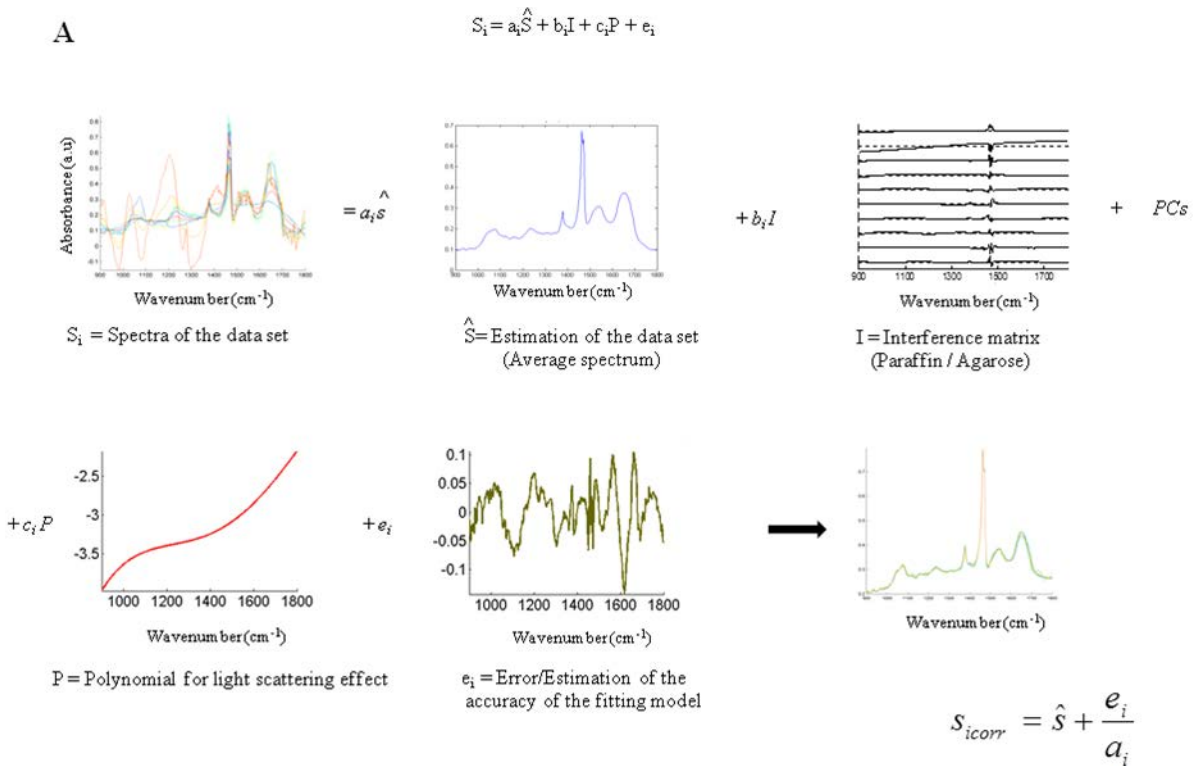


Figure 5: Infrared spectral features of paraffin, agarose and colonic tissue.

Average IR spectra of paraffin (A), paraffin and agarose mixed together (B), and a paraffinized colon TMA section (C) which includes spectral information from tissue, paraffin and agarose in the spectral range of 900 - 1800 cm^{-1}.

For efficient classification and understanding of the biochemical nature of the tissue, the variability of these contributions had to be neutralized and their influence circumvented, for which a modified EMSC algorithm was employed. The workflow of EMSC algorithm is depicted in figure 6.



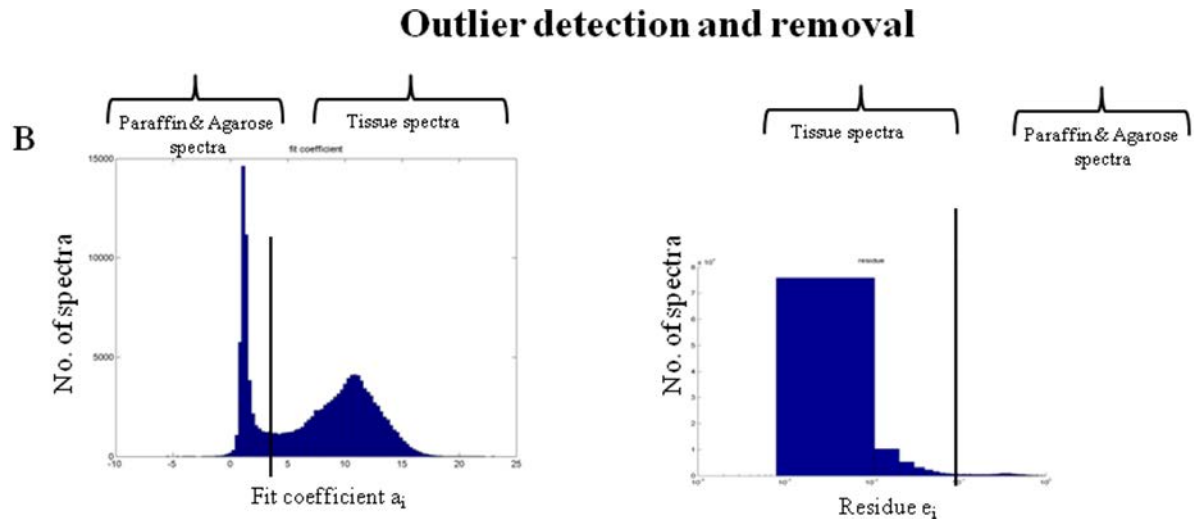


Figure 6: The depiction of the EMSC algorithm for mathematical neutralization of paraffin and agarose contributions (A), and the outlier spectra detection and removal (B).

EMSC models linearly each spectrum of the data set as:

$$\mathbf{s_i} = a_i\hat{\mathbf{s}} + \mathbf{b_i}\mathbf{I} + \mathbf{c_i}\mathbf{P} + \mathbf{e_i} \qquad (4)$$

where,

$\mathbf{s_i} \in \mathbb{R}^{1\times n}$ is the i[th] acquired spectrum of the data set, i.e., a vector composed of n points

$\hat{\mathbf{s}} \in \mathbb{R}^{1\times n}$ is the target spectrum that is chosen as the mean spectrum of the studied dataset

$\mathbf{I} \in \mathbb{R}^{k\times n}$ is the interference matrix composed of k components

$\mathbf{P} = \begin{pmatrix} \nu_1^0 & \dots & \nu_1^p \\ \vdots & \ddots & \vdots \\ \nu_n^0 & \dots & \nu_n^p \end{pmatrix}^T \in \mathbb{R}^{(p+1)\times n}$ is the transpose of the Vandermonde matrix of the n wavenumbers ν_j; this matrix is used to compute $\mathbf{c_i}\mathbf{P}$, a p-order polynomial function modeling for light scattering effect, i.e., for baseline correction.

$\mathbf{e_i} \in \mathbb{R}^{1\times n}$ is the model error vector

a_i is the scalar fitting coefficient of $\hat{\mathbf{s}}$ to $\mathbf{s_i}$

$\mathbf{b_i} \in \mathbb{R}^{1\times k}$ is the vector of the fitting coefficients of $\mathbf{I}$ to $\mathbf{s_i}$

$\mathbf{c_i} \in \mathbb{R}^{1\times(p+1)}$ is the vector of the fitting coefficients of $\mathbf{P}$ to $\mathbf{s_i}$ and represents the coefficient of the p-order polynomial function.

The coefficients a_i, $\mathbf{b_i}$ and $\mathbf{c_i}$ are estimated by the traditional least squares method in order to minimize the model error $\mathbf{e_i}$. The corrected spectra could be then represented by the equation

$$\mathbf{s_{i}}_{corr} = \hat{\mathbf{s}} + \frac{\mathbf{e_i}}{a_i} \qquad (5)$$

Figure 7 depicts the flowchart of the EMSC protocol that has been used to realize several corrections. Firstly, it corrects spectra from paraffin and agarose contributions. Secondly, it

corrects spectra for light scattering effects, and thirdly, it normalizes spectra on the mean spectrum $\hat{\mathbf{s}}$.

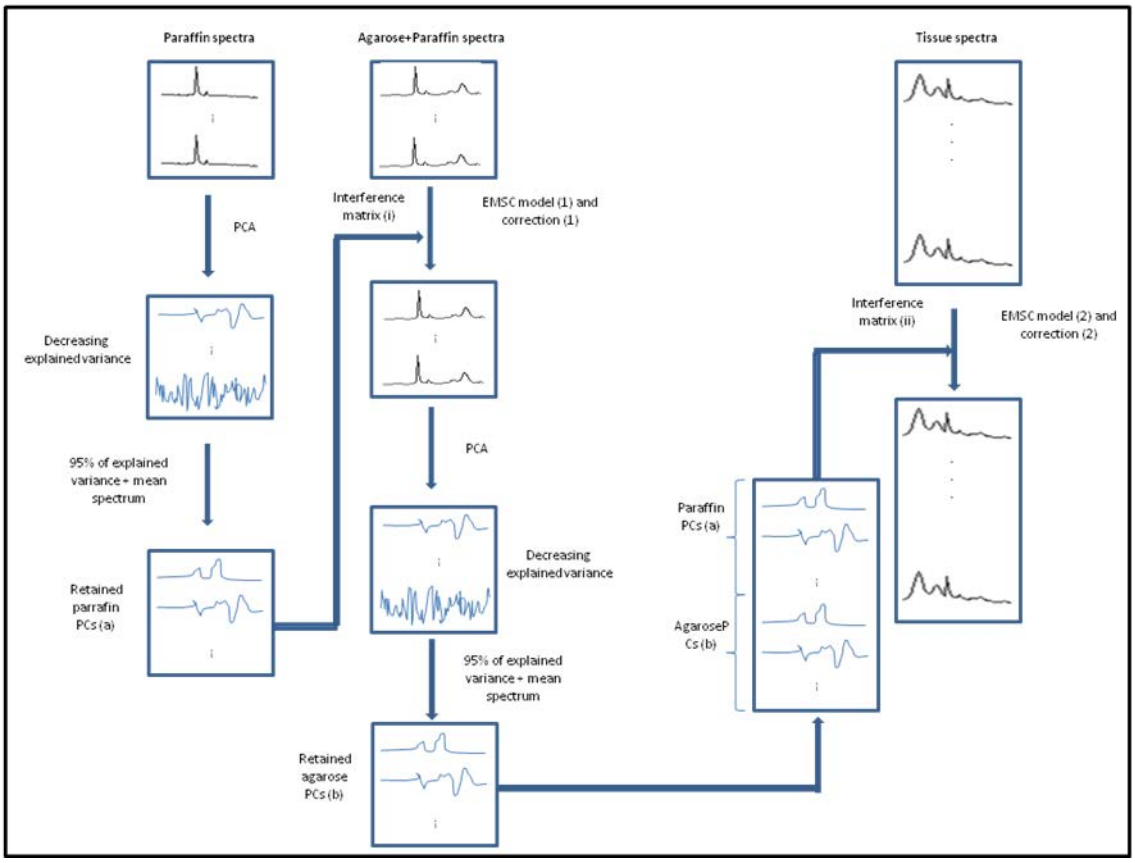


Figure 7: The depiction of a flowchart for the EMSC model for spectral pre-processing.

Briefly, in order to achieve these corrections, an IR image consisting of 13516 spectra was acquired from 10 µm thick paraffin (used for tissue embedding in the pathology laboratory) section using the same spectral parameters as that of the tissue array images. PCA was performed on these spectra to model them with few orthogonal components best explaining the variance of paraffin. The interference matrix $\mathbf{I}$ of model was constructed by retaining the first 10 principal components (PCs) and the mean spectrum of paraffin. Another IR image consisting of 15872 spectra was acquired from a 10 µm section of a mixture of paraffin and agarose, as agarose is a semisolid matrix (at 2% used for tissue array construction) and could not be sectioned alone. The spectra of this image were then modeled using equation (4) in which a fourth order polynomial function is assumed to construct $\mathbf{P}$ to model the baseline. Paraffin contributions were then neutralized from agarose, by applying the correction in

equation (5). Next, PCA was performed on these paraffin corrected agarose spectra in order to model the IR signal of agarose. The first 10 significant PCs and the mean spectrum of agarose were then added to the interference matrix $\mathbf{I}$. $\mathbf{I}$ is thus composed of 11 components modeling paraffin and 11 components modeling agarose. $\mathbf{I}$ being constructed and a fourth order polynomial function being still assumed for $\mathbf{P}$, the model (1) was applied to the colon IR spectral images acquired from the paraffinized biopsies. The entire data set was then corrected for the contributions of paraffin and agarose, baseline corrected, and normalized on the entire spectral range using equation (5). Furthermore, a thresholding of a_i and $E = \sum_{j=1}^{n} \left(\frac{\mathbf{e_i}(j)}{a_i} \right)^2$ permitted to detect the outlier spectra of paraffin and agarose, and to eliminate them from further analysis. In the k-means classified images (see below), the pixels corresponding to these outliers are colored white.

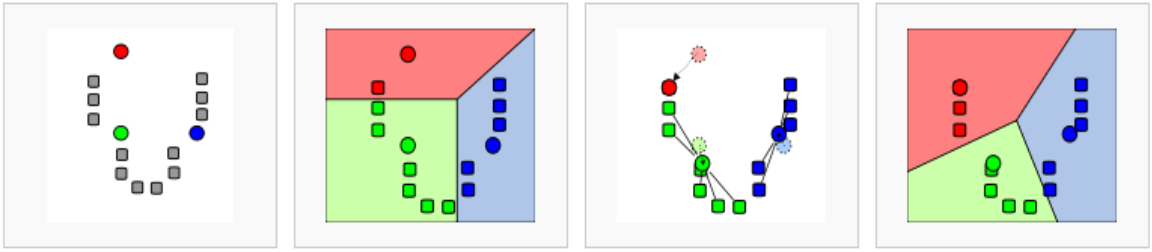
II.3.3: Pre-processing of IR spectra from other tissue types

Similar treatments of atmospheric correction, EMSC based neutralization of paraffin correction (without agarose model) along with baseline and normalization were performed on the IR spectral images of the breast tissues. However, for the frozen tissue samples since there were no paraffin interferences, EMSC was employed for baseline correction and normalization without the model for mathematical deparaffinization.

II.4: Multivariate data analysis and processing:

After pre-processing, the spectra were subjected to various multivariate statistical tests. Initially, k-means clustering algorithm was applied to partition the IR spectral images into clusters in order to recover the histological organization. This method iteratively partitions the spectra into different classes based on the spectral distances (Figure 8). In the initialization phase, K spectra (K is the number of searched clusters) are randomly chosen to represent initial centroids which model the mean spectrum of each cluster. Second, each spectrum is affected to the cluster with the nearest centroid according to the Euclidean distance. Third, each centroid is updated as the mean of the spectra belonging to its cluster.

Steps 2 and 3 are repeated until the convergence of the algorithm. Therefore, spectra with similar biological characteristics fall into the same cluster and spectra with different biological characteristics fall into different clusters.



1) *k* initial "means" (in this case *k*=3) are randomly selected from the data set (shown in color).

2) *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means

3) The centroid of each of the *k* clusters becomes the new means.

4) Steps 2 and 3 are repeated until convergence has been reached.

Figure 8: The demonstration of the k-means algorithm

To overcome the dependency of k-means on the initialization step for partitioning the data, six consecutive and independent runs were performed per image to classify the spectra. Out of these classifications, the cluster image showing the highest association to the reference histopathological image was selected for further analysis. In k-means, each spectrum belongs to a unique cluster and can thus be represented by a unique color distinct from those of the remaining clusters and a color coded image can be reconstructed for rapid and simple visual analysis of the clustering results. These were then compared to adjacent conventionally stained sections to annotate each spectral cluster to the tissue structural feature that it belongs to by an expert pathologist. K-means clustering provided label-free histological maps of the tissue samples. Further, it permitted to retrieve the tissue specific spectral signatures for biochemical interpretation via statistical analyses.

Statistical analysis

Different types of statistical tests were used in the study. From the k-means clustered images, the spectra representing various histological features were subjected to Mann-Whitney U test and principal component analysis (PCA) in order to find the most discriminant spectral regions. Further, the spectral differences between the compared groups were evaluated using

PC scores and PC loadings. PCA is one of the commonly employed spectral data processing method which reduces the size of the data still retaining the variance. This variance is represented by PCs and the first PC represents the maximum variance in a mean centered data.

The k-means clustered images were also used to develop a prediction model based on linear discriminant analysis (LDA) for automated recognition of tissue features, to enable identification and localization of the tumor in unknown samples. LDA is a supervised technique which aims at maximizing the between-class variance and minimizing the within-class variance. For the development and execution of the prediction model, spectral signatures representing tumor and other histological classes were identified from the initially classified k-means images. A representation of the LDA prediction procedure is shown in figure 9, and explained in detail in section III.4 of Chapter 3.

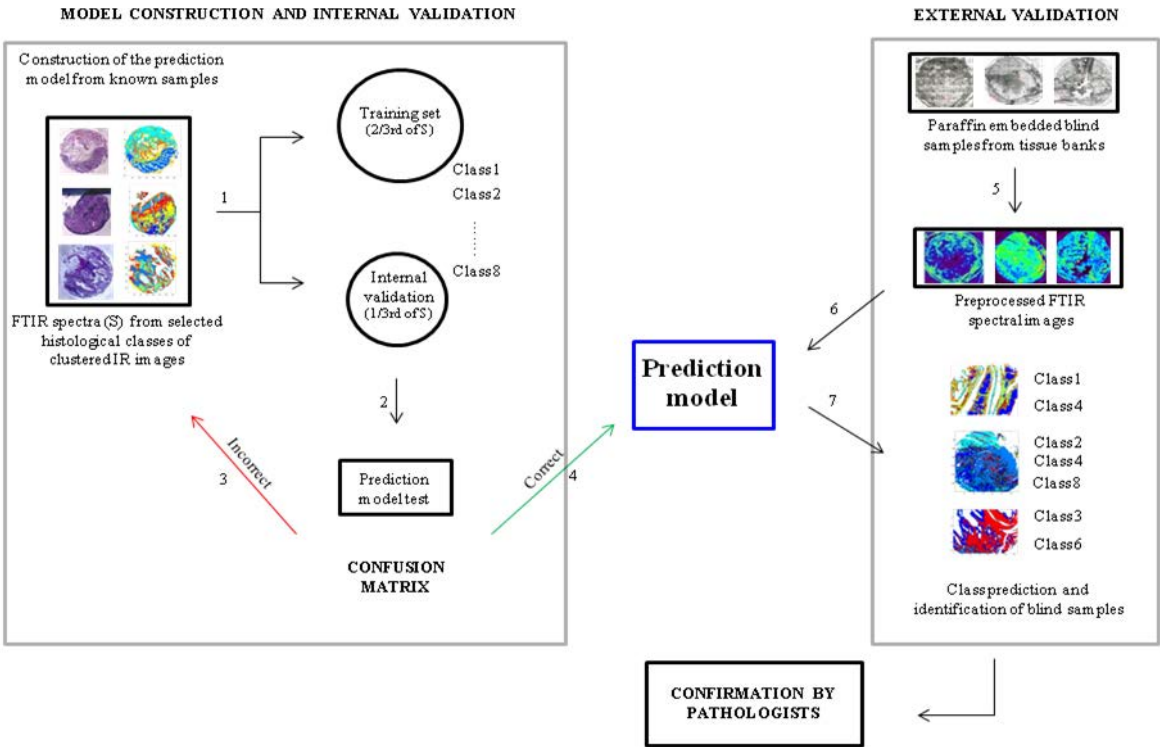


Figure 9: A schematic representation of construction and application of the prediction model based on linear discriminant analysis.

Similar spectra showing such signatures were grouped into a unique class with a unique label specific to the original histological structure that it belongs to, and added to the model. After several trials, the model presenting the best sensitivity was then applied in an external

validation on unknown samples that were secluded to the model. For the prediction, a posterior probability of 0.5 was used.

II.5: Immunohistochemistry (IHC):

IHC was used as a complementary tool (on adjacent sections) to enhance visibility of tumor budding (Anti-Human Cytokeratins-large spectrum Monoclonal Antibody, Clone KL 1, dilution 1/50, Immunotech, France) and precise the nature of the inflammatory cells: T-lymphocytes (CD3 Rabbit anti-Human Polyclonal Antibody, dilution 1/200, Dako, France), and B-lymphocytes (CD20 Mouse antibody, clone L6 mouse, dilution 1/400, Dako, France), in order to validate some of the important observations detected by IR spectral imaging. This was performed using the fully automated IHC staining protocol (XT ultraView DAB v3).

CHAPTER III

Results and discussion

III. Infrared spectral histopathology: Concept and application to colon cancers:

III.1: Résumé :

Le cancer est l'une des principales causes de mortalité dans le monde. De nouvelles technologies, capables de fournir un aperçu des signes qui se produisent au cours de la cancérogénèse, sont régulièrement testées afin de complémenter les méthodologies existantes de diagnostic. Parmi ces techniques, l'approche biophotonique par imagerie spectrale IR paraît être une méthode d'intérêt car elle mesure les vibrations des liaisons chimiques présentes au sein des cellules et des tissus, et est capable de fournir un aperçu des changements biochimiques qui ont lieu au cours des stades précoces de la cancérogénèse. Dans ce contexte, l'objectif principal de ce travail était de développer une méthodologie de diagnostic automatique basée sur l'imagerie spectrale IR pouvant être utilisée comme un outil complémentaire à l'histopathologie conventionnelle.

Le premier article de ce chapitre présente tout d'abord le test de faisabilité qui a été menée afin d'établir une méthodologie d'imagerie spectrale IR sur des coupes de tissus coliques congelées. Les images spectrales IR obtenues ont été traitées par une méthode de clustering non supervisée afin de classer les spectres en fonction des différentes structures histologiques des tissus coliques selon leurs caractéristiques intrinsèques biochimiques. Un test statistique a ensuite été utilisé afin d'introduire, pour la première fois dans cette étude, un nouveau concept de code-barres spectral IR, construit à partir des signatures spectrales de différentes classes histologiques. Le code-barres constitue un outil de représentation et d'interprétation des marqueurs spectraux discriminants entre les tissus normaux et tumoraux de manière facile et simplifiée. Cette approche, en combinaison avec une analyse multivariée, a été réalisée sur un petit nombre d'échantillon (N = 10).

Afin de poursuivre cette approche d'imagerie spectrale IR sur un nombre plus important d'échantillon provenant de la banque de tumeurs, la méthodologie, en combinaison avec l'analyse statistique multivariée, a été appliquée sur des tissus arrays de côlon paraffinés, représentant une riche source d'information pour une étude à haut débit.

Le deuxième article de ce présent chapitre porte sur la première partie du travail qui a été l'adaptation de la méthodologie d'imagerie spectrale IR sur un petit nombre de tissu arrays sélectionnés (N = 6). Étant donné que les échantillons ont été obtenus sous la forme de tissus stabilisés au sein d'une matrice d'agarose et paraffinés, un prétraitement des données est nécessaire afin de corriger les interférences spectrales de ces deux constituants. Pour cela, une méthode de correction appelée Extended Multiplicative Signal Correction (EMSC) a été appliquée de manière sélective dans le but de neutraliser les contributions parasites de l'agarose et de la paraffine, et ainsi conserver uniquement les informations biochimiques spécifique du tissu.

En raison de la nature complexe des spectres IR obtenus à partir des tissus, des analyses statistiques multivariées ont ensuite été utilisées pour extraire les informations les plus discriminantes. Initialement, des algorithmes de clustering ont été utilisés dans le but de construire des images selon un code couleur en fonction de la distribution spatiale et biochimique des principales caractéristiques histologiques des tissus du côlon. Ceci nous a permis d'attribuer des signatures spectrales spécifiques aux différentes structures histologiques. Sur la base de ces signatures, il a été possible de mettre en évidence les différences biochimiques qui se produisent au cours de la cancérogénèse.

Les spectres extraits des tissus normaux et tumoraux ont été comparées en utilisant le test de Mann-Whitney U et l'ACP, afin d'identifier les altérations biochimiques associées à la cancérogenèse. En effet, les variables spectrales discriminantes identifiées ont été corrélées à diverses biomolécules impliquées dans le cancer du côlon comme les phosphates, les glucides, les protéines etc. Cette approche représente la preuve d'un nouveau concept d'histopathologie spectrale IR appliqué aux tissu arrays de côlon.

Ce concept a ensuite été validé sur les tissu arrays restant (N = 80), tel que présentés dans le troisième article de ce chapitre. L'imagerie spectrale IR combinée à une analyse multivariée a été mise en œuvre pour construire un modèle de prédiction pouvant automatiquement prédire la présence d'un cancer dans des échantillons inconnus. Les mêmes prétraitements ainsi que les mêmes traitements par EMSC et par clustering non supervisé ont été utilisées respectivement pour éliminer les interférences parasites et pour extraire les signatures spectrales des différentes caractéristiques histologiques du côlon. Le modèle de prédiction ayant été développé à partir de ces signatures, a ensuite été testé pour effectuer un diagnostic des échantillons coliques inconnus. Cette approche d'histopathologie spectrale IR a permis

non seulement d'identifier l'histopathologie des tissus d'une manière automatisée, mais également de mettre en évidence certaines caractéristiques particulières importantes associées à la tumeur comme le phénomène de tumeur budding, l'interaction de la tumeur et du stroma, etc.

Enfin, pour affirmer la capacité du modèle de prédiction, l'analyse a été étendue à des échantillons adénomateux à bas et haut grade dysplasie. Les résultats sont prometteurs pour l'identification des changements moléculaires précoces associés à ces échantillons pré-cancéreuses.

III.2: Summary:

Cancer is one of the leading causes of mortality around the world. Modern technologies capable of providing insights into signs occurring during carcinogenesis are continuously under scrutiny in order to complement the existing diagnostic methods. Of these techniques, the biophotonic approach of IR spectral imaging is an interesting candidate as it measures the chemical bond vibrations in cells and tissues and thus provides insights into the biochemical changes occurring on the advent of carcinogenesis. In this perspective, the main aim of this work was to develop an automatic diagnostic methodology based on IR spectral imaging that could be used as a complementary aiding tool to conventional histopathology.

The first article in this chapter presents the initially feasibility test that was carried out in order to establish the IR spectral imaging methodology on frozen colonic tissue sections. The IR spectral images obtained were treated with an unsupervised clustering method in order to spectrally classify the various histological features of the colonic tissue based on their intrinsic biochemical features. A statistical test was then employed on the spectral signatures of the histological classes, based on which a new concept of IR spectral barcode was introduced for the first time in this study. The barcode constituted an easy-to-interpret representation of the discriminant spectral markers between the normal and the tumoral tissues. This approach in combination with multivariate analysis was carried out on a small sample size (N=10).

In order to pursue with the approach of IR spectral imaging on a larger sample size that was accessible from the tumor bank, the IR spectral imaging methodology in combination with multivariate statistical analysis was implemented on paraffinized colonic tissue arrays which represent a rich source of information for high-throughput studies.

The article 2 of this chapter presents the initial part of the work that deals with the IR spectral imaging methodological adaptation on a small number of selected tissue array cores (N=6). Since the samples were obtained in the form of paraffinized tissue arrays that were stabilized in an agarose matrix, a data pre-processing step was necessary for correction of spectral interferences. For this, a modified extended multiplicative signal correction (EMSC) was applied to selectively neutralize parasitic contributions of paraffin and agarose and to retain only the biochemical information originating from the tissue.

Due to the complex nature of the IR spectra obtained from tissues, multivariate statistical analyses were implemented to extract discriminant information. Initially, clustering algorithms were used for constructing color-coded spectral maps that provided the spatial and biochemical distribution of the main histological features of the colonic tissues. This then enabled to retrieve spectral signatures specific to different histological structures. On the basis of these signatures, it was possible to investigate the biochemical differences occurring with the development of cancers.

The retrieved spectra from the normal and the tumoral tissues were compared using the Mann-Whitney U test and PCA to identify these biochemical alterations. Indeed the identified discriminant variables were correlated to various biomolecules implicated in colon cancers such as phosphates, carbohydrates, proteins etc. This approach represented the proof-of-concept IR spectral histopathology of colon tissue arrays.

This concept was then validated on the remaining tissue arrays (N=80) as presented in the third article of this chapter. The IR spectral imaging combined with multivariate analysis was implemented to construct a prediction model that can automatically predict the presence of cancer in unknown samples. Similar pre-processing and processing by EMSC and unsupervised clustering were employed to eliminate parasitic interferences and to extract spectral signatures of the colonic histological features respectively. The prediction model was developed from these signatures, which was then tested to diagnose unknown colonic samples. This approach of IR spectral histopathology permitted not only to identify the histopathology of the tissues in an automated manner, but also to highlight certain important tumor-associated features like tumor budding, tumor-stroma association etc.

Finally, to affirm the applicability of prediction model, the analysis was extended to few adenomatous samples with low to high grade dysplasia. The results are promising in identifying the early molecular changes associated with these pre-cancerous samples.

III.3: Article 1

Infrared imaging as a cancer diagnostic tool: introducing a new concept of spectral barcodes for identifying molecular changes in colon cancers

(Submitted to the Journal of Cytometry Part A, July 2012)

Préambule à l'article 1

Contexte

A l'heure actuelle, l'histopathologie constitue la méthode de référence pour le diagnostic du cancer basé sur l'identification des modifications morphologiques dans les tissus symptomatiques. De nouvelles approches capables de fournir des informations biomoléculaires complémentaires à l'histopathologie conventionnelle sont en cours de développement. Dans cette démarche, une approche biophotonique basée sur la micro-imagerie spectrale infrarouge, combinée à une analyse statistique multivariée a été mise en œuvre sur les tissus du côlon.

Objectif

L'objectif de ce travail a été de développer un nouveau concept de code-barres spectral basé sur les caractéristiques intrinsèques biochimiques des cellules et des tissus grâce au couplage de l'imagerie infrarouge qui permet d'exploiter un grand volume de données spectrales avec l'analyse multivariée des images.

Matériels et Méthodes

Afin de mettre en œuvre ce concept, dix échantillons congelés de tissus de côlon provenant de 5 patients (un normal et un tumoral par patient) ont été analysés par micro-imagerie spectrale infrarouge de manière non-destructive. Les images spectrales ont ensuite été traitées par une méthode de classification multivariée (le clustering par k-means) afin de déterminer l'organisation histologique. Pour chaque patient, l'information spectrale correspondant à l'épithélium normal et tumoral est automatiquement récupérée et comparée à l'aide d'une méthode statistique (test de Mann-Whitney *U*). Ceci permet de faire ressortir des éléments discriminants qui sont ensuite utilisés pour construire des code-barres spectraux spécifiques de chaque tissu.

Résultats

Les code-barres spectraux représentant les nombres d'onde discriminants ont permis la caractérisation des altérations biochimiques d'une part, de la mucine associées à la malignité, et d'autres parts, des nucléotides, des glucides et des protéines. Cette approche a non seulement permis d'identifier des altérations biochimiques communes entre tous les patients atteints de cancer du côlon, mais a également révélé un gradient de différences au sein de chaque patient.

Conclusion

Ce nouveau concept de code-barres spectral issu de l'analyse d'images IR, apparaît comme une approche intéressante qui pourraît être automatisée pour le diagnostic rapide des tumeurs.

Title Page:

Infrared imaging as a cancer diagnostic tool: introducing a new concept of spectral barcodes for identifying molecular changes in colon cancers

Authors' names and institutional affiliation:

Jayakrupakar Nallala[1], Olivier Piot[1], Marie-Danièle Diebold[1, 2], Cyril Gobinet[1], Olivier Bouché[1, 3], Michel Manfait[1], Ganesh Dhruvananda Sockalingum[1*]

1. MéDIAN Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, FRE CNRS 3481 MEDyC, UFR de Pharmacie, SFR Cap Santé, 51 rue Cognacq-Jay, 51096 Reims cedex, France.

2. Laboratoire d'Anatomie et Cytologie Pathologiques, CHU Robert Debré, Avenue du Général Koenig, 51092 Reims Cedex, France.

3. Service d'Hépato Gastroentérologie et de Cancérologie Digestive, CHU Reims, Avenue du Général Koenig, 51092 Reims Cedex, France.

Tel: +33 32 69 18 12 8, Fax: +33 32 69 13 55 0

Running headline: Infrared spectral barcodes for colon cancer

Abstract:

At present, histopathology is the gold standard method for diagnosis of cancers which is based on the identification of morphological alterations in symptomatic tissues. Novel approaches capable of providing biomolecular information in complement to conventional histopathology are under scrutiny. In this perspective, a biophotonic approach based on infrared spectral micro-imaging combined with multivariate statistical analysis has been implemented on colon tissues. The ability of infrared imaging to investigate the intrinsic biochemical features of cells and tissues has been exploited to develop a new concept of spectral bar-coding. In order to implement this concept, ten frozen colon tissue samples (five normal and tumoral pairs from five patients) were imaged using IR spectral micro-imaging in a non-destructive manner. The spectral images were processed by a multivariate clustering method to identify the histolopathological organization in a label-free manner. Spectral information from the epithelial components was then automatically recovered on the basis of their intrinsic biochemical composition, and compared using a statistical method (Mann-Whitney *U* test) to construct spectral barcodes specific to each patient. The spectral barcodes representing the discriminant infrared spectral wavenumbers (900-1800 cm^{-1}) enabled characterization of malignancy associated biochemical alterations in mucin, nucleotides, carbohydrates and protein regions. This approach not only allowed identification of common biochemical alterations among all the colon cancer patients, but also revealed a difference gradient within individual patients. This new concept of spectral barcoding gives insight into the potential of infrared spectral micro-imaging as a complementary diagnostic tool to conventional histopathology, for biochemical level understanding of malignancy in colon cancers in an objective and label-free manner.

Key words: Infrared spectral imaging, colon cancer, spectral barcodes

Introduction:

Histopathological identification of tissue alterations is the current evaluation method for diagnosis of cancers (1). Several indicators are taken into account for diagnosing cancers which include architectural disorganization and cytological atypia, deeply stained nuclei, increased nucleus to cytoplasmic ratio, loss of differentiation etc,. This present gold standard histopathological analysis is based upon pre-requisite tissue staining for microscopic visualization (2).

Innovative diagnostic methods that provide indications, complementary to the conventional histopathology, in particular the early biomolecular alterations under malignant conditions are under scrutiny (3). One such candidate method is the infrared (IR) spectral imaging which has the potential to provide, in a non-destructive and label-free manner, a biochemical fingerprint of cells and tissues (4). As such its potentials have been exploited in various IR spectroscopic studies applied to cells and tissues from different organs (5-12). IR imaging provides spectral maps which when processed with appropriate multivariate statistical approaches enables to identify and recover the biomolecular information from the histological structures of tissues in normal and tumoral conditions (13).

In this perspective, IR spectral imaging in combination with multivariate statistical analysis has been applied to colon cancer tissues which are one of the highly incident cancers in terms of both incidence and mortality (14). The acquired IR spectral images were subjected to clustering algorithm that permitted classification of the colon histological organization based on the intrinsic biochemical composition, and to construct color-coded spectral images. In comparison to the conventional hematoxylin and eosin (HE) stained reference histological images, the cluster images permitted to retrieve specific IR spectral signatures representative of the normal and the tumoral epithelial components. Further, statistical tests were performed on these spectral signatures to identify discriminant spectral markers, which constituted the basis for a new concept of spectral barcodes.

The aim of the study was therefore to demonstrate the methodology employed to develop spectral barcodes based on IR spectral markers, which can provide rapid and easy to use information originating from the biochemical alterations of a malignant tissue. Although, several screening methods are available for colon cancers (15-17), the diagnosis is always confirmed by microscopic examination of excised tissues. Therefore, we hypothesize that the spectral barcodes can provide complementary biomolecular level information in a rapid,

objective and label free manner. Characterization of some of the important biomolecular constituents implicated in colon cancers has been achieved. The potential applications of this novel concept of spectral barcodes in histopathological diagnosis are discussed.

Materials and Methods:

Infrared spectral image acquisition: Ten frozen colon tissue samples (5 tumoral and 5 adjacent normal) from 5 patients were obtained from Reims University Hospital with the approval of the Institutional Review Board. The sample details are presented as supplementary information 1.

Supplementary information 1: Sample details

Patient No.	Code	Age	Sex	Sample type	Carcinoma location	TNM classification	Grade	Measured image size (µm)
1	11B02580	72	M	Tumoral	Right colon	T3N0	Moderately differentiated adenocarcinoma, colloidemucous	694x554
				Normal				992x1455
2	11B02743	54	M	Tumoral	Left colon	T3N2	Moderately differentiated adenocarcinoma	1313x1315
				Normal				1206x682
3	11B02759	70	M	Tumoral	Left colon	T4N1M1	Moderately differentiated adenocarcinoma	1510x1144
				Normal				1270x1590
4	11B04401	69	M	Tumoral	Left colon	T3N0	Moderately differentiated adenocarcinoma	1337x1858
				Normal				1662x1031
5	11B04415	80	M	Tumoral	**	T4N2	Less differentiated colon adenocarcinoma, aggressive	1047x1281
				Normal				1658x1156

The methodology employed for IR imaging of colonic tissues is shown in Figure 1. Two 10 micron thick consecutive sections were obtained from these samples. While the first section was HE stained and served as a morphological reference, the adjacent section was directly transferred onto a calcium fluoride (CaF_2) window for IR spectral imaging. Tissue zones of interest were selected by an expert pathologist after analyzing the third, 3 micron thick consecutive tissue section by conventional histopathological analysis. Imaging was performed on the Perkin Elmer Spectrum Spotlight 300 imaging system (Courtaboeuf,

France) equipped with nitrogen-cooled 16-element MCT detector at a pixel size of 6.25 x 6.25 μm^2 and a spectral resolution of 4 cm^{-1}, averaged over 16 scans, in the mid-IR range of 750 to 4000 cm^{-1}. Each pixel element (6.25 x 6.25 μm^2) contained a full spectrum. Throughout the measurements, the imaging system and the sample compartment were continuously purged with dry air. The background spectra acquired prior to image acquisition from the CaF_2 window was subtracted from the dataset automatically. On an average, each IR spectral image consisted of about 39100 spectra out of which 27700 spectra were guarded for analysis while the remaining corresponding to low signal to noise ratio were eliminated from the analysis.

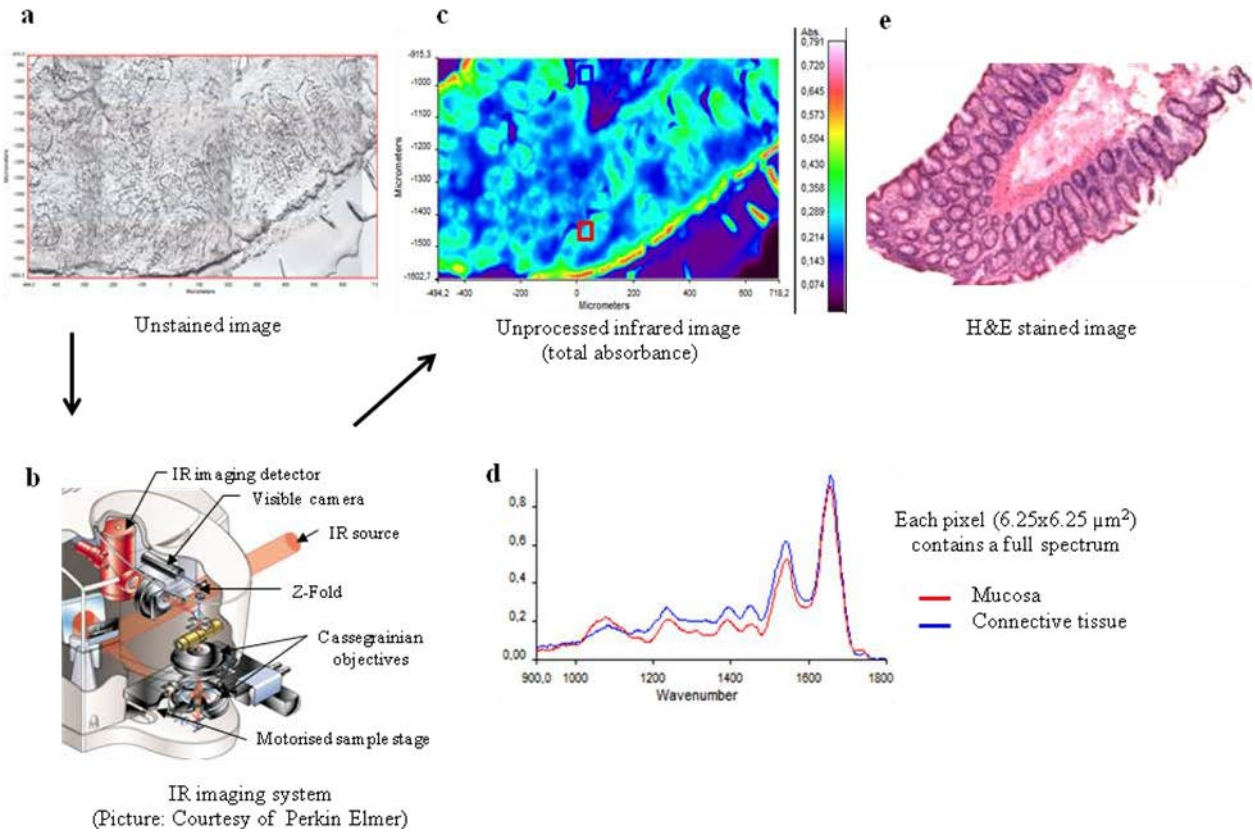


Figure 1: Infrared spectral imaging methodology of colonic tissues.

An unstained tissue section (a) is imaged by an IR imaging system (b) which provides the unprocessed infrared spectral image (c), in which each pixel (6.25µm) corresponds to a full spectrum (d). A conventionally stained (HE) image (e) is used as a morphological reference.

Data pre-processing: The raw IR spectra were initially corrected for atmospheric interferences of water vapour and CO_2 by the built in Perkin Elmer software of the Spectrum Spotlight 300 imaging system. Further pre-processing and processing of the IR spectra were accomplished using the in-house programmes written in Matlab 7.2 (The Mathworks, Natick, MA). A modified Extended Multiplicative Signal Correction (EMSC) method was used for eliminating the IR spectra with low signal to noise ratio (18). All the eliminated spectra were colored as white pixels in the k-means clustered images. Using the same EMSC algorithm, the spectra were also corrected for baseline and finally normalized.

Clustering analysis: With the objective to identify spectrally, different histological classes of the normal and the tumoral colonic tissues, each IR spectral image was partitioned using k-means clustering. This algorithm is an unsupervised clustering method which enables to partition the IR image into predefined number of clusters. Hence, spectra with similar biochemical characteristics group into the same cluster in an iterative manner where each cluster corresponds to a histopathological feature (19). The k-means generated clusters were then annotated into their corresponding histological classes by an expert pathologist using the HE stained images as a reference. The spectral distance between different clusters corresponding to the endogenous biochemical tissue signature of the histological classes was visualized in the form of a dendrogram obtained by hierarchical clustering analysis based on Ward's linkage algorithm. IR spectra from the normal and the tumoral epithelial clusters of the colonic tissue were then extracted for further statistical analysis.

Construction of spectral barcodes using statistical analysis: The Mann-Whitney *U* test was performed to identify the most discriminant IR spectral wavenumbers between the normal and the tumoral conditions. Figure 2 is a schematic representation for the construction of the spectral barcodes. The samples were selected in such a way that each time the statistical comparison was performed to compare IR spectra from the epithelial components the tissue pairs from the same patient to avoid the influence of inter-patient variability. In this way, independent comparisons on five sample pairs from five patients were performed. Three different P- values of significance were then used to consider a gradient of discriminant wavenumbers. The latter permits to constitute the spectral barcodes which reflected the biomolecular changes associated with the compared classes.

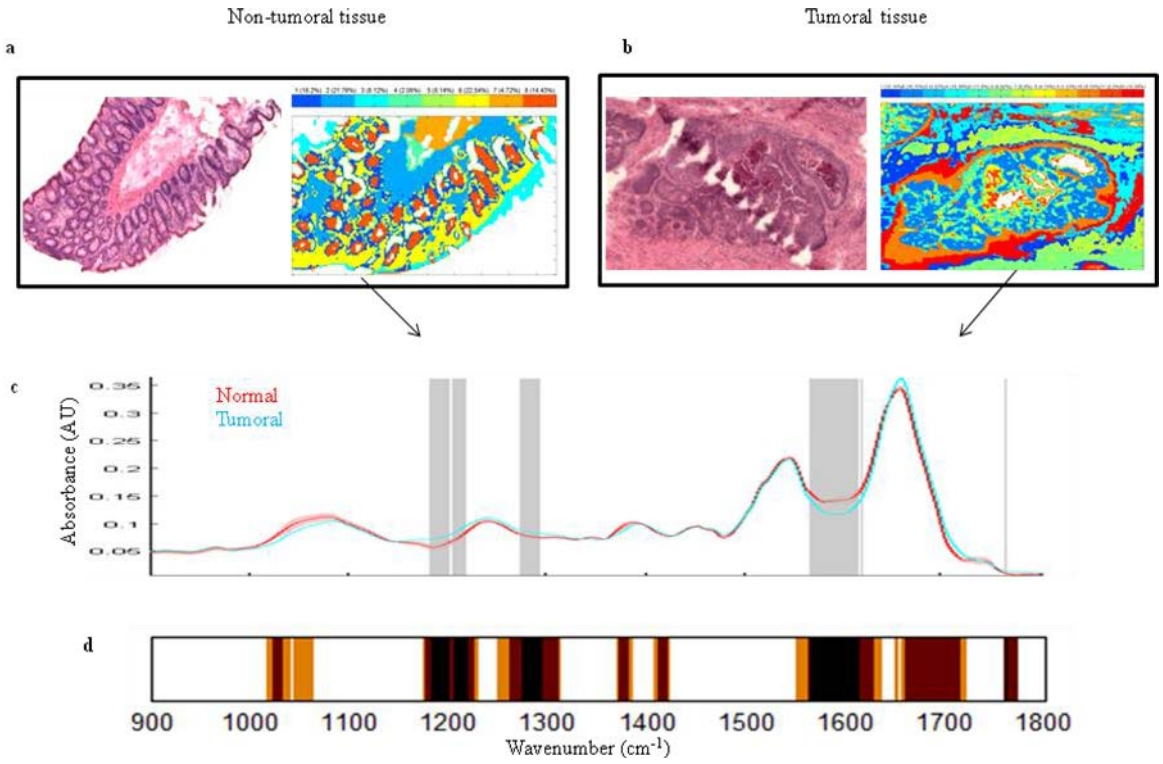


Figure 2: Construction of spectral barcode.

Infrared spectra corresponding to the normal and tumor epithelia (from a and b) are retrieved from the k-means cluster images and compared using a statistical test using different p-values of significance, to find out the significant discriminant wavenumbers (c). These wavenumbers sorted out in a gradient constitutes the spectral barcode (d) representing different biomolecular features.

Results:

Cluster analysis: Cluster analysis by k-means permitted to construct color-coded spectral images. Figure 3 shows representative IR spectral images after the unsupervised classification of a normal and a tumoral colonic tissue using 8 clusters and 12 clusters respectively. These cluster numbers permitted to recover the main histological features of the colonic tissue. Figure 3A displays the cluster analysis in comparison with the HE stained images highlighting a clear identification of the normal colonic features such as the normal epithelium (clusters 1 and 8) comprising respectively the outer and the inner parts of the

crypts, the connective tissue: the lamina propria (cluster 6) and the submucosa (clusters 4 and 7), the muscularis mucosa (cluster 2) and the secreted mucus (cluster 3). Cluster 5 appeared to be associated with the outer parts of the crypts. In contrast, for the malignant tissue the normal histological aspects were no longer discernible. As shown in figure 3B, the only aspects identifiable were the malignant epithelial component itself (clusters 2 and 5) and the associated stromal tissue (cluster 10).

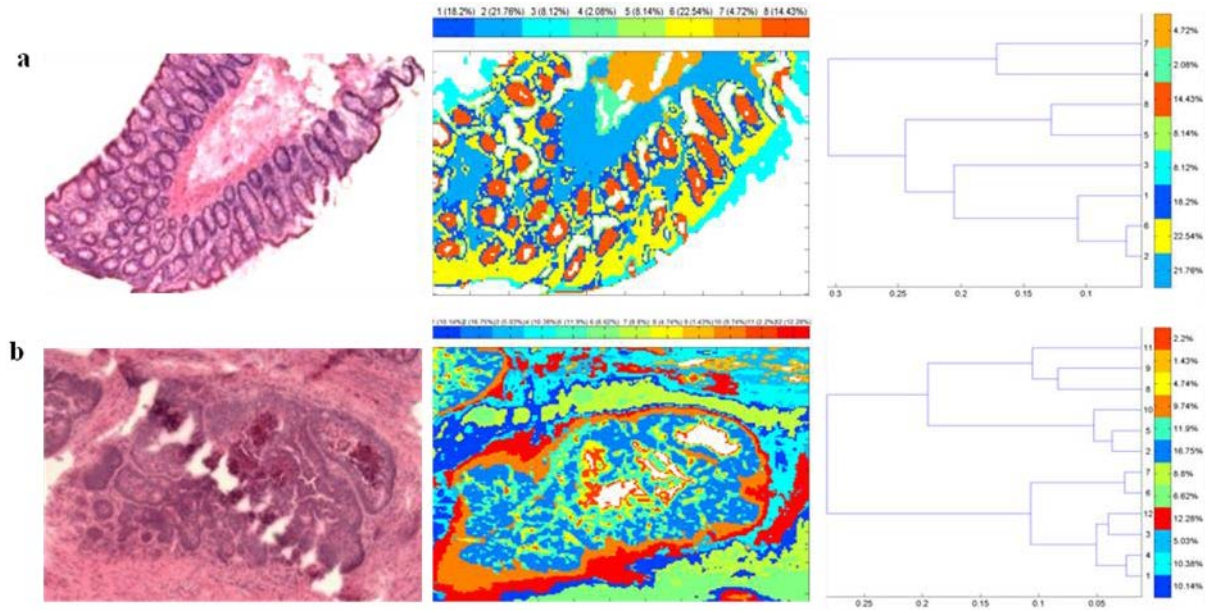


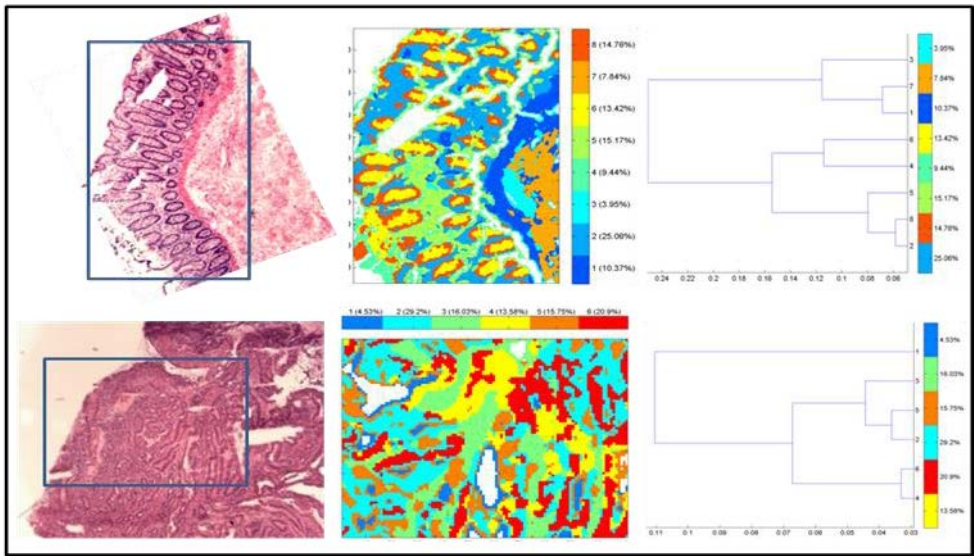
Figure 3: K-means clustering of normal and tumoral colonic FTIR spectral images (middle panel) with the respective dendrograms (right panel) compared to the HE stained sections (left panel), (Sample 2 of SI 1).

Normal colonic tissue section (a) clustered using 8 clusters representing the major normal colonic tissue features by random pseudo-colors. The representation is as follows: Clusters 1 and 8 - normal epithelium (central and peripheral parts of the crypts), cluster 2 - muscularis mucosa, cluster 3 - secreted mucus, clusters 4 and 7 - submucosa, cluster 6 - lamina propria, and cluster 5 - associated with the peripheral parts of the crypts. A moderately differentiated adenocarcinoma of a colon tissue section (b) classified using 12 clusters representing the major tumoral tissue features by random pseudo-colors. The representation is as follows: Clusters 2 and 5 - tumor epithelial component, and cluster 10 - stromal tissue. Remaining clusters are not attributed to any histological class. The HE images are at 5X magnification.

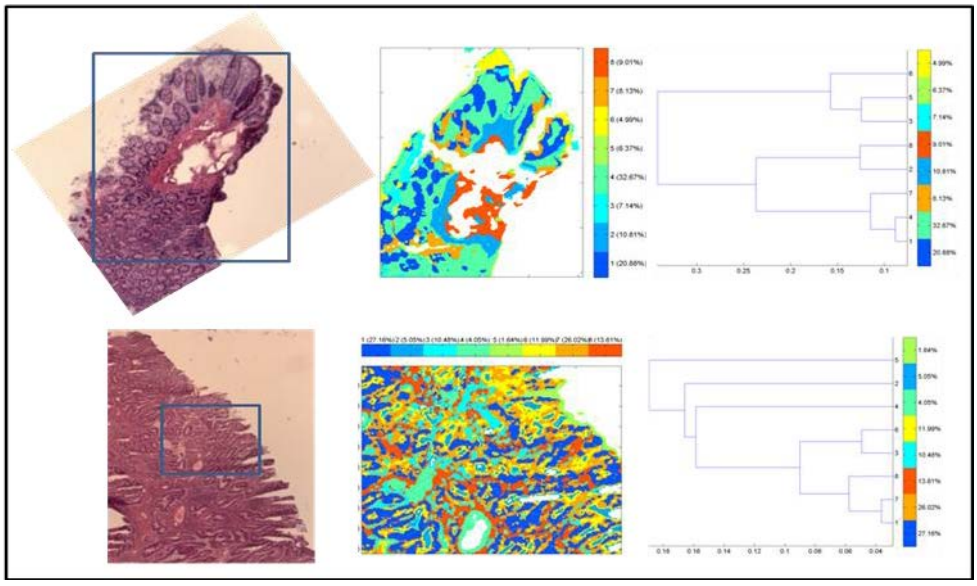
The remaining clusters seemed to be associated with the surrounding stromal tissue. The corresponding dendrograms show the spectral distance between different histological classes. For the tumoral sample, the spectral proximity between the tumor and the stroma (clusters 2 and 5, and 10) appeared clearly.

The k-means clustering enabled identification of the important histological classes of the colonic tissue, and permitted easy retrieval of spectral signatures corresponding to different histological classes for further analysis. The clustering results of the other samples used in the study are shown in the supplementary information 2.

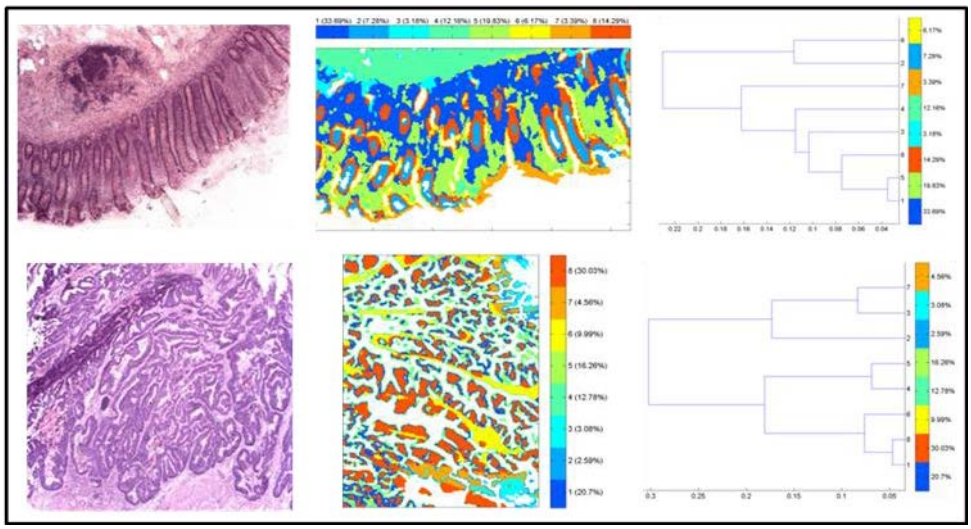
(1)



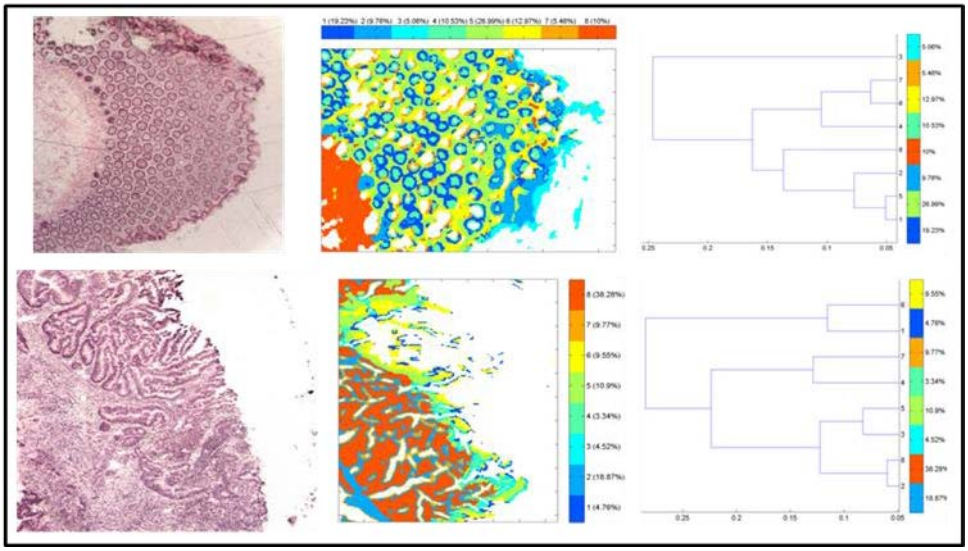
(3)



(4)



(5)



Supplementary information 2: K-means clustering results of all the tissue section included in the study (Sample number 1, 3, 4, and 5 of SI 1 respectively).

K-means clustering of the sample pairs of normal and tumoral colonic FTIR spectral images (middle panel) with the respective dendrograms (right panel) compared to the HE stained sections (left panel). The normal and the tumoral epithelium from each of the sample pair was used for constructing the spectral barcodes. The HE images are at 5X magnification.

Spectral barcodes for biochemical information: Spectra from the normal and the tumoral tissues were compared to identify the most discriminant wavenumbers in the IR spectral range of 900-1800 cm^{-1}, using the Mann-Whitney *U* test. These were arranged in the form of spectral barcodes for the colonic normal and the tumoral epithelia as presented in figure 4. The discriminant wavenumbers were sorted in a gradient of specific color code, using three different levels of statistical significance tests. The color code, black indicated the most discriminant wavenumbers (p=0.00016); orange indicated the less discriminant (p=0.01); and brown the intermediate (p=0.001). The analyzed IR spectral region of 900-1800 cm^{-1}, was divided into three zones as follows: 900-1300 cm^{-1} (zone 1), 1300-1500 cm^{-1} (zone 2), and 1500-1800 cm^{-1} (zone 3).

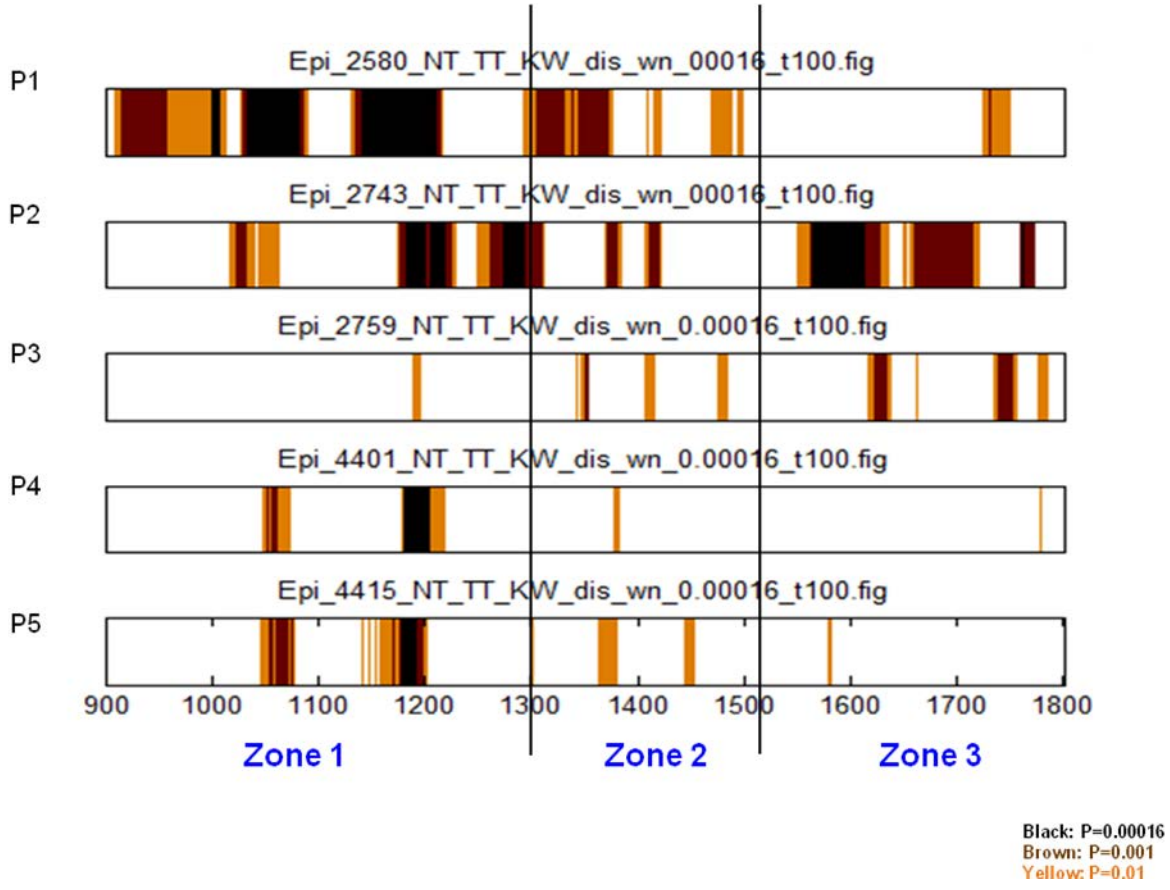


Figure 4: Infrared spectral barcodes constructed for five sample pairs (1-5 of SI 1).

Spectral barcodes are constructed by independently comparing the normal and the tumoral epithelia of five different patients in the spectral region of 900-1800 cm^{-1} which was further divided into three zones. The discriminant wavenumbers are color-coded based on the significance level (p<0.00016-black, p<0.001-brown, p<0.01-orange).

As an example, for patient 1 in figure 4, the most discriminant wavenumbers (black) were observed in zone 1. The intermediate discriminant wavenumbers were mostly present in zone 1 and 2, and the less discriminant wavenumbers were present throughout all the three zones. From these observations, a correlation between the occurrence of the discriminant wavenumbers and the biomolecular alterations in cancerous condition in the respective spectral zones was sought. The IR spectral zone 1 (900-1300 cm^{-1}) representing the highest discrimination wavenumbers (black) was correlated to the alterations involving some of the important biomolecules such as nucleotides, mucin, and carbohydrates as represented in table 1. In parallel, for patient 2, the black zones were predominantly observed in zone 1 (900-1300 cm^{-1}) and zone 3 (1500-1800 cm^{-1}). While the zone 1 was correlated to the biomolecules such as nucleotides, mucin, and carbohydrates, zone 3 was correlated to the protein alterations involving the amide I and the amide II vibrational bands (table 1). For ensemble of the samples the most discriminant wavenumbers (black) together with the intermediate discriminant wavenumbers (brown) were mostly present in zone 1 (4 out of 5 samples) followed by in zone 3 (1 out of 5 samples). Zone 2 included some of the intermediate to less discriminant wavenumbers, and the less discriminant wavenumbers were present throughout all the three zones.

Table 1: Infrared spectral peak attribution

Peak position (cm^{-1})	Biomolecular attribution
1080	PO_2^- symmetric stretch of nucleic acids
1240	PO_2^- asymmetric stretch of nucleic acids
1036	Mucin
1072	
1122	
1314	
1155	C-O stretch of Carbohydrates
1162	H-bonded C-O stretch of Proteins
1176	non-H-bonded C-O stretch of Proteins
1212	Collagen[33]
1280	
1654	Amide I of Proteins
1526 and 1534	Amide II of Proteins
1554 - 1568	
1724 - 1756	C=O stretch of Phospholipids

Discussion:

IR micro-spectroscopy and imaging has demonstrated its potentials in several studies for characterizing cells and tissues for diagnostic purposes. In this study, IR spectral imaging combined with multivariate statistical analysis was applied to colonic tissue in order to identify the spectral markers representing the biochemical changes associated with malignancy, and this forms the basis of a novel concept of spectral barcodes.

To implement this concept, IR spectral imaging of non-tumoral and tumoral sample pairs from different patients was performed independently. Cluster analysis using k-means in comparison to the HE stained images permitted to identify the normal and the tumoral epithelial components of the colonic tissue. K-means clustering has been implemented in several studies for its rapidity and large data classification (19). Using this, the organizational levels of the colonic tissue such as the crypts, the lamina propria, and the submucosa were identified. The normal colonic tissues were characterized by well-differentiated glands in the form of crypts that constituted the mucin filled inner region, and the nuclei rich outer region. In contrast, the tumoral samples were devoid of this organization due to loss of differentiation of the glands. Here the dual parts of the crypts were no longer recognizable and the tumoral tissue was characterized by only two spectral zones corresponding to the epithelial component and the adjacent stroma.

In order to identify statistically the spectral and hence the biochemical differences, comparison of IR spectra corresponding to the normal epithelium and the epithelial malignant component (adenocarcinoma) was undertaken. Results were visualized in the form of spectral barcodes.

The biochemical differences appeared to show a strong association with zone 1 followed by zone 3. Some earlier IR studies on colorectal cancers have associated biochemical alterations with the spectral regions of 900-1300 cm^{-1} (20, 21). In accordance, the color code indicated similar strong association of the most discriminant wavenumbers ($p=0.00016$) in this IR spectral zone that reveal some of the important biomolecules implicated in colon cancers such as nucleotides, mucin, and carbohydrates. Other regions such as the amide regions of proteins are also known to undergo alterations during carcinogenesis (22) thereby changing the overall over all spectral profiles (23). In our study, these biochemical differences originating from the amide regions also appeared discriminant in these cancers ($p=0.001$).

Besides the easy-to-interpret representation of the spectral barcodes, the employment of a specific color coding further enabled to look at the gradient of differences among the biochemical alterations across all the patients included in this study. Furthermore, biochemical alterations occurring in a single patient were also represented that allow extracting patient specific information.

This study is a proof-of-concept for the introduction of IR spectral barcodes carried out on a small sample population, and at this point, we foresee potential applications of this concept. In this study we compared only the normal and the tumoral epithelial components that are the important regions where the cancerous signatures develop. However, barcodes for several other tumor associated features such as tumor and stroma, tumor and inflammation, adenoma and adenocarcinoma, etc., can be developed and deciphered, which can throw light into the biochemical changes associated with each of these tissue states. Also, since tumors are organ specifically heterogeneous, spectral barcodes can be constructed to a particular cancer type constituting a fingerprint of the tumor.

Based on the fact that the spectral barcodes allows visualizing intra-patient biochemical variability, constructing spectral barcodes for the primary tumors with the normal counterpart, and the secondary tumors with the same normal reference, can explore the connectivity of IR spectra between the primary and the metastatic tumor thereby potentially provide information on the metastatic properties of the tumor. Finally, one of the important features of this concept is the possibility to digitalize the spectral barcodes, which can be used to archive data from various tissues sources and also can be used as an identity-card for their identification and retrieval.

However, there remain certain important challenges which need to be addressed, before this concept can be used to its full potential. Although the inter-patient variability is overcome by comparing each time the normal and the tumoral tissues from the same patient, the inherent heterogeneity of different patients and different tumors according to their genotypes, localization, etc can introduce differences in the spectral profiles. This might be the reason as to why in this study, no common spectral profile could be found to be the most discriminant among the studied samples with the same p value of significance. Hence, large scale studies need to be undertaken in order to archive spectral signatures characterizing tissues, and validated in order to construct a larger database using a selected patient population based on the different properties of malignancy in terms of evolution stage, histological and clinical

aggressiveness, and genotypes. We envisage that the IR imaging based spectral barcodes could provide valuable molecular information complementary to histopathology for diagnostic purposes.

The authors declare no conflict of interest.

Acknowledgments

This study was supported by a grant of Institut National du Cancer (INCa) and Canceropôle Grand Est. We would like to thank Ligue contre le Cancer, Conférence de Coordination Interrégionale du Grand-Est, and CNRS Projets Exploratoires Pluridisciplinaires, for financial support. Plateforme IBiSA "Imagerie Cellulaire et Tissulaire", and the Tumorotheque, Champagne-Ardenne is also acknowledged. NJ is a recipient of doctoral fellowship from the Région Champagne-Ardenne.

References:

1. Kendall C, Isabelle M, Hegemark FB, Hutchings J, Orr L, et al. Vibrational spectroscopy: a clinical tool for cancer diagnostics. Analyst, 2009, 134, 1029–1045

2. Thomas D. Wang, George Triadafilopoulos, James M. Crawford, Lisa R. Dixon, Tarun Bhandari, Peyman Sahbaie, Shai Friedland, Roy Soetikno, and Christopher H. Contag. Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. 15864–15869_PNAS_October 2, 2007_vol. 104_no. 40

3. Kendall C, Stone N, Shepherd N, Geboes K, Warren B, Bennett R, et al. Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus.J Pathol 2003; 200: 602–609.

4. Martin FL, Kelly JG, Llabjani V, Hirsch PM, Patel H, Trevisan J, et al. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. Nat Protoc 2010; 5(11), 1748-1760.

5. Pijanka JK, Kumar D, Dale T, Yousef I, Parkes G, Untereiner V, et al. Vibrational spectroscopy differentiates between multipotent and pluripotent stem cells. Analyst. 2010 Dec;135(12):3126-32.

6. Schubert JM, Bird B, Papamarkakis K, Miljkovic M, Bedrossian K, et al. Spectral cytopathology of cervical samples: detecting cellular abnormalities in cytologically normal cells. Laboratory Investigation 2010; 90, 1068–1077.

7. Pijanka JK, Kohler A, Yang Y, Dumas P, Chio-Srichan S, Manfait M, et al. Analyst 2009; Jun;134(6):1176-81.
Spectroscopic signatures of single, isolated cancer cell nuclei using synchrotron infrared microscopy.Epub 2009 Mar 11.

8. Travo A, Piot O, Wolthuis R, Gobinet C, Manfait M, Bara J, et al. IR spectral imaging of secreted mucus: a promising new tool for the histopathological recognition of human colonic adenocarcinomas. Histopathology 2010; 56(7), 921-931.

9. Sebiskveradze D, Vrabie V, Gobinet C, Durlach A, Bernard P, Ly E, et al. Automation of an algorithm based on fuzzy clustering for analyzing tumoral heterogeneity in human skin carcinoma tissue sections. Lab Invest 2011; 91(5), 799-811.

10. Beljebbar A, Amharref N, Le´ ve` ques A, Dukic S, Venteo L, Schneider L, et al. Modeling and Quantifying Biochemical Changes in C6 Tumor Gliomas by Fourier Transform Infrared Imaging. Anal. Chem. 2008; 80, 8406–8415.

11. Kwak JT, Hewitt SM, Sinha S, Bhargava R, et al. Multimodal microscopy for automated histologic analysis of prostate cancer. BMC Cancer 2011, 11:62.

12. Fernandez DC, Bhargava R, Hewitt SM, Levin IW. Infrared spectroscopic imaging for histopathologic recognition. Nat Biotechnol 2005; 23(4), 469-474.

13. Lasch P, Haensch W, Lewis E, Kidder L, Naumann D. Characterization of Colorectal Adenocarcinoma Sections by Spatially Resolved FT-IR Microspectroscopy. Appl Spectrosc 2002; 56(1).

14. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010; 127(12), 2893-2917.

15. Miyoshi H, Oka M, Sugi K, Saitoh O, Katsu K, Uchida K. Accuracy of detection of colorectal neoplasia using an immunochemical occult blood test in symptomatic referred patients: comparison of retrospective and prospective studies. Intern Med 2000; 39(9), 701-706.

16. Rex DK. Colon tumors and colonoscopy. Endoscopy 2000; 32(11), 874-883

17. Zuber TJ. Flexible sigmoidoscopy. Am Fam Physician 2001; 63(7), 1375-1380, 1383-1378.

18. Kohler A, Kirschner C, Oust A, Martens H. Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin. Appl Spectrosc 2005; 59(6), 707-716.

19. Lasch P, Haensch W, Naumann D, Diem M. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. Biochim Biophys Acta 2001; 1688(2), 176-186.

20. Sahu RK, Argov S, Walfisch S, Bogomolny E, Moreh R, Mordechai S. Prediction potential of IR-micro spectroscopy for colon cancer relapse. Analyst, 2010; 135, 538–544.

21. Katukuri VK, Hargrove J, Miller SJ, Rahal K, Kao JY, Wolters R, et al. Detection of colonic inflammation with Fourier transform infrared spectroscopy using a flexible silver halide fiber. Biomedical Optics Express 2010; Vol. 1, No. 3, 1014.

22. Conti C, Giorgini E, Rubini C, Sabbatini S, Tosi G, Anastassopoulou J, et al. FT-IR microimaging spectroscopy: A comparison between healthy and neoplastic human colon tissues. Journal of Molecular Structure 2008; 881(46-51).

23. Rigas B, Morgello S, Goldman IS, Wong PT. Human colorectal cancers display abnormal Fourier-transform infrared spectra. Proc Natl Acad Sci U S A, 1990; 87(20), 8140-8144.

III.4: Article 2

Infrared spectral imaging as a novel approach for histopathological recognition in colon cancer diagnosis

(Submitted to the Journal of Biomedical Optics,

Major revisions were done and the final decision is awaited)

Préambule à l'article 2

Contexte

A l'heure actuelle, le développement de méthodes innovantes de diagnostic semble nécessaire afin de complémenter les méthodes existantes d'histopathologie conventionnelle pour le diagnostic du cancer.

Objectif

Dans cette perspective, nous proposons un nouveau concept basé sur l'histopathologie spectrale, par micro-imagerie infrarouge, directement appliqué sur des tissu arrays paraffinés de côlon stabilisé dans une matrice d'agarose sans déparaffinage chimique.

Matériels et Méthodes

Afin de corriger les interférences spectrales de la paraffine et de l'agarose, un prétraitement mathématique a été mis en œuvre. Les images spectrales corrigées (N=6) ont ensuite été traitées par une méthode de classification par analyse multivariée afin de récupérer automatiquement, sur la base de la composition moléculaire intrinsèque des tissus, les principales classes histologiques composant les tissus normaux et tumoraux du côlon. Les signatures spectrales des différentes classes histologiques des tissus du côlon ont été analysés en utilisant des méthodes statistiques (test du Mann-Whitney *U* et Analyse en Composantes Principales) afin identifier les caractéristiques discriminantes.

Résultats

Ces informations discriminantes ont permis de mettre en évidences certaines modifications biomoléculaires associées à la malignité. Ainsi, par l'intermédiaire d'une seule analyse, sans marquage et de manière non-destructive, les principaux changements liés aux nucléotides et aux glucides, ainsi que les caractéristiques du collagène, ont pu être identifiés simultanément en comparant les tissus normaux avec les tissus cancéreux.

Conclusion

Cette étude démontre clairement le prévue de concept de l'imagerie spectrale IR comme outil moderne et complémentaire à l'histopathologie conventionnelle, pour un diagnostic de cancer objectif directement à partir de tissu arrays paraffinés. Elle établit ainsi les bases du concept d'histopathologie spectrale.

Title:

Infrared spectral imaging as a novel approach for histopathological recognition in colon cancer diagnosis

Authors' names and institutional affiliation:

Jayakrupakar Nallala[1], Cyril Gobinet[1], Marie-Danièle Diebold[1, 2], Valérie Untereiner[1], Olivier Bouché[1, 3], Michel Manfait[1], Ganesh Dhruvananda Sockalingum[1], Olivier Piot[1*]

1. MéDIAN Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, FRE CNRS 3481 MEDyC, UFR de Pharmacie, SFR Cap Santé, 51 rue Cognacq-Jay, 51096 Reims cedex, France.

2. Laboratoire d'Anatomie et Cytologie Pathologiques, CHU Robert Debré, Avenue du Général Koenig, 51092 Reims Cedex, France.

3. Service d'Hépato Gastroentérologie et de Cancérologie Digestive, CHU Reims, Avenue du Général Koenig, 51092 Reims Cedex, France.

Tel: +33 32 69 18 12 8, Fax: +33 32 69 13 55 0

Abstract:

Innovative diagnostic methods are the need of the hour that could complement conventional histopathology for cancer diagnosis. In this perspective, we propose a new concept based on spectral histopathology, using IR spectral micro-imaging, directly applied to paraffinized colon tissue array stabilized in an agarose matrix without any chemical pre-treatment. In order to correct spectral interferences from paraffin and agarose, a mathematical procedure was implemented. The corrected spectral images were then processed by a multivariate clustering method to automatically recover, on the basis of their intrinsic molecular composition, the main histological classes of the normal and the tumoral colon tissue. The spectral signatures from different histological classes of the colonic tissues were analyzed using statistical methods (Mann-Whitney *U* test and Principal Component Analysis) to identify the most discriminant IR features. These features allowed characterizing some of the biomolecular alterations associated with malignancy. Thus, via a single analysis, in a label-free and non-destructive manner, main changes associated with nucleotide, carbohydrates and collagen features could be identified simultaneously between the compared normal and the cancerous tissues. The present study demonstrates the potential of IR spectral imaging as a complementary modern tool, to conventional histopathology, for an objective cancer diagnosis directly from paraffin-embedded tissue arrays.

Key words: Infrared spectral imaging, colon cancer, paraffinized tissue arrays, spectral histopathology

Introduction:

Over the last decade, several biophotonic approaches have been undertaken in view of developing innovative diagnostic methods to complement conventional histopathology. These techniques are foreseen as non-destructive helping tools for pathologists in their routine clinical practice. Among these, infrared (IR) spectroscopy is regarded as one of the candidate methods that could be of valuable interest for cancer diagnosis. This technique allows acquiring spectra from IR active biomolecules present in cells and tissues, whose chemical bonds undergo changes in their electric dipole moment during vibrations thus providing a highly specific "vibrational fingerprint" [(1)]. The spectral information obtained in label-free and non-destructive manner offers insights into the presence of these biomolecules, as well as into their structural and metabolic changes, occurring on the onset and during the course of the disease.[(2)] Combined with a micro-imaging device, IR spectroscopy can rapidly give spatially-resolved biochemical information of different tissue structures, where each pixel of an IR image provides a complete spectrum.[(3)] Via this modality, several studies have exploited IR spectroscopy as a helpful tool with a potential diagnostic value in various cancers like, but not limited to, skin,[(4)] breast,[(5)] cervix,[(6)] colon,[(7)] prostate,[(8, 9)] lung,[(10)] esophagus,[(11)] thyroid,[(12)] brain.[(13)] These IR studies were performed on tissues that were either fresh;[(11, 12)] frozen,[(5, 10, 13)] or formalin-fixed paraffin-embedded (FFPE).[(6, 8)] Until recently, IR studies of FFPE tissues necessitated chemical dewaxing prior to image acquisition because of the strong contribution of IR absorption peaks of paraffin, which interfere with the biochemical information originating from the tissue. However, this procedure is time- and reagent- consuming, and has been shown to result in an incomplete deparaffinization.[(14)] An alternative way to circumvent chemical dewaxing is to perform a numerical deparaffinization directly on the IR spectral image. Thus, for the first time, the feasibility of IR imaging combined with numerical deparaffinization to paraffinized colon tissue arrays that are stabilized in an agarose matrix, without any chemical deparaffinization, was undertaken. In addition to paraffin, the agarose matrix also contributes to the confounding spectral interferences. Therefore, an algorithm based on Extended Multiplicative Signal Correction (EMSC) was implemented to neutralize these spectral interferences from paraffin and agarose. The processed IR images were then analyzed with a clustering method to identify and classify the constituent tissue structures based on their intrinsic molecular composition. This statistical approach permitted to construct color-coded images that were then compared with conventional histology for morphological recognition. From this

procedure, identification of characteristic spectral signatures representing the biomolecular changes, useful for differentiating between normal and tumoral conditions, and tumor and tumor-associated stroma, was also undertaken. To demonstrate the proof-of-concept of spectral histopathology, we selected one of the highly incident cancers namely the colorectal cancer, that has an incidence of 1.2 million cases and 608,000 deaths worldwide in 2008.[15] Although fecal occult blood test (FOBT),[16] colonoscopy,[17] and sigmoidoscopy[18] are used for colorectal cancer screening and detection, presently the diagnosis is settled upon microscopic examination which remains the gold standard for cancer diagnosis. Nevertheless, the staining and morphological analyses do not allow interpretation of the molecular changes occurring within the cancerous tissue at that particular time. In such scenario, IR imaging could be a valuable complementary tool for conventional histopathological cancer tissue examination.

Materials and Methods

Tissue array preparation: Tissue arrays are paraffinized tissue blocks in which chosen tissue cores have been assembled. The tissue array blocks were paraffinized, and stabilized in an agarose matrix to reduce the common problem of tissue loss during sectioning, and were manually prepared in the University pathology laboratory. Each tissue array block consisted of 13 tissue cores of approximately 3 mm in diameter from normal and tumoral colonic tissue. Samples were selected by an expert pathologist using the hematoxylin, phloxine and saffron (HPS) stained image as the reference. In this study, IR imaging analysis has been implemented on six samples (three normal and three tumoral) of the colon tissue array obtained from three different patients. From each patient, a sample pair of normal and tumoral tissues was obtained to avoid inter-patient variability, in order to optimize this novel methodology. All the tumoral samples corresponded to moderately differentiated adenocarcinoma and the normal samples from the adjacent normal mucosa. This study was approved by the Institutional Review Board of CHU Reims.

Fourier transform infrared (FTIR) image acquisition: The methodology for IR imaging of a tissue array is shown in Figure 1. Three and 10 µm thick adjacent microtome sections were cut from the tissue array block. While the 3 µm section was used by the pathologist for conventional histopathological analysis via HPS staining, the first 10 µm section was used for

IR imaging analysis and the second for additional histopathological comparison. The HPS stained sections were chemically deparaffinized while the adjacent 10 µm paraffinized unstained tissue section was mounted on an IR compatible calcium fluoride (CaF_2) window. This was directly imaged without deparaffinization, by an IR imaging system (Spotlight 300, Perkin Elmer, Courtaboeuf, France) equipped with nitrogen-cooled 16-element MCT detector at a pixel size of 6.25 µm and spectral resolution of 4 cm^{-1}, averaged to 16 scans, in the mid-IR range of 750 to 4000 cm^{-1}. These acquisition parameters provided good quality data with good enough spatial and spectral resolutions for tissue investigation. The instrument and the sample compartment were continuously purged with dry air and parameters like relative humidity and water vapor were kept constant throughout the image acquisition time.

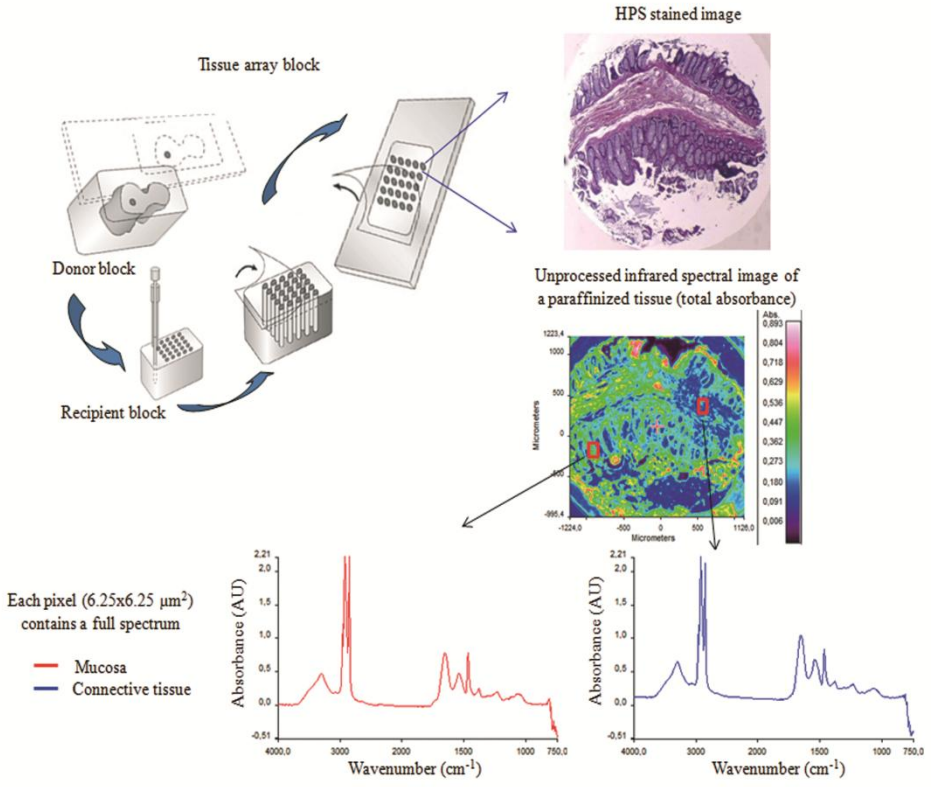


Figure 1: Infrared spectral imaging methodology of colon tissue arrays.
A paraffinized tissue array core is imaged directly by infrared imaging system that constitutes the unprocessed infrared spectral image, which harbors a full spectrum at each pixel size of 6.25 µm, using a conventionally stained image as a morphological reference.

The background spectrum from the CaF_2 window, acquired prior to image acquisition, was subtracted from the dataset automatically. Each tissue array-IR image of one circular spot (3 mm in diameter) consisted of around 130 000 spectra, and each pixel element of 6.25 µm contained a full spectrum.

Pre-processing of IR spectra: The spectra from the IR images included: atmospheric absorptions of water vapor and CO_2, chemical absorptions of paraffin and agarose, and biochemical absorptions from the tissue itself. In order to preserve only the biochemical information, stringent pre-processing steps were employed to neutralize the contributions of non-informative spectra. For this, atmospheric correction was performed to remove contribution from water vapour and CO_2 by the built-in software of Spectrum Image (Perkin Elmer). Further analyses were performed using in-house algorithms written in Matlab 7.2 (The Mathworks, Natick, MA). EMSC was used for correcting paraffin, agarose, and baseline, followed by normalization. Pre-processing, processing and analysis of the IR spectra were carried out on spectral images in the IR absorption range of 900-1800 cm^{-1} considered as the most informative region[19, 20] as far as the tissue features are concerned.

Construction of EMSC model: EMSC was developed initially to correct the spectra from the physical light scattering effects that are different from the chemical light absorbance effects.[21, 22] IR spectra of paraffinized colon tissue array sections, along with the biochemical information originating from the tissue, showed absorption bands of paraffin (1378 cm^{-1} and around 1467 cm^{-1}) and agarose (1072 cm^{-1} and minor peaks at 932 cm^{-1}, 1155 cm^{-1} and 1185 cm^{-1}) in the 900-1800 cm^{-1} spectral region (Figure 2; box 1). For efficient classification and understanding of the biochemical nature of the tissue, the variability of these contributions (paraffin and agarose) had to be reduced and their influence circumvented, for which EMSC algorithm was employed in this novel approach as shown in the form of a flowchart in the Figure 2, box 2. According to our previous study[23] EMSC models linearly each spectrum of the data set as:

$$\mathbf{s_i} = a_i\hat{\mathbf{s}} + \mathbf{b_i}\mathbf{I} + \mathbf{c_i}\mathbf{P} + \mathbf{e_i} \qquad (1), \text{ where,}$$

$\mathbf{s_i} \in \mathbb{R}^{1\times n}$ is the i^{th} acquired spectrum of the data set, i.e., a vector composed of n points,

$\hat{\mathbf{s}} \in \mathbb{R}^{1\times n}$ is the target spectrum that is chosen as the mean spectrum of the studied dataset,

$\mathbf{I} \in \mathbb{R}^{k \times n}$ is the interference matrix composed of k components,

$\mathbf{P} = \begin{pmatrix} \nu_1^0 & \cdots & \nu_1^p \\ \vdots & \ddots & \vdots \\ \nu_n^0 & \cdots & \nu_n^p \end{pmatrix}^T \in \mathbb{R}^{(p+1) \times n}$ is the transpose of the Vandermonde matrix of the n wavenumbers ν_j; this matrix is used to compute $\mathbf{c_i P}$, a p-order polynomial function modeling for the baseline,

$\mathbf{e_i} \in \mathbb{R}^{1 \times n}$ is the model error vector,

a_i is the scalar fitting coefficient of $\hat{\mathbf{s}}$ to $\mathbf{s_i}$,

$\mathbf{b_i} \in \mathbb{R}^{1 \times k}$ is the vector of the fitting coefficients of $\mathbf{I}$ to $\mathbf{s_i}$,

$\mathbf{c_i} \in \mathbb{R}^{1 \times (p+1)}$ is the vector of the fitting coefficients of $\mathbf{P}$ to $\mathbf{s_i}$ and represents the coefficients of the p-order polynomial function.

The coefficients a_i, $\mathbf{b_i}$ and $\mathbf{c_i}$ are estimated by the traditional least squares method in order to minimize the model error $\mathbf{e_i}$. The corrected spectra could be then represented by the equation

$$\mathbf{s}_{\mathbf{i}corr} = \hat{\mathbf{s}} + \frac{\mathbf{e_i}}{a_i} \quad (2)$$

The aim of EMSC is to estimate the model coefficients a_i, $\mathbf{b_i}$ and $\mathbf{c_i}$ in order to minimize the error $\mathbf{e_i}$, knowing $\hat{\mathbf{s}}$, $\mathbf{I}$ and $\mathbf{P}$. EMSC can also be viewed as a fitting of the recorded spectra on the mean spectrum. Thus, the biochemical differences of different pixel spectra are modeled in the error $\mathbf{e_i}$. The interference matrix and the Vandermonde matrix are uniquely used in the EMSC model to adjust the paraffin and agarose signals and baseline of the recorded spectra to the mean spectrum. The EMSC protocol has been used to realize several corrections; firstly, it corrects spectra from paraffin and agarose contributions. Secondly, it corrects spectra for light scattering effects, and thirdly, it normalizes spectra on the mean spectrum $\hat{\mathbf{s}}$. Briefly, in order to achieve these corrections, an IR image consisting of 13516 spectra was acquired from 10 µm thick paraffin (used for tissue embedding in our laboratory) section using the same spectral parameters as that of the TMA images. Principal component

analysis (PCA) was performed on these spectra to model them with orthogonal components best explaining the variability of paraffin. The interference matrix **I** of model (1) was constructed by retaining the first 10 principal components (PCs) and the mean spectrum of paraffin. Another IR image consisting of 15872 spectra was acquired from a 10 µm thick section of a mixture of paraffin and agarose, as agarose is a semisolid matrix (at 2% used for TMA construction) and could not be sectioned alone. The spectra of this image were then modeled using equation (1) in which a fourth order polynomial function is assumed to construct **P** to model baseline. Paraffin contributions were then neutralized from agarose, by application of correction (2). Next, PCA was performed on these paraffin corrected agarose spectra in order to model the IR signal of agarose. The first 10 significant PCs and the mean spectrum of agarose were then added to the interference matrix **I** . **I** is thus composed of 11 components modeling paraffin and 11 components modeling agarose. **I** being constructed and a fourth order polynomial function being still assumed for **P**, the model (1) was applied to the colon IR spectral images acquired from the biopsies. The entire data set was then corrected from the contributions of paraffin and agarose, baseline corrected and normalized on the entire spectral range using equation (2). Furthermore, a thresholding of a_i and $E = \sum_{j=1}^{n} \left(\frac{\mathbf{e_i}(j)}{a_i} \right)^2$ permitted to detect the outlier spectra (spectra with high paraffin and agarose contributions or spectra with a poor tissue contribution) of paraffin and agarose, and to eliminate them from further analysis. In the k-means classified images, the pixels corresponding to these outliers are colored white.

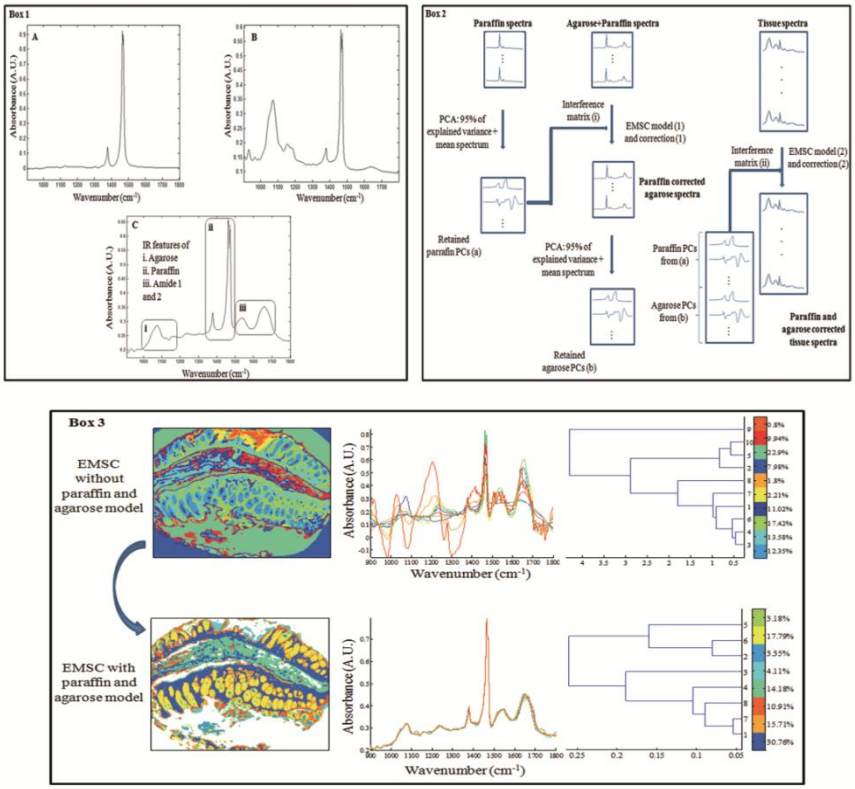


Figure 2: EMSC preprocessing

Box 1: Average IR spectra of paraffin (A), paraffin and agarose together (B) and a paraffinized colon tissue array section (C) which includes spectral information from tissue, paraffin, and agarose, in the spectral range of 900 - 1800 cm^{-1}.

Box 2: Flowchart of the EMSC protocol. Interference matrix 1 constructed from pure paraffin spectra (PCA + mean spectrum) and modeled into EMSC, is employed on paraffin-agarose spectra to neutralize the paraffin influence and retain only the agarose spectra. Interference matrix 2 is constructed from the paraffin corrected agarose spectra (PCA + mean spectrum) and modeled into EMSC. Interference matrices 1 and 2 are then employed on the tissue spectra to neutralize both paraffin and agarose influences and retain only the biochemical information.

Box 3: Comparison of the application of EMSC, with and without paraffin and agarose corrections, by k-means clustering of a FTIR spectral image (left panel). EMSC corrected pixels are colored in white. Corresponding cluster centroids (middle panel) and the dendrogram (right panel) show the differences due to the influence of spectral interferences (paraffin, agarose and other interferences represented by clusters 2, 5, 8 and 9).

Image clustering: The large numbers of IR spectra from each image were classified using an unsupervised k-means clustering method owing to its capability of rapid and huge data classification.[24] This method iteratively partitions the spectra into different classes based on the spectral signatures. First, K spectra (K is the number of searched clusters) are randomly chosen to represent initial centroids which model the mean spectrum of each cluster. Second, each spectrum is affected to the cluster with the nearest centroid according to the Euclidean distance. Third, each centroid is updated as the mean of the spectra belonging to its cluster. Steps 2 and 3 are repeated until the convergence of the algorithm. Therefore, spectra with similar biological characteristics fall into the same cluster and spectra with dissimilar biological characteristics fall into different clusters. In k-means, each spectrum belongs to a unique cluster and can thus be represented by a unique color distinct from those of the remaining clusters and a color coded image can be reconstructed for rapid and simple visual analysis of clustering results. These were then compared to adjacent HPS stained sections to annotate each spectral cluster to the tissue structural feature that it belongs to by an expert pathologist.

Statistical tests: Mann-Whitney *U* test was performed on individual spectra from two clusters and the wavenumbers that were significantly discriminant ($p<0.001$) were retained. These are shown as grey bars in the Figure 4a. In parallel, Principal component analysis (PCA), one of the commonly used spectral data processing method, was applied on the same two clusters (mean-centered data) for validation of the KW observations and better visualization of the spectral separation.

Results

Neutralization of paraffin and agarose contributions using EMSC: Spectral interferences from paraffin and agarose were estimated and corrected on the colonic tissues. Figure 2; box 3 shows a representative k-means cluster image before and after the application of the correction model for paraffin and agarose. In the unprocessed image constructed using 10 clusters, spectra corresponding to these outlier spectra were seen around the tissue array sample spot. Clustering analysis of this unprocessed image showed less accurate correlation with the adjacent HPS stained reference image (Figure 3a) and features such as the colonic epithelium could not be deciphered accurately even when increasing the number of clusters

(data not shown). The cluster centroids showed the contribution of outliers to the image (specifically clusters 2, 5, 8 and 9) which is also reflected in the dendrogram which separates the tissue features from the outliers. In the EMSC corrected image, all the outlier spectra mostly corresponding to the paraffin and agarose contributions are retrieved from the data analysis and are shown as white pixels, which can be found around, and within the clefts of the tissue array sample spot. The resulting high degree of correlation of the FTIR image using 8 clusters to the HPS stained reference image is shown in Figure 3a that demonstrates the importance of neutralizing the spectral interferences.

It has to be noted that although the IR tissue spectra still exhibited the characteristic paraffin and agarose bands (1378 cm^{-1} and around 1467 cm^{-1} for paraffin and, 1072 cm^{-1} and minor peaks at 1155 cm^{-1} and 1185 cm^{-1} for agarose), the influence of their spectral variability is neutralized in the clustering scheme by EMSC. Therefore, the EMSC model does not completely remove the spectral features of paraffin and agarose, but neutralizes them. Thus, in the image analysis by chemometric methods, only the biochemical information is taken into account. The signals from paraffin and agarose are disregarded. Along with the neutralization of intra-sample variability arising from paraffin and agarose contributions, the inter-sample variability is avoided by using a single common target spectrum (the average spectrum on which the spectra are fitted) for all the samples.

IR image clustering: After EMSC correction, k-means clustering was employed to partition the spectra of paraffinized normal and tumoral colonic tissue sections. Figure 3a and b show the corresponding k-means images of these samples classified into eight and fourteen clusters respectively. These cluster numbers permitted to retrieve the principal histological structures, when compared to the HPS stained images. For example, as shown in Figure 3a, it was possible to identify mucosa of the normal colon that comprises of; the lamina propria (cluster 1), the loose connective tissue in which the crypts are organised; crypts (cluster 6 and 7) comprising the central part and the peripheral parts, the functional glands of a colon composed of various epithelial cell populations like goblet cells, absorptive cells, endocrine cells, or stem cells. Mucus (cluster 2) as seen in the crypt lumen and also secreted out of the crypts, submucosa (cluster 4) the fibrous connective tissue usually rich in collagen, and the blood vessels (cluster 8) in the submucosa, were also identified. Finally, clusters 3 and 5 present in minute percentage were not assigned to any specific histological structure and

seem to represent extra mucos structures (appear on the periphery of the mucosa, or tired out mucosa).

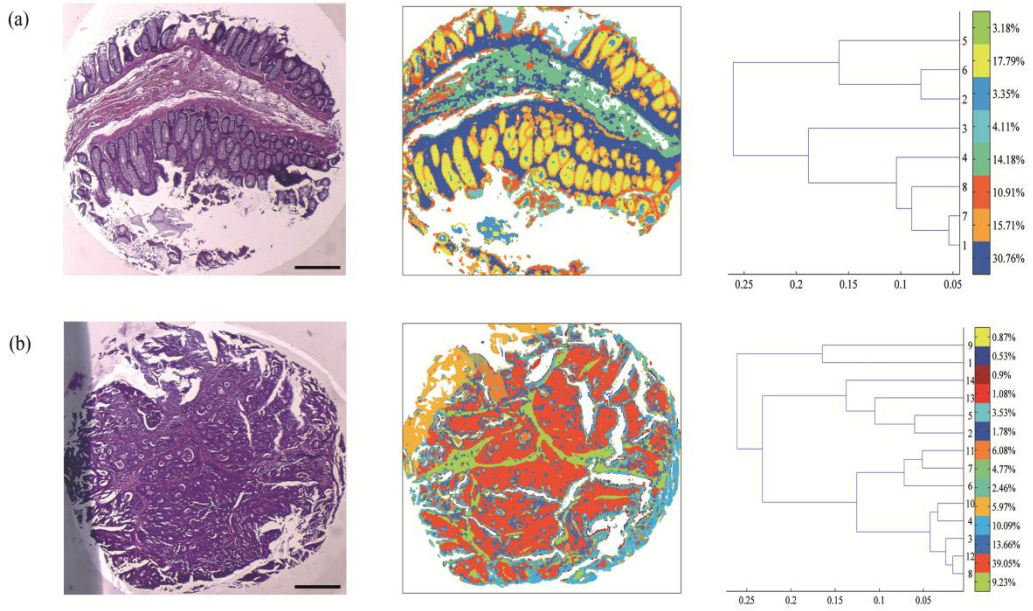
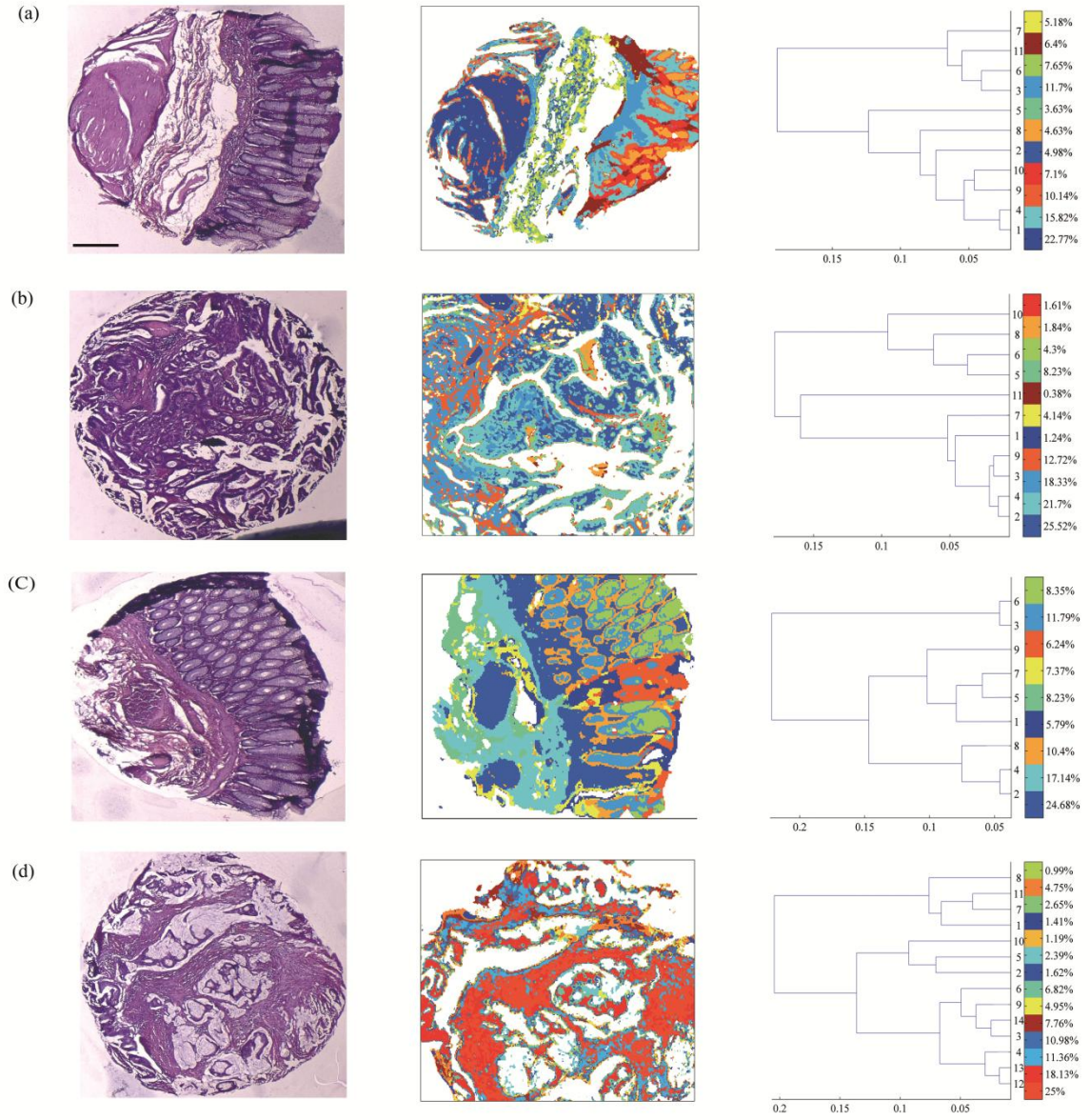


Figure 3: K-means clustering of FTIR spectral images (middle panel) with the respective dendrograms (right panel) compared to the HPS stained colon tissue sections (left panel). Normal colonic tissue section (a) classified using 8 clusters representing the major normal colonic tissue features by random pseudo-colors. The representation is as follows: Cluster 1 - lamina propria, cluster 2 - mucus, cluster 4 - submucosa, clusters 6 and 7- crypts (central and the peripheral parts), cluster 8 - blood vessel and other undefined tissue. Clusters 3 and 5 - extra mucous structures. A moderately differentiated adenocarcinoma of a colon tissue section (b) classified using 14 clusters representing the major tumoral tissue features by random pseudo-colors. The representation is as follows: Cluster 8 represents tumor-associated stroma, and cluster 12 represents tumor epithelial component. Remaining clusters are not attributed to any histological class. Scale bar indicates 500 µm.

The spectral distances between the 8 cluster centroids are computed and shown in the form of a dendrogram (figure 3, right panel). In the case of tumoral tissue, characterization by spectral imaging was illustrated in a sample of moderately differentiated colon adenocarcinoma as shown in figure 3b. K-means clustering using 14 clusters permitted to

highlight two informative clusters: one attributed to the epithelial component (cluster 12) and the other to tumor-associated stroma (cluster 8). The latter, clearly demarcated from tumor, necessitated a minimum of 14 clusters to be segregated out of the tumor. The close spectral signature of the epithelial component to its associated stroma is clearly demonstrated by the corresponding dendrogram. Increasing the number of clusters did not provide any further exploitable information for spectral histology. The k-means clustering results of the other samples used in the study are shown in supplementary figure 1.

From spectral data to biomolecular level information: From the k-means images, it was possible to assign specific spectral signatures to histological structures that were then exploited to gain insight into the biomolecular characteristics of the normal and the tumoral colonic tissues. For this, statistical data processing using the KW test was performed on individual spectra from two clusters of interest each time, to find the spectral differences. Complementarily, PCA was also performed to confirm these differences by considering the two first principal components (PC1 and PC2) that carried the highest explained variance. Figure 4a shows the most discriminating spectral regions identified by the KW test (grey bars) superimposed over the PCA loadings (PC1 and PC2) for the following pair-wise comparisons: normal crypts with adenocarcinoma corresponding to the epithelial components (left panel); adenocarcinoma with the associated stroma, which is the seat of the changes associated with the tumor environment during carcinogenesis and progression (middle panel); and in the normal tissue, lamina propria with submucosa (right panel).



Supplementary figure 1: K-means clustering of FTIR spectral images (middle panel) with the respective dendrograms (right panel) compared to the HPS stained colon tissue sections (left panel). Normal colonic tissue sections (a and c) are clustered using11 and 9 clusters respectively representing the major normal colonic tissue features by random pseudo-colors. The moderately differentiated colon adenocarcinoma tissue sections (b and d) clustered using 11 and 14 clusters respectively representing the major tumoral tissue features by random pseudo-colors. The tumoral tissue (c) is a mucinous tumor. Scale bar indicates 500 µm.

The discriminant wavenumbers identified by KW test corresponded principally to the first PC that was found to be visually the most discriminant in the pair-wise comparisons of right and left panels, and the second PC that was most discriminant in pair-wise comparison of the middle panel, as also represented in the PCA score plots in Figure 4b.

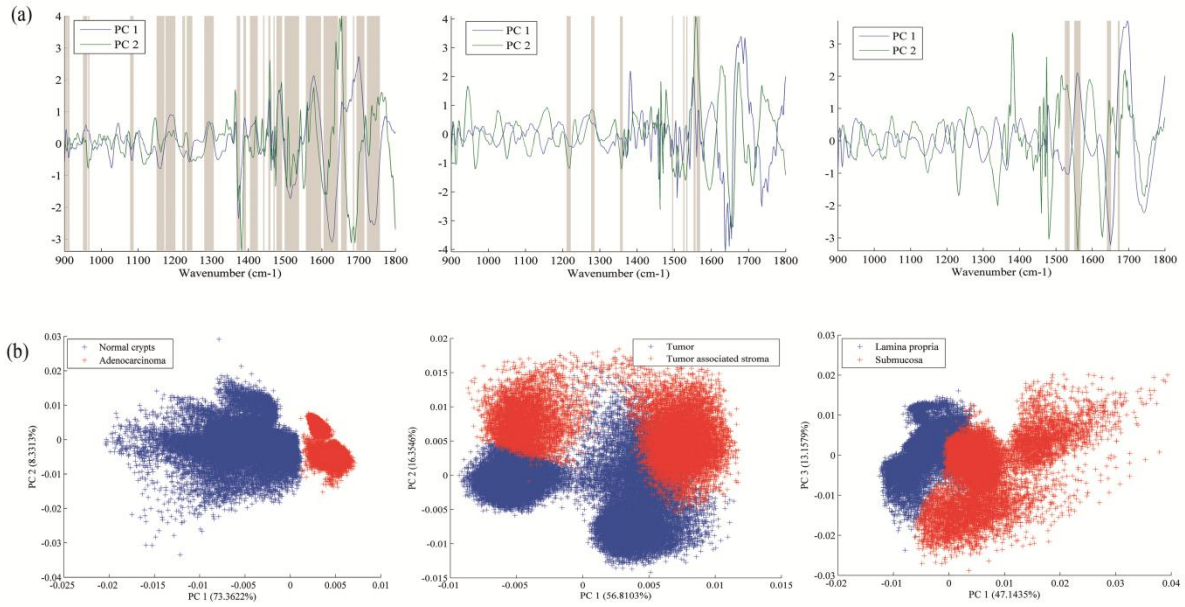


Figure 4: Discrimination of tissue features obtained by the Mann-Whitney *U* test and validated by PCA between the following pair-wise comparisons: normal crypts with adenocarcinoma (left panel); adenocarcinoma with the associated stroma (middle panel); lamina propria with submucosa (right panel). The most discriminant spectral wavenumbers between the compared clusters identified by the Mann-Whitney *U* test ($p<0.001$) are represented as gray bars. They are superimposed by PCA loadings showing the two first PCs with the highest explained variance (a). The PCA score plot showing the separation between the compared clusters (b).

The most clear-cut discrimination as shown in the score plot of Figure 4b, left panel (in the form of separation between the two clouds) was observed between the normal crypts and adenocarcinoma that reflect the overall biochemical alterations in this malignancy. When comparing the adenocarcinoma cluster with its associated stroma (middle panel), or the lamina propria and the submucosa (right panel), the separation is possible but with some spectral overlapping between the clouds. From the wavenumbers identified as discriminant

by the KW test for all comparisons, we have tentatively attempted to correlate some of the IR vibrations to the biomolecular information contained in the colonic tissues as shown in

Table 1: Infrared spectral peak attribution

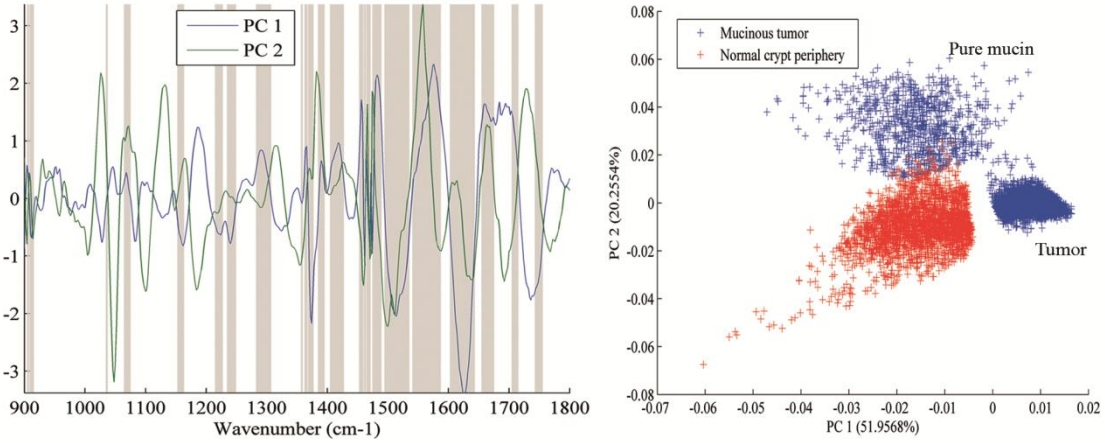
Normal crypts - Adenocarcinoma		Adenocarcinoma - Tumor associated stroma		Lamina propria - Submucosa	
Peak position (cm^{-1})	Biomolecular attribution	Peak position (cm^{-1})	Biomolecular attribution	Peak position (cm^{-1})	Biomolecular attribution
1080	PO_2^- symmetric stretch of nucleic acids[9]	1212	Collagen[33]	1526 - 1536	Amide II of Proteins
1240	PO_2^- asymmetric stretch of nucleic acids[33]	1280		1552 - 1566	
1155	C-O stretch of Carbohydrates[34]	1526 and 1534	Amide II of Proteins	1642 - 1650	Amide I of Proteins
1162	H-bonded C-O stretch of Proteins[34]	1554 - 1568		1672 - 1674	
1176	non-H-bonded C-O stretch of Proteins[34]				
1654	Amide I of Proteins				
1724 - 1756	C=O stretch of Phospholipids[9]				

Other spectral attributions	
1036 1072 1122 1314	Mucin [2, 7, 24, 33]
1378 1467	Paraffin
932 1072 1155 1185	Agarose

At the same time, the PC scores and loadings were also exploited to interpret the differences of spectral intensities between the compared classes. As an example, in the case where the first PC is the most discriminant (figure 4b, left panel), the spectra are mathematically approximated by the first PC loading weighted by the first PC score. Thus, the representative peak at 1658 cm^{-1} (amide I region) of the first PC loading and the first PC scores of adenocarcinoma being positive, their product is positive and hence correspond to higher spectral intensity. On the contrary, the PC scores of the normal crypts being negative their product with the positive peak at 1658 cm^{-1} of the first PC loading, is negative, and hence represents a decrease of spectral intensity. This spectral difference attribution becomes more complex when there is more than one discriminant PC (figure 4b, right panel) as several PCs can have an opposing contribution to the peak intensity.

For the normal crypts and the adenocarcinoma, the discriminant spectral features were particularly attributed to PO_2^- symmetric and asymmetric stretching vibrations of nucleic acids which exhibited relatively higher intensities in the normal crypts. Other differences included those originating from the phospholipids (C=O stretching vibrations); and those from the carbohydrates (C-O stretching vibrations). These signals were relatively more intense in normal crypts than in adenocarcinoma while the opposite tendency was observed for the amide I band of proteins. The hydrogen bonded C-O groups of proteins in the normal tissue was seen to decrease in the tumoral tissue.

It was examined if the discrimination potential of the methodology between the normal and the tumoral tissues is influenced by the tumor type with respect to certain biomolecules. For this, the spectra originating from the tumor and the secreted mucin clusters, of one of the tumoral samples that was mucinous adenocarcinoma were compared with the spectra from the non-mucinous regions of the normal crypts (crypt periphery) of the same patient. The statistical analysis revealed appearance of mucin peaks (1036 cm^{-1}, 1072 cm^{-1}) as discriminant vibrations that were of high intensity in the tumoral tissue as shown in the Supplementary figure 2.



Supplementary figure 2: Discrimination of tissue features obtained by the Mann-Whitney *U* test and validated by PCA between the pair-wise comparisons of normal crypt periphery with mucinous tumor. The most discriminant spectral wavenumbers between the compared clusters identified by the Mann-Whitney *U* test (p<0.001) are represented as gray bars. They are superimposed by PCA loadings showing the two first PCs with the highest explained variance (left panel). The PCA score plot showing the separation between the compared clusters (right panel).

When comparing adenocarcinoma and tumor-associated stroma, the discriminating spectral features corresponded to collagen features, and amide II of proteins. For lamina propria and submucosa clusters, the amide regions of proteins appeared to contribute to the discriminant wavenumbers.

Discussion

Very few studies have combined IR imaging with tissue microarray (TMA) technology[25, 26] and none have involved direct analysis of the paraffinized tissue arrays or, tissue arrays stabilized in an agarose matrix.[27] This study is a first attempt to apply IR spectral imaging to a paraffinized tissue array stabilized in an agarose matrix, without any chemical deparaffinization, for comparing normal and tumoral colonic tissue samples. EMSC initially developed to correct light scattering effects,[21, 22] and water vapor and carbon dioxide,[28] has also been previously implemented by our group to neutralize paraffin contributions in

paraffinized tissues.[23, 29, 30] In this study, it was employed for the first time, a step ahead to neutralize spectral interferences from both paraffin and agarose, projecting EMSC as a 'custom made correction method' which could be adapted to correct a variety of spectral interferences and permit to test tissues in different embedding materials.

K-means classification of the EMSC corrected IR spectral images allowed identification of various histological structures of the normal and the tumoral colonic tissues. The colonic tissue structures like the lamina propria, the submucosa, the crypts and the blood vessels were easily identified in the normal histological and the spectral images. The spectral signatures associated with the biomolecular differences between these histological groups were highlighted by the KW test and confirmed by PCA analysis. In the normal tissue, k-means clustering differentiated well between the lamina propria and the submucosa, which are both connective tissues. Based on the multivariate statistical analyses, the biomolecular discrimination can be associated to the changes in the spectral profiles of the amide regions of proteins.

Normal crypts are the functional glands of the colonic mucosa, where the molecular transformations in the event of carcinogenesis take place. The k-means cluster image allowed to clearly distinguish both the central and the surrounding nuclear part of the epithelial glands and the lamina propria in which the glands were organized. In the case of malignant tissue, the crypts were no longer well-differentiated, and no particular cluster could be attributed to either the central or the nuclear part. The mucosal structures were no longer individualized, and only two components could be distinguished: the epithelial one and the associated stroma.

By comparing the normal crypts and the adenocarcinoma, surprisingly the IR spectra of normal crypts were associated with relatively higher intensities of nucleic acids than in the adenocarcinomatous epithelial component. This is in contrast to other studies that have showed increased nucleic acid intensities in tumoral samples when compared to the normal samples.[9] Another study has shown decreased intensity of PO_2^- asymmetric stretch of nucleic acids in tumoral tissue while increased intensity of PO_2^- symmetric stretch of nucleic acids.[31]

One of the possibilities for this observation is likely that the spectral alterations involving nucleic acids are less marked since the normal colon cells themselves are highly proliferative and have high mitotic rate and, in tumors that are moderately differentiated, the cellular proliferation is only slightly increased.[24] Interestingly, there are also studies which have shown that the spectral differences observed between a normal and a tumoral tissue actually

may correspond to the differences originating from the different phases of cell cycles, since the opacity of DNA to IR radiation is based on the cell cycle phase which is related to the DNA packing and condensing. [32]

Usually, the normal colon crypts are rich in mucin. However, its corresponding peaks were not discriminatory when the all the normal and the tumoral samples were compared. This could be explained from the fact that the presence of a mucinous tumor diminishes the spectral differences between the mucin rich normal crypts and the tumoral tissues. Interestingly, in comparison of the mucinous adenocarcinoma tissue with the non-mucinous regions of the normal crypts, mucin corresponding peaks reappeared as discriminant features. These results which corroborated with the histopathology show the ability of IR spectroscopy in identifying biomolecular changes in respect to the analyzed tissue types based on the spectral characteristics. The identification of subtle changes involving mucin could be used to characterize tumor types in colon cancers.

The same tendency of higher intensities was observed for carbohydrate and phospholipids between the normal and the tumoral tissues. On the other hand, higher amide I intensities were associated with adenocarcinoma probably indicating greater accumulation of proteins during carcinogenesis and progression.

Another interesting observation arises from changes in the relative intensities of the vibrations involving the H-bonded C-O and non-H-bonded C-O bond vibrations of proteins. While the former is more pronounced in the normal tissues, the latter is more in the tumoral tissues. Similar changes have been observed in earlier studies on colon cancers that probably indicate the molecular changes associated with the amino acid side chains involving tyrosine, serine and threonine.[31, 33, 34] Finally, the observed difference in the spectral profiles of nucleotides, proteins, phospholipids and carbohydrates, between the benign and the malignant tissues appears as an interesting discriminating feature in moderately differentiated colon cancers.

For characterizing the tumoral tissue (figure 3B), 14 clusters were necessary to identify the tumor together with its associated stroma. These two clusters showed very close spectral profiles, an observation that supports the view that stroma is intimately associated to its tumor. In spite of this, the highly sensitive statistical methods enabled to depict subtle differences that could be probably associated with the spectral profiles of collagen features together with the amide II regions of the proteins, and other stroma-associated proteins in malignancy.

Conclusion

This study demonstrates the potential of IR spectral imaging for identifying and differentiating various histological features of normal and tumoral paraffin-embedded colon tissue arrays. An important aspect is that large spots (3 mm-diameter) of the paraffinized tissue array stabilized in an agarose matrix could be directly analyzed without chemical dewaxing thus simplifying the experimental protocol. This procedure was enabled by the implementation of an optimized version of the EMSC algorithm permitting to numerically neutralize both paraffin and agarose spectral contributions. Additionally, using multivariate analysis, complementary information on the changes associated with the biochemical properties between normal and malignant tissues could be also recovered, in a single measurement and in a label-free manner. The translation of this methodology of IR imaging is envisaged to paraffinized tissue microarrays that can enable high-throughput, molecular level analysis of large tissue archives. These optimistic results open a new way for developing spectral biomarkers and libraries which could be used, in complement to conventional histopathology, for early diagnosis, and also potentially for prognosis and theranostics of cancers.

Acknowledgments

This study was supported by a grant of Institut National du Cancer (INCa) and Canceropôle Grand Est. We would like to thank Ligue contre le Cancer, Conférence de Coordination Interrégionale du Grand-Est, and CNRS Projets Exploratoires Pluridisciplinaires, for financial support. NJ is a recipient of doctoral fellowship from the Région Champagne-Ardenne.

References

1. F. L. Martin, J. G. Kelly, V. Llabjani, P. L. Martin-Hirsch, Patel, II, J. Trevisan, N. J. Fullwood and M. J. Walsh, "Distinguishing cell types or populations based on the computational analysis of their infrared spectra," Nat Protoc 5(11), 1748-1760 (2010)

2. R. K. Sahu, S. Argov, E. Bernshtain, A. Salman, S. Walfisch, J. Goldstein and S. Mordechai, "Detection of abnormal proliferation in histologically 'normal' colonic biopsies using FTIR-microspectroscopy," Scand J Gastroenterol 39(6), 557-566 (2004)

3. H. W. Lasch P, Lewis E, Kidder L, Naumann D, "Characterization of Colorectal Adenocarcinoma Sections by Spatially Resolved FT-IR Microspectroscopy," Appl Spectrosc 56(1), (2002)

4. A. Tfayli, O. Piot, A. Durlach, P. Bernard and M. Manfait, "Discriminating nevus and melanoma on paraffin-embedded skin biopsies using FTIR microspectroscopy," Biochim Biophys Acta 1724(3), 262-269 (2005)

5. H. Fabian, N. A. Thi, M. Eiden, P. Lasch, J. Schmitt and D. Naumann, "Diagnosing benign and malignant lesions in breast tissue sections by using IR-microspectroscopy," Biochim Biophys Acta 1758(7), 874-882 (2006)

6. W. Steller, J. Einenkel, L. C. Horn, U. D. Braumann, H. Binder, R. Salzer and C. Krafft, "Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging," Anal Bioanal Chem 384(1), 145-154 (2006)

7. A. Travo, O. Piot, R. Wolthuis, C. Gobinet, M. Manfait, J. Bara, M. E. Forgue-Lafitte and P. Jeannesson, "IR spectral imaging of secreted mucus: a promising new tool for the histopathological recognition of human colonic adenocarcinomas," Histopathology 56(7), 921-931 (2010)

8. M. J. Nasse, M. J. Walsh, E. C. Mattson, R. Reininger, A. Kajdacsy-Balla, V. Macias, R. Bhargava and C. J. Hirschmugl, "High-resolution Fourier-transform infrared chemical imaging with multiple synchrotron beams," Nat Methods 8(5), 413-416 (2011)

9. M. J. German, A. Hammiche, N. Ragavan, M. J. Tobin, L. J. Cooper, S. S. Matanhelia, A. C. Hindley, C. M. Nicholson, N. J. Fullwood, H. M. Pollock and F. L. Martin, "Infrared spectroscopy with multivariate analysis potentially facilitates the segregation of different types of prostate cell," Biophys J 90(10), 3783-3795 (2006)

10. K. Yano, S. Ohoshima, Y. Gotou, K. Kumaido, T. Moriguchi and H. Katayama, "Direct measurement of human lung cancerous and noncancerous tissues by fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool?," Anal Biochem 287(2), 218-225 (2000)

11. T. D. Wang, G. Triadafilopoulos, J. M. Crawford, L. R. Dixon, T. Bhandari, P. Sahbaie, S. Friedland, R. Soetikno and C. H. Contag, "Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy," Proc Natl Acad Sci U S A 104(40), 15864-15869 (2007)

12. X. Zhang, Y. Xu, Y. Zhang, L. Wang, C. Hou, X. Zhou, X. Ling and Z. Xu, "Intraoperative Detection of Thyroid Carcinoma by Fourier Transform Infrared Spectrometry," J Surg Res (2010)

13. C. Krafft, S. B. Sobottka, K. D. Geiger, G. Schackert and R. Salzer, "Classification of malignant gliomas by infrared spectroscopic imaging and linear discriminant analysis," Anal Bioanal Chem 387(5), 1669-1677 (2007)

14. E. O. Faolain, M. B. Hunter, J. M. Byrne, P. Kelehan, H. A. Lambkin, H. J. Byrne and F. M. Lyng, "Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents," J Histochem Cytochem 53(1), 121-129 (2005)

15. J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," Int J Cancer 127(12), 2893-2917 (2010)

16. H. Miyoshi, M. Oka, K. Sugi, O. Saitoh, K. Katsu and K. Uchida, "Accuracy of detection of colorectal neoplasia using an immunochemical occult blood test in symptomatic referred patients: comparison of retrospective and prospective studies," Intern Med 39(9), 701-706 (2000)

17. D. K. Rex, "Colon tumors and colonoscopy," Endoscopy 32(11), 874-883 (2000)

18. T. J. Zuber, "Flexible sigmoidoscopy," Am Fam Physician 63(7), 1375-1380, 1383-1378 (2001)

19. M. Khanmohammadi, A. B. Garmarudi, K. Ghasemi, H. K. Jaliseh and A. Kaviani, "Diagnosis of colon cancer by attenuated total reflectance-Fourier transform infrared microspectroscopy and soft independent modeling of class analogy," Med Oncol 26(3), 292-297 (2009)

20. M. Khanmohammadi, A. Bagheri Garmarudi, S. Samani, K. Ghasemi and A. Ashuri, "Application of linear discriminant analysis and Attenuated Total Reflectance Fourier Transform Infrared microspectroscopy for diagnosis of colon cancer," Pathol Oncol Res 17(2), 435-441 (2010)

21. H. Martens, J. P. Nielsen and S. B. Engelsen, "Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures," Anal Chem 75(3), 394-404 (2003)

22. A. Kohler, C. Kirschner, A. Oust and H. Martens, "Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin," Appl Spectrosc 59(6), 707-716 (2005)

23. E. Ly, O. Piot, R. Wolthuis, A. Durlach, P. Bernard and M. Manfait, "Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies," Analyst 133(2), 197-205 (2008)

24. P. Lasch, W. Haensch, D. Naumann and M. Diem, "Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis," Biochim Biophys Acta 1688(2), 176-186 (2004)

25. D. C. Fernandez, R. Bhargava, S. M. Hewitt and I. W. Levin, "Infrared spectroscopic imaging for histopathologic recognition," Nat Biotechnol 23(4), 469-474 (2005)

26. J. T. Kwak, R. Reddy, S. Sinha and R. Bhargava, "Analysis of variance in spectroscopic imaging data from human tissues," Anal Chem 84(2), 1063-1069 (2012)

27. P. Schraml, J. Kononen, L. Bubendorf, H. Moch, H. Bissig, A. Nocito, M. J. Mihatsch, O. P. Kallioniemi and G. Sauter, "Tissue microarrays for gene amplification surveys in many different tumor types," Clin Cancer Res 5(8), 1966-1975 (1999)

28. S. W. Bruun, A. Kohler, I. Adt, G. D. Sockalingum, M. Manfait and H. Martens, "Correcting attenuated total reflection-Fourier transform infrared spectra for water vapor and carbon dioxide," Appl Spectrosc 60(9), 1029-1039 (2006)

29. D. Sebiskveradze, V. Vrabie, C. Gobinet, A. Durlach, P. Bernard, E. Ly, M. Manfait, P. Jeannesson and O. Piot, "Automation of an algorithm based on fuzzy clustering for analyzing tumoral heterogeneity in human skin carcinoma tissue sections," Lab Invest 91(5), 799-811 (2011)

30. R. Wolthuis, A. Travo, C. Nicolet, A. Neuville, M. P. Gaub, D. Guenot, E. Ly, M. Manfait, P. Jeannesson and O. Piot, "IR spectral imaging for histopathological characterization of xenografted human colon carcinomas," Anal Chem 80(22), 8461-8469 (2008)

31. B. Rigas, S. Morgello, I. S. Goldman and P. T. Wong, "Human colorectal cancers display abnormal Fourier-transform infrared spectra," Proc Natl Acad Sci U S A 87(20), 8140-8144 (1990)

32. S. Boydston-White, T. Gopen, S. Houser, J. Bargonetti and M. Diem, "Infrared spectroscopy of human tissue. V. Infrared spectroscopic studies of myeloid leukemia (ML-1) cells at different phases of the cell cycle," Biospectroscopy 5(4), 219-227 (1999)

33. F. P. Conti C, Giorgini E, Rubini C, Sabbatini S, Tosi G, Anastassopoulou J, Arapantoni P, Boukaki E, Konstadoudakis S, Theophanides T, Valavanis C, "FT-IR microimaging spectroscopy: A comparison between healthy and neoplastic human colon tissues," Journal of Molecular Structure 881(46-51 (2008)

34. S. L. Patrick T. T. Wong, Hossein M. Yazdi, "Normal and Malignant Human Colonic Tissues Investigated by Pressure-Tuning FT-IR Spectroscopy," in Applied Spectroscopy (1993).

Disclosure/Conflict of Interest

The authors declare no conflict of interest.

III.5: Article 3

Infrared spectral histopathology for cancer diagnosis; a novel approach for automated pattern recognition of colon adenocarcinoma

(Submitted to the Journal of Cancer Research, July 2012)

Préambule à l'article 3

Contexte

L'histopathologie reste la méthode de référence pour le diagnostic du cancer du côlon. De nouvelles approches complémentaires sont régulièrement testées et évaluées en complémentarité de cette méthode de référence.

L'imagerie infrarouge apparaît comme une méthode très intéressante parce qu'elle permet la mise en évidence des liaisons chimiques intrinsèques présentes dans un tissu, fournissant une «signature spectrale» spécifique de la composition biochimique et une cartographie moléculaire des différentes structures.

Objectif

L'histopathologie spectrale IR, qui associe l'imagerie IR et les méthodes multivariées de traitement de données, a été mise en œuvre. Les caractéristiques biochimiques et structurales des tissus ont été identifiées en vue de réaliser un modèle de prédiction qui pourrait être utilisé à des fins diagnostics.

Matériels et Méthodes

Quatre-vingts coupes de tissus de côlon paraffinées sous la forme de tissu microarrays ont été analysées par imagerie IR. Pour éviter l'étape de déparaffinage chimique des coupes, une méthode de correction mathématique appellée « Extended Multiplicative Signal Correction » (EMSC) a été utilisée pour neutraliser les interférences spectrales de la paraffine et de l'agarose. La méthode de clustering par k-means a ensuite été utilisée pour classer les spectres et ainsi établir une image en fausses couleurs en fonction des différentes structures tissulaires (cryptes, lamina propria, tumeur, etc). Des coupes adjacentes colorées par HPS sont utilisées comme référence. L'analyse discriminante linéaire (LDA) a ensuite été employée pour construire un modèle de prédiction basé sur les résultats du k-means, en utilisant 9 échantillons pour la validation interne. Ce modèle a ensuite été appliqué à 71 échantillons en validation externe.

Résultats

Comparé aux images colorées par HPS, les images spectrales, après l'attribution d'un code couleur, révèlent non seulement des caractéristiques communes représentatives de la composition biochimique des tissus, mais permettent également de mettre en évidence des caractéristiques supplémentaires comme le phénomène de tumeur budding, le stroma associé à la tumeur, etc sans étape de marquage au préalable.

Conclusion

Cette nouvelle approche d'imagerie spectrale infrarouge sur des biopsies de tissus paraffinées a permis la détection et la différenciation des tissus normaux et tumoraux du côlon en se basant uniquement sur leurs caractéristiques biochimiques intrinsèques avec une sensibilité de 100 %. Cette méthodologie, ne nécessitant aucun marquage des coupes, combinée à une analyse statistique multivariée des images, apparaît comme un outil prometteur pour le diagnostic du cancer du côlon et confirme le potentiel du concept d'histopathologie spectrale.

Title:

Infrared spectral histopathology for cancer diagnosis; a novel approach for automated pattern recognition of colon adenocarcinoma

Authors' names and institutional affiliation:

Jayakrupakar Nallala[1], Marie-Danièle Diebold[1, 2], Cyril Gobinet[1], Olivier Bouché[1, 3], Ganesh Dhruvananda Sockalingum[1], Olivier Piot[1], Michel Manfait[1*]

1. MéDIAN Biophotonique et Technologies pour la Santé, Université de Reims Champagne-Ardenne, FRE CNRS 3481 MEDyC, UFR de Pharmacie, SFR Cap Santé, 51 rue Cognacq-Jay, 51096 Reims cedex, France.

2. Laboratoire d'Anatomie et Cytologie Pathologiques, CHU Robert Debré, Avenue du Général Koenig, 51092 Reims Cedex, France.

3. Service d'Hépato Gastroentérologie et de Cancérologie Digestive, CHU Reims, Avenue du Général Koenig, 51092 Reims Cedex, France.

Tel: +33 32 69 18 12 8, Fax: +33 32 69 13 55 0

Abstract:

Histopathology remains the gold standard method for colon cancer diagnosis. Novel complementary approaches are being tested and evaluated for molecular level diagnosis of the disease. Infrared (IR) imaging could be a good candidate method as it probes the intrinsic chemical bonds present in a tissue, and provides a "spectral fingerprint" of the biochemical composition and the structures. To this end, IR spectral histopathology, which combines IR imaging and data processing techniques, in order to identify tissue biochemical and structural characteristics in view of implementing a prediction model which could be used for diagnostic purposes, has been implemented. Eighty paraffinized colon tissue sections in the form of tissue microarrays were analyzed directly by IR imaging. To avoid chemical deparaffinization, a modified Extended Multiplicative Signal Correction (EMSC) method was used to digitally neutralize the spectral interferences of paraffin. K-means clustering was then used to partition the spectra and construct color-coded images, for assigning spectral clusters to various tissue structures (crypts, lamina propria, tumor, etc) using the adjacent HPS stained sections as reference. Based on the k-means results, Linear Discriminant Analysis (LDA) was then used to construct a stringent prediction model using an internal validation set (9 samples). This model was then applied to an external validation set (remaining 71 samples). When compared to HPS stained images, color-coded spectral images not only reveal common features representative of the biochemical make up of the tissues, but also highlight additional features like tumor budding, tumor associated stroma,. etc in a label-free manner. This novel approach of IR imaging on paraffinized tissue biopsies allowed detection and differentiation of normal and tumoral colon tissues based only on their intrinsic biochemical features. This label-free methodology combined with multivariate statistical image analysis appears as a promising tool for colon cancer diagnosis and opens the way to the new concept of numerical spectral histopathology.

Introduction:

Colorectal cancer has one of the highest incidence and mortality among all the cancers affecting both sexes, of which the type adenocarcinoma is the most common (1). Radiation therapy, chemotherapy and surgical intervention have improved the life expectancy of cancer patients, but the outcome of these methods is dependent upon the stage and the accuracy in diagnosis (2). Currently different detection and screening methods are employed for colorectal cancers, including fecal occult blood test (FOBT) (3), sigmoidoscopy (4), colonoscopy (5), etc. However, the final diagnosis is settled upon the microscopic examination of the symptomatic tissue with the 'gold standard' histopathology in which preferential stains are used to enhance visualization of the tissue morphological aberrations. Such aberrations (pre-cancerous or cancerous) are the manifestations of the biomolecular changes that have already undergone the provocative changes for malignancy. However, the ongoing state of the tissue molecular changes during the onset or progression of malignancy, without any morphological signatures, poses challenge for identification. In certain cases, immunohistochemistry (IHC) is used to identify specific proteins of interest which can give a molecular level understanding of the malignant condition. Histopathology requires precise human expertise which limits high-throughput diagnosis. Although, the histopathological diagnosis is based on morphological examination, it has successfully served in cancer diagnosis over several years. Additionally, if it is combined with approaches that could provide complementary biochemical information in a rapid, cost effective manner and reducing human involvement, the efficacy of the histopathological diagnosis can be completed.

In this regard, optical spectroscopic approach of infrared (IR) imaging appears as a potential candidate for routine tissue characterization, and has been exploited as a diagnostic tool on various tissues (6-16). IR spectroscopy probes intrinsic chemical bond vibrations of biomolecules and thus provides a biochemical fingerprint of the tissues. Combined with an imaging set-up, spectral images can be obtained rapidly in a label-free manner, in which each pixel element harbors an IR spectrum containing biochemical information at each wavenumber. Such IR images can be exploited using computer based multivariate cluster analysis to generate digitally stained morphological maps of the tissue histology. Since the constituent IR spectra of each digitally stained histological class represent its biochemical signature, such as collagen features in the connective tissue, specific spectral signatures can be identified from different histological classes. Such signatures can be used to train

predictive algorithms for identification of unknown tissues in a rapid and user-friendly manner. One of the important possibilities of using this methodology is automation of this protocol which can reduce human involvement and provide an objective biochemical based diagnostic approach.

In this regard, we carried out spectral histopathology based on IR imaging in conjunction with multivariate analysis. The main objectives were to digitally detect and identify malignancy and its associated features on unknown tissues without any chemical staining, constituting an automated diagnosis for colon adenocarcinoma. For this, 80 human colon tissues from normal and moderately differentiated adenocarcinoma were analyzed, in the form of paraffinized tissue arrays that were stabilized in an agarose matrix. The tissue arrays are increasingly used in pathological studies since they constitute a humongous source of information and permit high-throughput analysis for modern histological practices (17). An innovative processing of digital deparaffinization was specially implemented to avoid chemical dewaxing, and also to reduce toxic chemical treatments and time consumption (18). Then, a prediction model representing the main colonic histological classes was constructed and its robustness was evaluated on subsequent number of tissue array cores. Digital annotation using this model facilitated characterization of malignancy, and malignancy associated features such as tumor budding, and tumor-stroma association.

Materials and Methods:

Sample preparation: Eighty formalin fixed paraffin embedded (FFPE) colon tissue samples (47 tumoral and 33 non-tumoral) from 35 cancer patients were obtained from the Reims University Hospital, with the approval of the Institutional Review Board. All the tumoral samples were moderately differentiated colon adenocarcinoma with the TNM grade ranging from T3N0M0 to T4N2M0. The sample details are presented in supplementary information 1. Several paraffinized tissue arrays that were stabilized in an agarose matrix were manually prepared from these samples using the pathology laboratory protocols as described in (Article 1 of Chapter III).

Supplementary information 1: Sample details.

Patient No.	Sex/Age	Tumor			Normal	TMA No.	Colon Carcinoma location	TNM classification	Non-tumoral tissue location
		Front (A)	Lateral (B)	Middle (C)	Distant (D)				
1	F, 67				TG	08_TMA_DC_1_1	L	T3N0	L
2	M, 71	**		TG	TG	08_TMA_DC_1_2	L	T3 N0	L
3	F, 74	**	**		**	08_TMA_DC_1_4	L	T3 N0	L
4	F, 48				LF	08_TMA_DC_7_1	L	T3 N0	L
5	F, 61				**	08_TMA_DC_7_2	R	T3 N0	R
6	M, 70	**			**	08_TMA_DC_7_3	L	T3 N1	L
7	M, 76					08_TMA_DC_7_4	L	T3 N1	L
8	F, 62				TG	08_TMA_DC_13_1	L	T3 N0	L
9	F, 66					08_TMA_DC_13_2	L	T3 N2	L
10	F, 72	**	TG		LF	08_TMA_DC_13_3	R	T4 N0	R
11	M, 51				LF	08_TMA_DC_13_4	R	T3 N1	R
12	F, 51		TG	TG	TG	08_TMA_DC_14_1	L	T3 N1	L
13	M, 66					08_TMA_DC_14_2	L	T3 N0	L
14	F, 57	**				08_TMA_DC_14_3	L	T3 N1	L
15	F, 41		TG		LF	08_TMA_DC_18_1	R	T3 N1	R
15	As above	LF				08_TMA_DC_18_2	R	T3 N1	R
16	M, 73					08_TMA_DC_18_3	L	T3 N2	L
17	F, 46				**	08_TMA_DC_18_4	R	T4 N2	R
18	F, 91					12_TMA_DC_1_1A			R
19	F, 78					12_TMA_DC_1_1B			R
20	M, 41					12_TMA_DC_1_1C			R
21	M, 72					12_TMA_DC_1_2A			R
22	M, 54				LF	12_TMA_DC_4_PT_1A			Peri-tumoral
23	M, 68				LF	12_TMA_DC_4_PT_1C			Peri-tumoral
24	F, 82				LF	12_TMA_DC_4_PT_2A			Peri-tumoral
25	F, 90				LF	12_TMA_DC_4_PT_2B			Peri-tumoral
26	M, 64					12_TMA_DC_2_1B			Sigmoid
27	M, 64					12_TMA_DC_2_1C			Sigmoid
28	M, 82				LF	12_TMA_DC_2_2B			Sigmoid
29	M, 69				LF	12_TMA_DC_2_2C			Sigmoid
30	F, 53					12_TMA_DC_6_1A			Sigmoid
31	F, 79					12_TMA_DC_6_1C			R
32	F, 73				LF	12_TMA_DC_6_2B			Sigmoid
33						TMA_2_NT			
34						TMA_2_2D			
35						TMA_2_3B			

	Analyzed
**	Not analyzed
TG	Training group
LF	Inflammation
L	Left
R	Right

A single sample spot in the tissue array block was approximately 3 mm in diameter. For each tissue array consisting around 12-16 spots, 3 and 10 µm thick sections (adjacent in most cases) were obtained. While the 3 µm section was used by the pathologist for conventional histopathological analysis via hematoxylin, phloxine, and saffron (HPS) staining, the first 10 µm unstained section was used for IR imaging analysis and the second stained section for additional histopathological comparison. The HPS stained sections were chemically deparaffinized while the unstained tissue section for IR imaging was mounted on an IR compatible calcium fluoride (CaF_2) support without any chemical deparaffinization. A schematic representation of the IR imaging methodology of the tissue arrays is presented in figure 1.

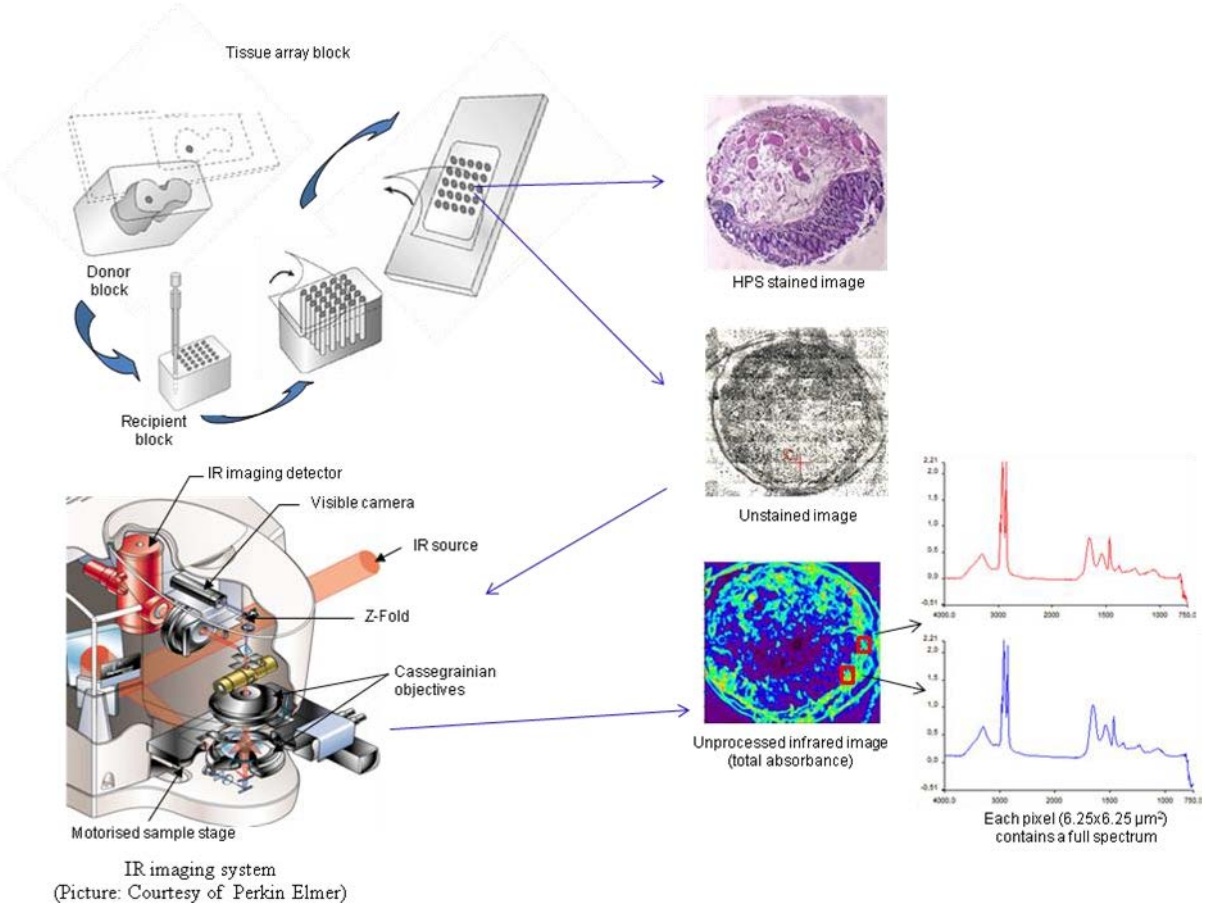


Figure 1: Schematic representation of infrared spectral imaging applied to paraffinized tissue arrays.

Instrumentation and FTIR data collection: IR images were acquired, by an IR imaging system (Spotlight 300, Perkin Elmer, Courtaboeuf, France) equipped with liquid nitrogen-cooled 16-element MCT detector, at 6.25 x 6.25 μm^2 pixel size, and 4 cm^{-1} spectral resolution averaged to 16 scans, in the mid-IR range of 750 to 4000 cm^{-1}. The system was continuously purged with dry air. The background spectrum from the CaF_2 support was recorded each time prior to image acquisition, using same parameters as that of the IR image. A total of 8540899 IR spectra were recorded from 80 images at an average of 106761 per image owing to the large size of the tissue array spots, and the high spatial resolution selected for imaging.

Data pre-processing: Raw IR data was corrected from various spectral interferences. An atmospheric correction was performed to remove contribution from water vapour and CO_2 by the built-in Perkin Elmer Spotlight software and further processing was carried out using programmes written in Matlab 7.2 (The Mathworks, Natick, MA). The spectra were reduced to the IR absorption range of 900-1800 cm^{-1} that contains several informative biochemical vibrations (19, 20) as far as the tissue features are considered. Neutralization of paraffin and agarose contributions was carried out using a modified Extended Multiplicative Signal Correction (EMSC) algorithm as detailed in (Article 1 of Chapter III). The IR spectra were also corrected for baseline and then normalized using the same algorithm. Outliers (N=3468314 spectra) in the form of paraffin and agarose spectra, and spectra with poor signal-noise ratio were eliminated from the analysis and were depicted as white pixels in all the IR images.

Data processing: The pre-processed data (N=5072585 spectra) was subjected to multivariate statistical prediction analysis. For this, spectral data from the non-tumoral and the tumoral samples was separated into a training group (SI 1 sample # TG), and a validation group. While the training group, representing the IR spectral signatures indicative of malignancy and other histological structures, was used for construction of a prediction model based on linear discriminant analysis (LDA), the validation group (external validation) was used for validating the model on unknown samples for automatic recognition of tissue features, to enable identification of malignancy. LDA is a multivariate supervised statistical technique that aims at maximizing the between-class variance and minimizing the within-class variance (21) and has been exploited in various studies (20, 22).

Cluster analysis for LDA training: The huge number of IR spectra from each image corresponding to the training group was subjected to unsupervised k-means clustering

method owing to its capability of rapid and huge data clustering (23). This method iteratively partitions the spectra into different clusters based on the spectral signatures from the intrinsic biochemical composition of the tissue. Therefore, spectra with similar biochemical characteristics group into the same cluster. In k-means, each spectrum belongs to a unique cluster and can thus be represented by one color. K-means clustering performed using defined cluster numbers resulted in the construction of digital color-coded images. These were then compared to adjacent HPS stained sections to annotate by an expert pathologist, each spectral cluster to the tissue structural feature that it corresponds to. The spectral distance between different k-means clusters was visualized in a dendrogram obtained by hierarchical clustering analysis using Ward's linkage algorithm.

Prediction model: The initially k-means clustered and annotated spectra were used as inputs for the LDA model. Training group spectra (SI 1 sample # TG) from 9 samples across 6 different patients were considered for the model, to take into account the inter-patient variability. The prediction model consisted of 8 classes with different number of spectra, representing various histological features of non-tumoral and tumoral tissues: the normal epithelium defined by the crypt inner-part (crypt-IP) (N = 8377) and the crypt outer-part (crypt-OP) (N = 3567), the lamina propria (N = 14106), the submucosa (N = 3964), the tumor epithelium (N = 35083), the tumor-associated stroma (N = 16409), the blood vessel (N = 782) and the muscularis propria (N = 4514). These spectra (N=86802) constituting one-third of the spectra from each class were used to train the model and the other two-third were used for an internal validation to optimize the model. The prediction model was then applied in an external validation on different unknown samples, the spectra from which were secluded to the model, to evaluate its robustness. The external validation consisted of 71 samples encompassing a large scale spectral data base of 3620287 spectra that were to be identified. All the predictions were carried out at a posterior probability of 0.5 and in the IR spectral range of 1080 cm^{-1} - 1300 cm^{-1} as discussed later.

Spectral information to biochemical information (spectral analysis): Since the spectral signatures are based on the biochemical properties of the tissue features, it was attempted to characterize, the biochemical alterations characteristic of malignancy, and the relationship of malignant tissue with the surrounding stroma. For this, the Mann-Whitney *U* test was applied

to compare spectra from selected cluster groups used in the prediction model training in order to identify the most discriminant wavenumbers.

Immunohistochemistry (IHC): IHC was used as a complementary tool (on adjacent sections) to enhance visibility of tumor budding (Anti-Human Cytokeratins-large spectrum Monoclonal Antibody, Clone KL 1, dilution 1/50, Immunotech, France) and precise the nature of the inflammatory cells: T-lymphocytes (CD3 Rabbit anti-Human Polyclonal Antibody, dilution 1/200, Dako, France), and B-lymphocytes (CD20 Mouse antibody, clone L6 mouse, dilution 1/400, Dako, France), in order to validate some of the important observations detected by IR spectral imaging. This was performed using the fully automated IHC staining protocol (XT ultraView DAB v3).

RESULTS:

Cluster analysis:

K-means clustering was used to identify the spectral signatures characteristic of the main histological features of the non-tumoral and the tumoral colonic tissues, which permitted to construct digitally stained images. For the non-tumoral as well as the tumoral tissues, this approach permitted to identify, and to recover automatically the important histological components in comparison to the adjacent HPS stained images as shown in the figure 2 (SI 1 sample # 1D and 12C). As an example, for the non-tumoral colonic tissue (figure 2A) 8 clusters permitted to view the important histological structures representing the colonic tissue organization. They included the colonic mucosa constituted by well-differentiated crypts (cluster 8 - inner part-IP and cluster 6 - outer part-OP); and the lamina propria (cluster 1), the supportive loose connective tissue in which the crypts are organized.

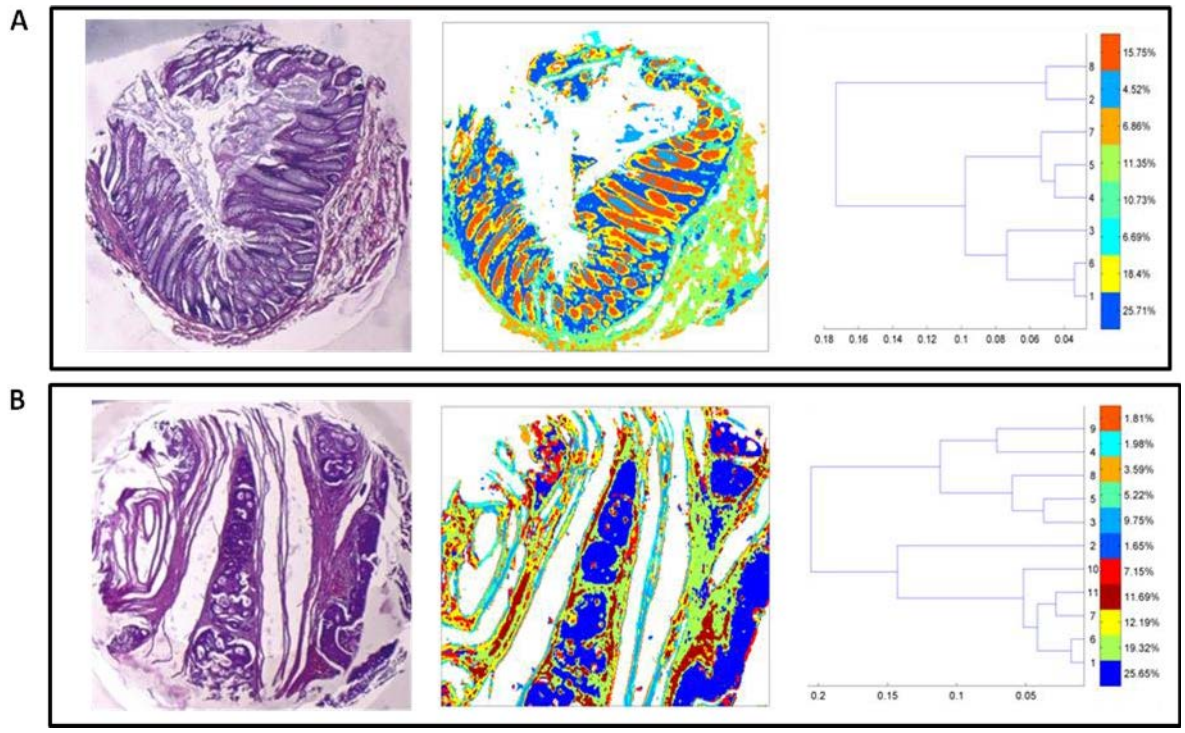


Figure 2: K-means classification and digital staining of FTIR spectral images with random pseudo-colors.

Left panel: HPS stained colon tissues (SI 1 sample # 1D and 12C). Middle panel: K-means classification and digital staining of FTIR spectral images with random pseudo-colors. Right panel: Dendrograms corresponding to the respective cluster images.

A is a non-tumoral colonic tissue classified using 8 clusters representing the major normal colonic tissue features. The cluster representation is as follows: Cluster 1 - lamina propria, cluster 2 - mucus, clusters 4, 5 and 7 - submucosa, cluster 6 - crypt (outer part-OP), cluster 8 - crypt (inner part IP) and cluster 3 - undefined tissue.

B is a moderately differentiated colonic adenocarcinoma classified using 11 clusters. The important histological classes are cluster 1 - tumor, clusters 6, 7, and 11 - tumor-associated stroma. Remaining clusters were attributed to the fibrous stroma. The HPS images are at 5X magnification.

The residual mucin (cluster 2) was observed to be localized within the crypt lumen while a small amount was seen secreted outside. The submucosa, attributed to clusters 4, 5 and 7 was distinguished effectively from the lamina propria by the clustering method. Finally cluster 3 appeared to represent the blood vessels. On the contrary, in the typical adenocarcinomatous tissue (figure 2B), the only important histological classes retrieved were the tumor epithelium (cluster 1) and its associated stroma in the tumor vicinity (cluster 6). Most of the other clusters represented the fibrous stromal tissue. The corresponding dendrogram showed the close spectral nature of the tumor associated stroma to its tumor where they are very closely grouped (clusters 1 and 6) while the stroma that is not in direct contact with the tumor epithelium appear more distant. A total of 11 clusters were required to identify these features. In both cases, considering the overall colonic tissue organization, increasing the number of clusters did not add any further retrievable histological information. The k-means clustering is an efficient method to identify IR spectral markers specific to different histological components of non-tumoral and tumoral colonic tissues. On the basis of these spectral signatures, the diagnostic potential of IR spectral imaging has been evaluated using a LDA based prediction model as schematically represented in figure 3.

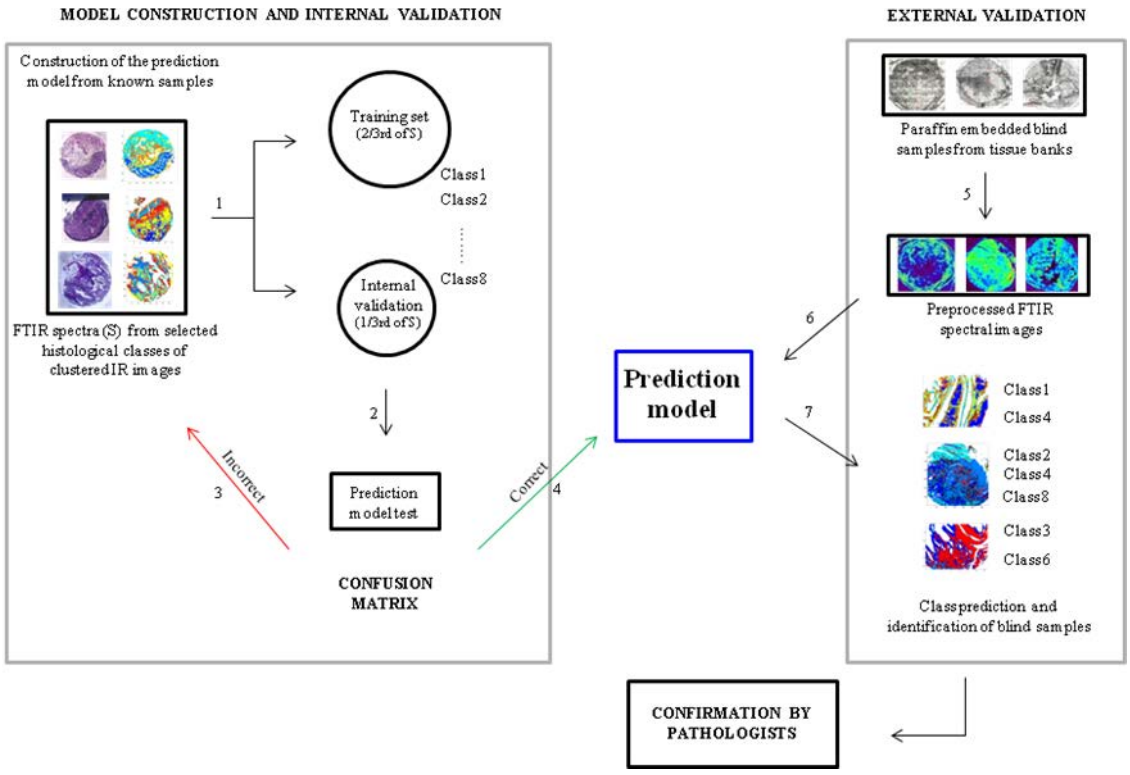


Figure 3: Schematic representation of construction and application of the prediction model based on linear discriminant analysis.

Optimization of the prediction model - internal validation group: The LDA based prediction model developed from 9 samples (6 patients) with 8 different classes comprising a total of 86802 spectra was trained, and tested in an internal validation. The sensitivity of the prediction model in the internal validation can be evaluated from the confusion matrix which shows the confrontation between the histopathological class annotation (real class) and the IR spectral prediction (predicted class) (Table 1). Different spectral regions were tested and the highest sensitivity (average 89.38%) was obtained for the region between 1080 cm^{-1} to 1300 cm^{-1}. It has to be noted that for the class tumor epithelium a specificity of 96.4 % was reached, and showed no confusion with the class normal epithelium (comprising crypt inner part and crypt outer part).

Table 1: The confusion matrix representing the sensitivity of the infrared spectral imaging based prediction model developed using 8 classes, to the gold standard histopathological attribution, in the spectral range of 1080 cm^{-1} to 1300 cm^{-1}. The table shows an average sensitivity of 89. 49 %.

Predicted class (infared imaging)

Histopathological class	'Tumor_epithelium'	'Crypt_OP'	'Lamina_propria'	'Musclaris_propria'	'TAS'	'Crypt_IP'	'Submucosa'	'Blood_vessel'
'Tumor_epithelium'	96,45	0	1,32	0,07	2,03	0	0	0,1
'Crypt_OP'	0,1	88	6,16	0,28	0	4,42	0,5	0,22
'Lamina_propria'	2	1,27	83	0,14	11,34	0,09	1,8	0,45
'Musclaris_propria'	0,1	0,04	0	98,22	1,1	0,04	0	0,04
'TAS'	16,33	0	1	0,08	81,31	0,02	1,24	0
'Crypt_IP'	0,04	6	0,04	0,04	0	93,7	0,16	0,04
'Submucosa'	0,1	0	7,5	0,05	14,42	0	77,54	0,35
'Blood_vessel'	0	0	0	0	2,3	0	0	97,7
No. of spectra used in the model	35083	3567	14106	4514	16409	8377	3964	782

Total = 86802

Tumor detection and tissue characterization in unknown samples - external validation group: The external validation was performed on the remaining 71 blind samples involving a large scale spectral bank of 3620287 spectra and showed 100 % sensitivity for the tumor class. Along with tumor class, other histological classes were also identified with high correlation to the conventional histology. A representative demonstration of prediction on unknown non-tumoral and tumoral samples (SI 1 sample # 14D and 7C) is shown in figure 4. The figure 4A histologically corresponded to a non-tumoral colonic tissue in which the prediction model correctly identified its characteristic features with utmost homology to the

histological image. Counterpart to the normal tissue, histologically the figure 4B corresponded to a typical moderately differentiated colon adenocarcinoma. In this, the spectral characteristics of the normal mucosa were absent and the only distinguished ones were malignant epithelial component with its associated stroma. Additionally, identification of features difficult to discern using conventional techniques, such as tumor budding was facilitated.

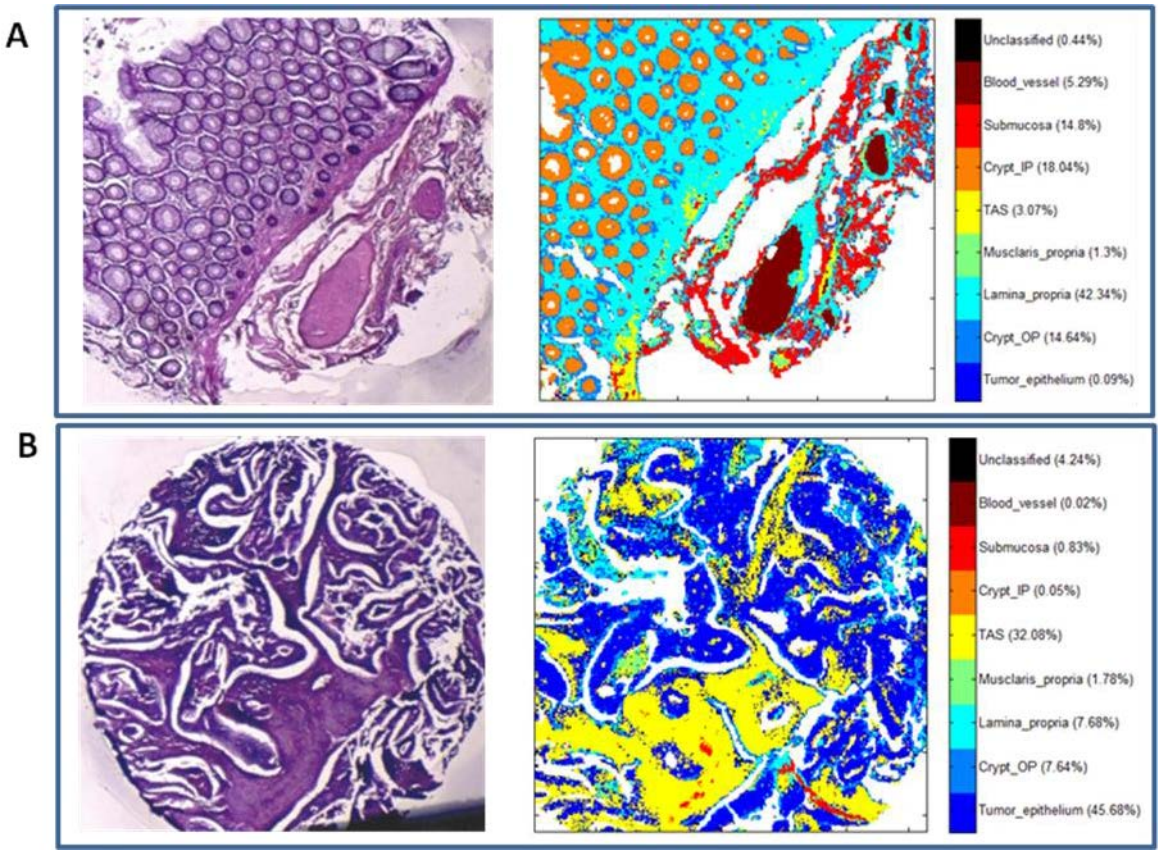


Figure 4: Performance of the prediction model: Identification of unknown colonic tissues by spectral histopathology. Left panel: HPS stained colon tissues (SI 1 sample # 14D and 7C). Right panel: LDA predicted image. A is a non-tumoral colonic tissue section in which all the important normal colonic histological features are well-identified by the model. There is presence of well-differentiated normal epithelium (crypt-IP and crypt-OP). Normal connective tissue is identified, also in which blood vessels are dispersed throughout.

B is a moderately differentiated colon adenocarcinoma in which the tumor epithelium is well-identified together with its associated stroma. There is a complete absence of normal epithelium. The HPS images are at 5X magnification.

Detection and characterization of malignancy associated features:

Tumor budding: Budding is characterized by small clusters of isolated tumor cells which become detached from the neoplastic epithelium and migrate into the stroma, and is an indication of high tumor invasiveness in colorectal cancers. Although this morphological phenomenon is detectable in conventional histopathology at high power magnification, IHC may be employed for better visualization. The IR prediction model was able to clearly identify this tumor particularity even in the presence of abundant stroma as shown in the figure 5, (SI 1 sample # 9B). In the same tumoral sample, along with the malignant epithelium, there was presence of some normal epithelial component together with normal connective tissue, and all these features were identified by the prediction model. Importantly, both the malignant and the non-malignant epithelial cells were selectively stained and discriminated using a specific color-code. The positive staining of the epithelial cells can be seen in the IHC image (see figure 5C).

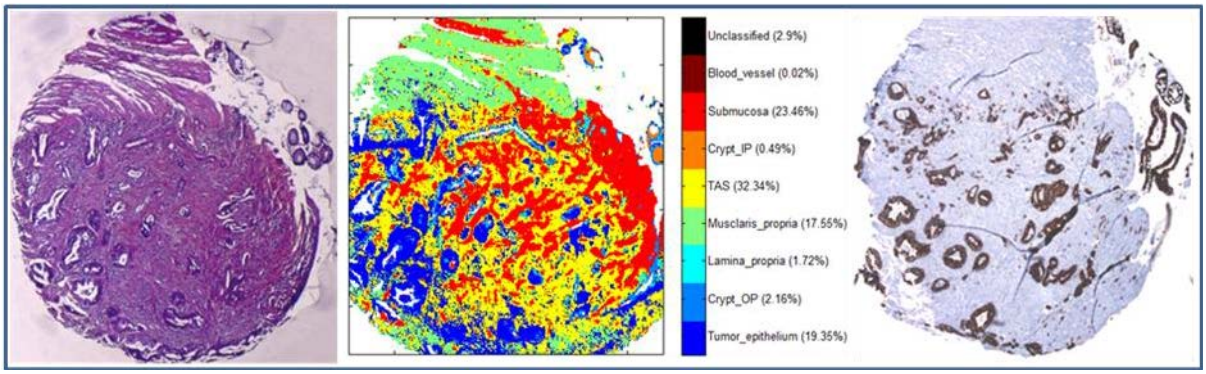
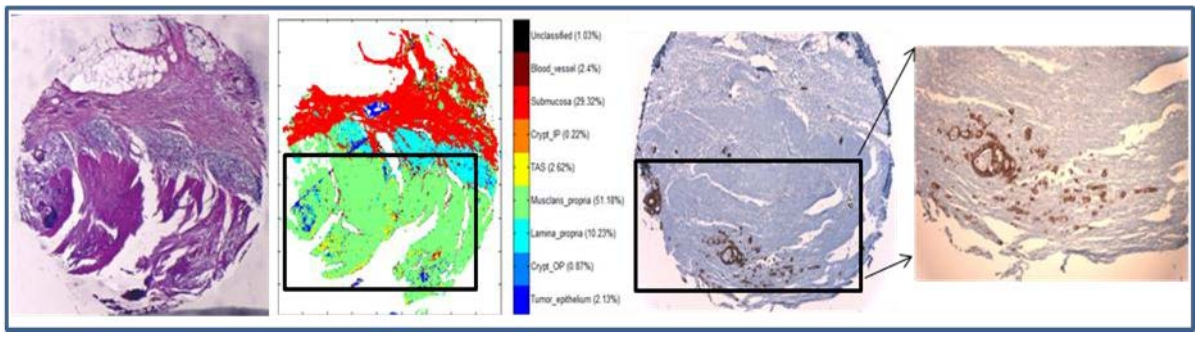


Figure 5: Identification of tumor budding in an unknown colonic tissue.

Left: HPS stained colonic tissue (SI 1 sample # 9B). Middle: LDA predicted image. Right: KL 1 immuno-stained image. The sample is a moderately differentiated colon adenocarcinoma in which the cancerous glands are identified along with the tumor-associated stroma. Small isolated tumor clusters representing tumor-budding are identified branching out into the stroma. The tumor-stromal boundary is also well-identified and clearly demarcated from the normal connective tissue (muscularis propria). In the same sample, few normal colonic glands are seen in the top-right position identified by presence of normal epithelium. The HPS and the IHC images are at 5X magnification.

Another tissue section obtained from different position (SI 1 sample # 9A) of the same tumor also showed tumor budding in a stroma dominant environment, and each time it was identified by the prediction model, which was later confirmed by IHC studies (SI 2).



Supplementary information 2: Identification of tumor budding in an unknown colonic tissue.

Left to right: HPS stained colon tissue (SI 1 sample # 9A), LDA predicted image, KL 1 immuno-stained image, and zoomed area of the same image.

The sample is a moderately differentiated colon adenocarcinoma with tumor-budding branching out into the stroma. The presence of even very few tumor cells sparsely visible in the HPS image seems to be predicted correctly, as can be verified from the immuno-stained image. The HPS and the IHC images are at 5X magnification.

Tumor stroma association:

The tumor-stroma association was also reported using IR spectral imaging. The confusion matrix (table 1) highlighted the spectral proximity of tumor and its associated stroma in which, indeed 16.3 % of tumor associated stroma pixels were classified in the tumor class. Complementarily, in the predicted images (SI 1 sample # 11B) these two classes appeared in geographic proximity (figure 6).

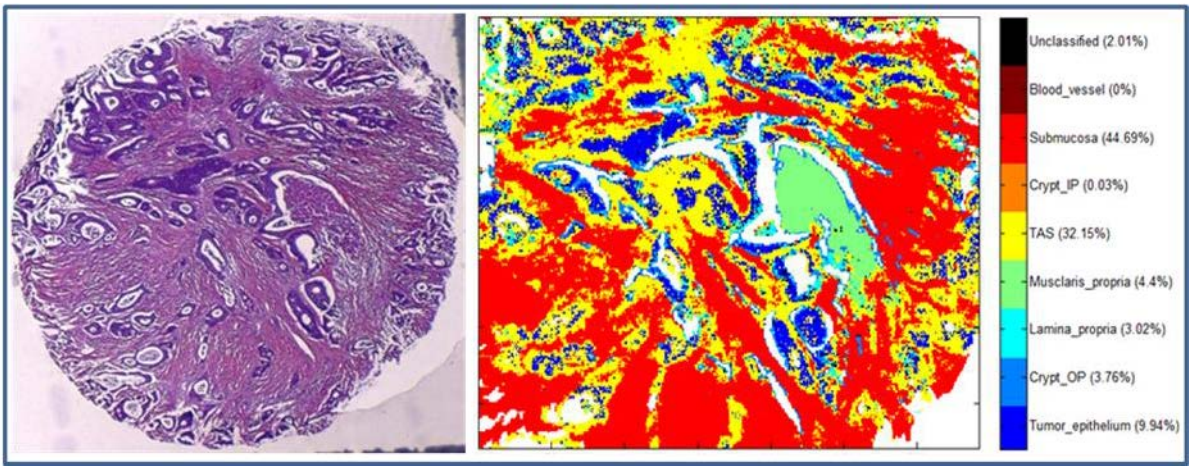
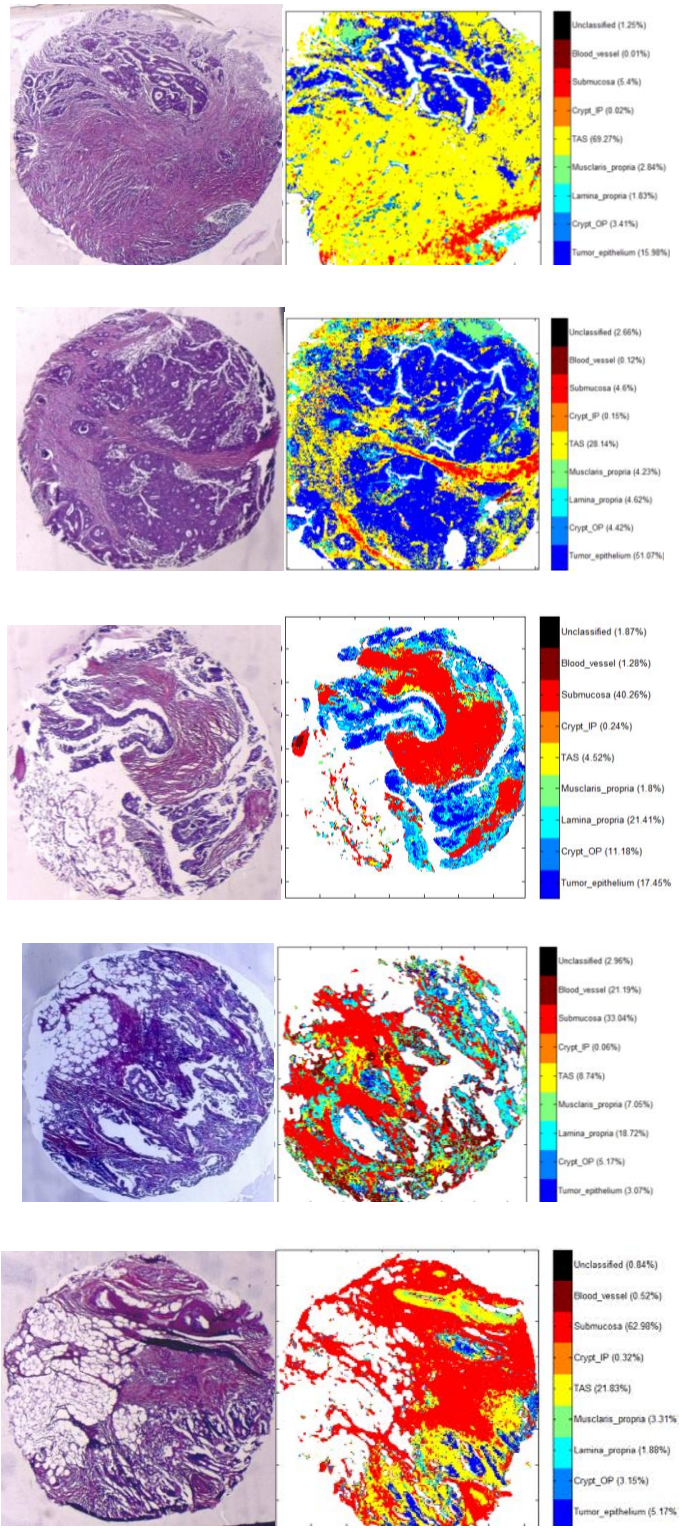


Figure 6: Tumor stroma geographical proximity

The sample is a moderately differentiated colonic adenocarcinoma with its associated stroma (SI 1 sample # 11B). The image is predicted correctly by the model. The nature of the connective tissue into which the tumor has infiltrated is also identified. The HPS image is at 5X magnification.

In the same image, distinction between the tumor associated stroma and the normal connective tissue corresponding to the submucosa was attained, while in the histological stained section, this was indistinguishable. The above mentioned tumor-stroma features were also observed in the other tumoral samples (SI 1 sample # 11A, 11C, 12A, 13A, and 15A) as shown in SI 3 including the cases of budding (fig 5).



Supplementary information 3: Tumor stroma geographical proximity. The samples are moderately differentiated colonic adenocarcinoma with its associated stroma with infiltration into the adjacent connective tissue (SI 1 sample # 11A, 11C, 12A, 13A, and 15A). Along with tumor identification, the nature of the connective tissue into which the tumor has infiltrated is also identified. The HPS images are at 5X magnification.

Vibrational analysis of spectroscopic markers: In this study, the k-means clustering was performed using the IR spectral range of 900 cm^{-1} - 1800 cm^{-1} that enabled identification and attribution of the important colonic histological classes. For unknown sample prediction, this zone was narrowed down to 1080 cm^{-1} to 1300 cm^{-1} harboring some of the important biomolecular vibrational modes implicated in colon cancers, and which showed the best prediction outcome for all the classes together. As shown in figure 7, the most discriminant wavenumbers within this zone were identified by the Mann-Whitney *U* test performed on the individual spectra and represented on the average spectra for the following pair-wise comparisons: normal epithelium with malignant epithelium (adenocarcinoma) for understanding the molecular alterations characteristic of malignancy, and adenocarcinoma with its associated stroma to understand the tumor induced alterations in the stromal tissue.

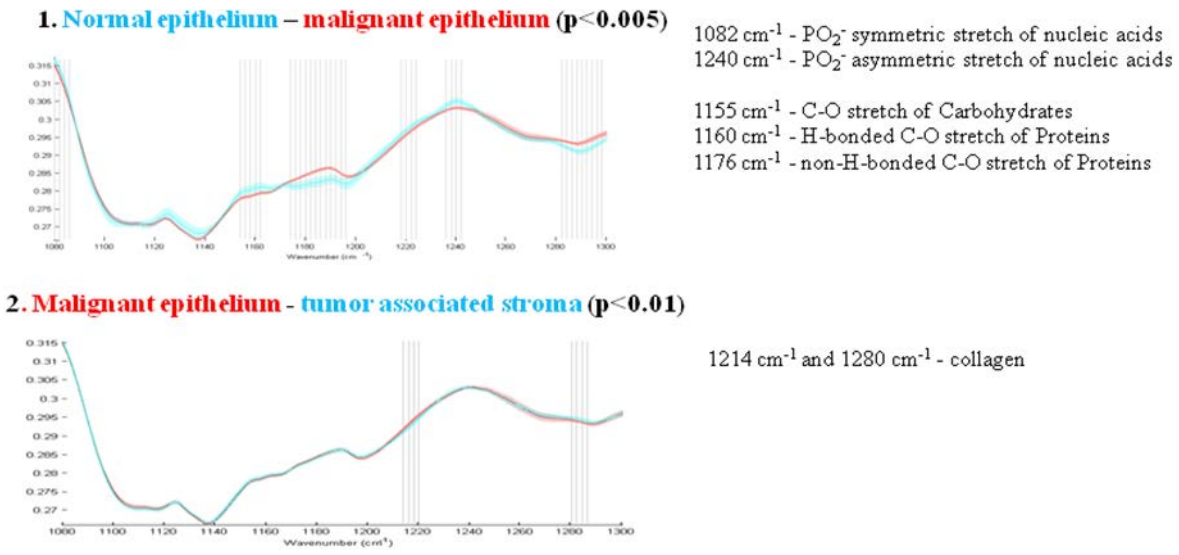


Figure 7: Most discriminant IR spectral vibrations identified by Mann-Whitney *U* test.

The test was performed for 1: Normal epithelium versus tumoral epithelium (p<0.005), and 2: Tumor epithelium versus tumor associated stroma (p<0.01).

From the discriminant wavenumbers identified for all comparisons, a tentative correlation of IR vibrations to the biomolecular information was attempted as shown in Table 2. Importantly, comparing the normal epithelium with the tumoral epithelium, the main

differences in the IR peaks were attributed to symmetric and asymmetric PO_2^- vibrations of the nucleic acids that demonstrated relatively higher intensities in normal than the tumoral tissues. Similarly, the C-O stretching vibration corresponding to carbohydrates was relatively more intense in normal than the tumoral tissues.

Table 2: Correlation of some of the most discriminant IR spectral vibrations

Normal epithelium - malignant epithelium		**Malignant epithelium - Tumor associated stroma**	
Peak position	**Biomolecular attribution**	**Peak position**	**Biomolecular attribution**
1082	PO_2^- symmetric stretch of nucleic acids[12]		
1240	PO_2^- asymmetric stretch of nucleic acids[2]		
		1214	Collagen[2]
1155	C-O stretch of Carbohydrates[39]	1280	
1160	H-bonded C-O stretch of Proteins[39]		
1176	non-H-bonded C-O stretch of Proteins[39]		

At the same time the hydrogen bonded C-O groups of proteins in the normal epithelium was observed to be decreased in the tumoral epithelium, while the opposite tendency was observed for the non-hydrogen bonded C-O groups of proteins. Secondly, when comparing adenocarcinoma with tumor associated stroma, the discriminating spectral features appeared to be contributed principally from collagen features.

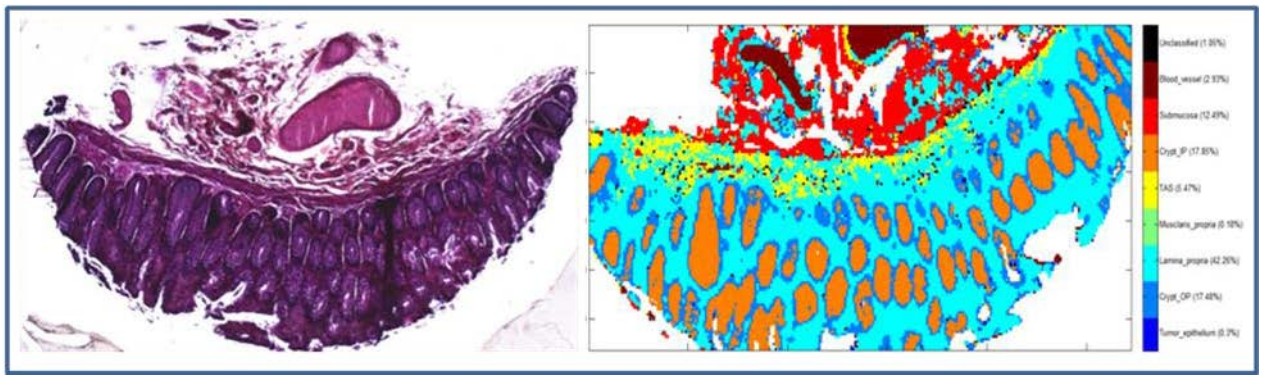
DISCUSSION:

Spectral histopathology based on IR imaging has been carried out to develop an innovative label-free diagnostic methodology directly on FFPE tissue arrays embedded in an agarose matrix without any chemical pre-treatments. EMSC that has been initially developed to separate light scattering effects from light absorbance effects, has also been used for accomplishing neutralization of paraffin contributions in IR spectral analysis (24-27). In this study, both paraffin and agarose interferences on the IR spectral images have been neutralized digitally without the use of any chemicals, using an improved EMSC algorithm (Article 1 of Chapter III).

Clustering: K-means clustering provided a rapid way to classify the IR spectral images into their constituent histological classes in comparison to the chemically stained conventional images. While the non-tumoral colonic tissues were characterized by well-differentiated architecture with both inner and the outer cryptal parts clearly distinguishable together with the connective tissue, the malignant tissues which were all of the advanced colonic cancer types, were characterized by the loss of differentiation of the normal colonic glands with no visible lumen; and presence of stromal tissue. The digital staining of each k-means cluster formed the basis for spectral marker assignment comprising the malignant colonic characteristics, along with the normal tissue features, at different organizational levels of the colonic wall. Based on this spectral database from as little as 11 % of the samples, a prediction model was trained for automatic detection of malignancy in unknown specimens independently of conventional histopathology.

Prediction: Some of the earlier IR imaging studies have tested prediction algorithms on different tissue types (19, 20). However, the number of spectra used for constructing the model was limited compromising the robustness of the model. In our study, the high image acquisition parameters applied to tissue arrays (3 mm diameter) constituted a huge bank of 86802 spectra in the model, representative of the real biochemical signatures of distinct colonic structures, making the model highly robust. Only one such IR imaging study on prostate tissues has used such a robust model for prediction on unknown tissues (28). In this study, 8 classes were included that described the colonic tissue organization in non-tumoral and tumoral samples. Some of these histological structures may share certain similar molecular constituents with other histological classes present in the model (tumor and tumor

associated stroma), or not present in the model (muscularis mucosa and tumor associated stroma). The spectral proximity arising from this leads to misclassification between such classes as shown in the figure SI 4 concerning the muscularis mucosa (visible in the HPS image) which is identified as tumor associated stroma (SI 1 sample # 27). It has to be noted that there was no class for the muscularis mucosa in the model. This attribution can be presumed to have arisen from the residual normal muscularis mucosa signatures present in the tumor associated stroma from which the corresponding class was constructed in the prediction model. This prediction error appeared predominantly in non-tumoral samples where there is an intact muscularis mucosa. Despite these misclassifications, an overall high correlation between the predicted spectral classes and the corresponding histological structures is observed in the confusion matrix.



Supplementary information 4: Confusion between muscularis mucosa and stroma.

The sample is a non-tumoral colonic tissue in which the thin layer of muscularis mucosa is identified as tumor associated stroma by the prediction model seen as yellow pixels (SI 1 sample # 27). The HPS image is at 5X magnification.

External validation: The remaining 89 % of IR spectral images were identified by the prediction model without any a priori knowledge on their histopathology (external validation). These blind samples constituted a huge number of 3620287 spectra that were scanned and annotated by the automated computer trained prediction algorithm. The diagnosis was confirmed by an expert pathologist by using the conventional histological images based on which a 100 % accuracy of the prediction model was obtained for tumor diagnosis. This high sensitivity after scanning such a huge number of unknown spectra

signifies the potential of the current methodology as a diagnostic tool. The prediction analysis also facilitated simultaneously some important malignancy associated features.

Tumor budding:

The phenomenon of tumor budding is of crucial clinical importance in colorectal cancers since it has been shown to be a strong adverse prognostic marker (29). As such, studies have correlated its occurrence with aggressiveness and lymph node metastasis (30). In this study, the prediction model facilitated the identification of tumor budding in a stroma-dominant environment in an automated manner. This rapid and selective detection of small clusters of isolated tumor cells in an abundant stroma environment demonstrates the sensitivity and the applicability of the methodology avoiding the need of any histological or immunological markers. This envisages an important prospect since the tumor de-differentiation in the form of budding is being acknowledged as a key component in the metastatic process even in well- and moderately differentiated tumors (31, 32). At the same time, the color code based selective staining of the epithelial counter parts in the same tissue shows the discriminatory ability and the biomolecular specificity of this methodology.

Spectral Analysis:

The IR spectral region from 1000 cm^{-1} to 1300 cm^{-1} has been reported to carry important biochemical vibrations implicated in colon cancers and have been used for differentiating the malignant tissues from their normal counterparts (33, 34). In this study, the most discriminant spectral wavenumbers were associated with relatively decreased intensities of symmetric and asymmetric PO_2^- vibrations of the nucleic acids in the tumoral epithelium when compared to the non-tumoral tissues. On contrary to the expected increased nucleic acid intensities as shown in several studies, these spectral changes corresponding to the biochemical alterations corroborate with some of the previous studies on colon cancers where the nucleic acid intensities were shown to be reduced in malignant conditions (23, 35). It may be likely that the spectral changes involving nucleic acids are small in moderately differentiated tumors when compared to normal colonic epithelial cells which themselves are highly proliferative in nature. One study has stated that decreased phosphate content in malignant colon tissues may be due to decrease in carbohydrate content (36), which in our study was also indicated by the

relatively less intense C-O stretching vibration corresponding to carbohydrates in the tumoral tissue than the normal. At the same time, the relative intensities of H-bonded C-O vibrations of proteins were observed to be more pronounced in the normal epithelium than the tumoral, while the non-H-bonded C-O bond vibrations were more pronounced in the tumor. These changes may be indicative of the molecular alterations associated with the amino acid side chains concerning tyrosine, serine and threonine (2, 23, 36, and 39). The molecular changes involving adenocarcinoma and tumor-associated stroma appear principally due to collagen features.

Tissue inflammation influences the model specificity:

In 12 out of 26 samples histologically described as non-tumoral (SI 1 sample # LF); tumoral characteristics (over 4 % of pixels) were observed either regionally clustered or dispersed in the lamina propria. The HPS images gave insight into the regionally clustered tumor pixels as corresponding to lymphoid follicles in the colonic tissue. These structures showed spectral signatures close to the tumor group relative to the other classes. However, the tumor pixels dispersed in the lamina propria could not be accounted for as no visible correspondence between them and any histological feature could be found in the HPS images. Since these tissues showed high inflammatory infiltration, immuno-staining for T-lymphocytes (CD 3), B-lymphocytes (CD 20) and macrophages (KP 1) was performed to verify if the dispersed pixels corresponded to the inflammatory cells. The positive staining indicated that these pixels indeed corresponded mainly to interstitial T-lymphocytes as representatively shown in the figure 8A (SI 1 sample # 32 and 31). In parallel, the B-lymphocytes were seen assembled in lymph follicles. Non-tumoral tissues without any marked inflammation as confirmed by the IHC showed no tumor pixels in the IR spectral images (figure 8B). Since the model did not take into account inflammatory conditions (because of the tissue complexity arising from polymorphisms of the inflammatory infiltrates in colon cancers: polymorph predominant, mononuclear predominant, mixed or rich in lymphoid follicles, and the difficulty to have a representative spectral signature), these features were attributed to the spectrally nearest class which turned out to be the tumor class.

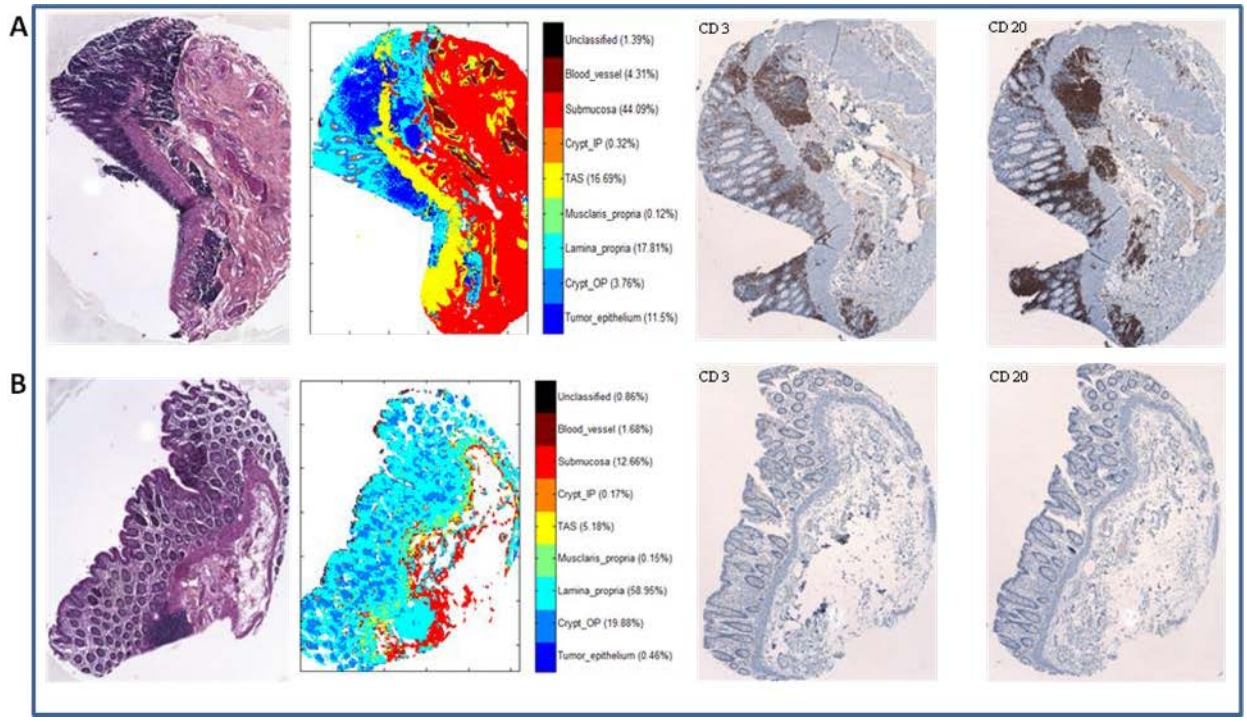


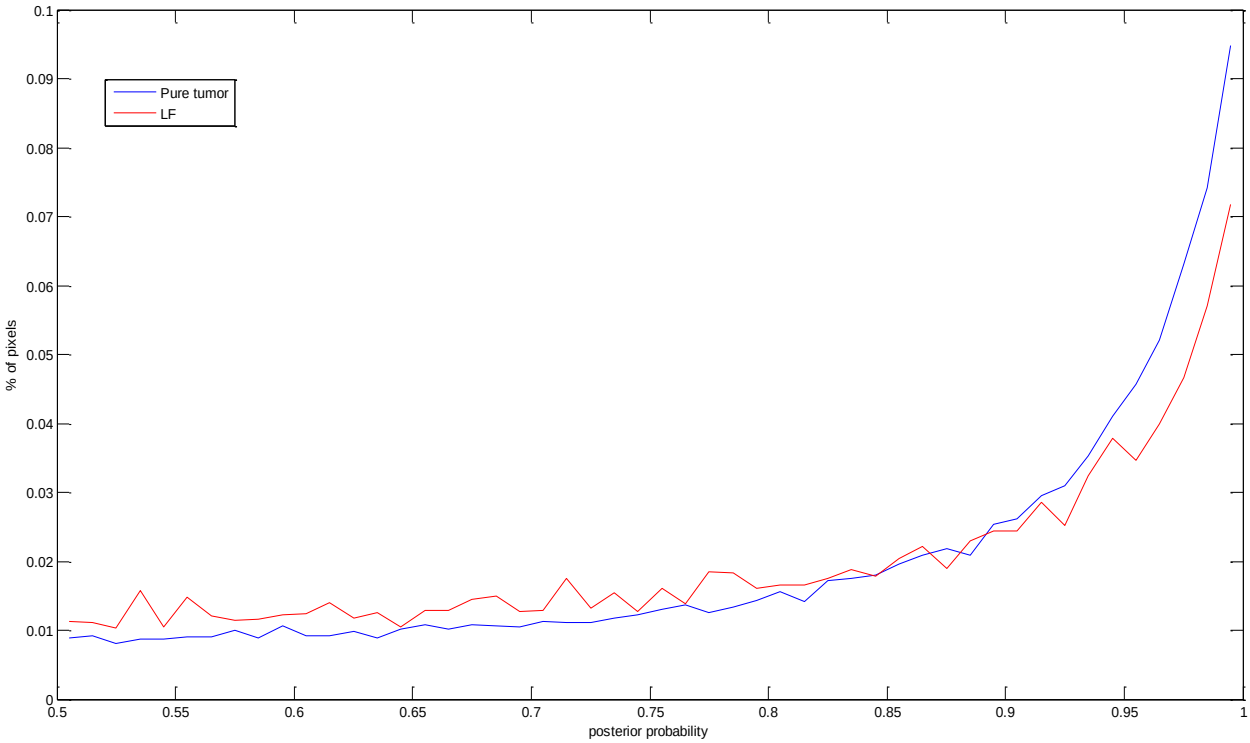
Figure 8: Influence of tissue inflammation on the prediction model.

Left to right: 1. HPS stained colon tissues (SI 1 sample # 32 and 31), 2. LDA predicted images, 3. Immuno-stained images for CD3 marker and 4. 3. Immuno-stained images for CD20 marker.

A is a non-tumoral colonic tissue with typical normal glands. The mucosa is partially populated by lymphoid follicle as seen in the HPS image. The prediction model identified the regions in the mucosa as tumor. Immuno-staining for CD3 and CD 20 markers revealed that the tumor class in the predicted images actually corresponded to inflammatory signatures. B is another non-tumoral tissue which is negative for CD 3 and CD 20 indicating absence of inflammatory signature, and is predicted as as verified by the immuno-staining which shows no positive staining for CD 3 and CD 20. The HPS and the IHC images are at 5X magnification.

A recent IR imaging study on cervical cancer tissues also quoted the influence of inflammatory signatures on the prediction model sensitivity and specificity (37). To have a broader insight into this aspect, we further looked at the spectral class attribution threshold for the tumor class. It turned out that the majority of the spectra corresponding to the inflammatory signatures have lesser threshold values compared to the tumor (SI 5).

Altogether, the IR signatures from the inflammatory regions appeared to class spectrally closer to tumor than other classes of the prediction model indicating an intermediate stage between normal and malignant condition, as was shown in an earlier study (38).



Supplementary information 5: Histogram for tumor pixel attribution in tumoral and non-tumoral sample.

The current work of IR spectral imaging on colon tissues provides automated diagnosis of malignancy on unknown samples. Various diagnostic features associated with malignancy which provides complementary information are also characterized. Important features such as tumor budding, tumor-stroma association are dealt with in a non-destructive and label-free manner. The analysis of such a large spectral database makes the study all the more representative. All these features have never been dealt together in colon cancer diagnosis using IR spectral imaging of paraffinized tissues in any of the previous studies. IR spectral imaging presents an optimistic overture for cancer knowledge in modern histopathology.

The current prediction model representing the important histological features of a colonic tissue certainly holds aspects for amelioration. The spectral attribution identified the

inflammatory signatures classed close to the tumor. Since these specific biochemical signatures were picked up by the model, the inflammatory infiltration, which pose risk of developing into cancers, could be incorporated into the model for an automated evaluation and direct diagnostic approach for inflammatory diseases. Aspects like genotype specific tumoral signatures and their treatment response sensibility unknown till now could open a new additional classification. Further, an automated quantification can be achieved for features like amount of tumor presence, or the amount of tumor budding, only limit being the use of adjacent tissue sections which may present slight variations from the reference tissue.

Conclusion:

The IR spectral imaging combined with multivariate statistical analyses appears as an optimistic diagnostic approach for colon cancers in complement to conventional histopathology. This innovative imaging approach enabled direct analysis of paraffinized tissue arrays and, via the employment of mathematical deparaffinization the need for chemical pretreatments was reduced. The prediction model permitted identification of unknown samples with a very high sensitivity, while the false positive prediction in the non-tumoral samples has put forth the influence of the inflammatory component. This very large scale spectral data base analyzed both in terms of training and validation shows the potentials of the IR spectral imaging methodology for automated diagnostic purposes. Moreover, it eliminated the need for sample staining and a priori knowledge of the sample to be analyzed. These optimistic results open a new way for developing spectral biomarkers and libraries which could be used, in complement to conventional histopathology, for early diagnosis, and also potentially for prognosis and theranostics of cancers.

References:

1. J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," Int J Cancer 127(12), 2893-2917 (2010)

2. C. Conti et al, Journal of Molecular Structure, 881, 46–51, (2008)

3. H. Miyoshi, M. Oka, K. Sugi, O. Saitob, K. Katsu et al. Accuracy of detection of colorectal neoplasia using an immunochemical occult blood test in symptomatic referred patients: Comparison of retrospective and prospective studies. Intern Med. 2000; 39:701–6.

4. T. J. Zuber. Flexible sigmoidoscopy. Am Fam Physician. 2001; 63: 1375–92.

5. D. K. Rex. Colon tumors and colonoscopy. Endoscopy, 2000; 32:874–83.

6. A. Tfayli, O. Piot, A. Durlach, P. Bernard and M. Manfait, "Discriminating nevus and melanoma on paraffin-embedded skin biopsies using FTIR microspectroscopy," Biochim Biophys Acta 1724(3), 262-269 (2005)

7. B. Bird, M. Miljkovic, S. Remiszewski, A. Akalin A, Kon M, et al. Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer. Laboratory Investigation, 2012; 1–16.

8. H. Fabian, N. A. Thi, M. Eiden, P. Lasch, J. Schmitt and D. Naumann, "Diagnosing benign and malignant lesions in breast tissue sections by using IR-microspectroscopy," Biochim Biophys Acta 1758(7), 874-882 (2006)

9. W. Steller, J. Einenkel, L. C. Horn, U. D. Braumann, H. Binder, R. Salzer and C. Krafft, "Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging," Anal Bioanal Chem 384(1), 145-154 (2006)

10. A. Travo, O. Piot, R. Wolthuis, C. Gobinet, M. Manfait, J. Bara, M. E. Forgue-Lafitte and P. Jeannesson, "IR spectral imaging of secreted mucus: a promising new tool for the histopathological recognition of human colonic adenocarcinomas," Histopathology 56(7), 921-931 (2010)

11. M. J. Nasse, M. J. Walsh, E. C. Mattson, R. Reininger, A. Kajdacsy-Balla, V. Macias, R. Bhargava and C. J. Hirschmugl, "High-resolution Fourier-transform infrared chemical imaging with multiple synchrotron beams," Nat Methods 8(5), 413-416 (2011)

12. M. J. German, A. Hammiche, N. Ragavan, M. J. Tobin, L. J. Cooper, S. S. Matanhelia, A. C. Hindley, C. M. Nicholson, N. J. Fullwood, H. M. Pollock and F. L. Martin, "Infrared spectroscopy with multivariate analysis potentially facilitates the segregation of different types of prostate cell," Biophys J 90(10), 3783-3795 (2006)

13. K. Yano, S. Ohoshima, Y. Gotou, K. Kumaido, T. Moriguchi and H. Katayama, "Direct measurement of human lung cancerous and noncancerous tissues by fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool?," Anal Biochem 287(2), 218-225 (2000)

14. T. D. Wang, G. Triadafilopoulos, J. M. Crawford, L. R. Dixon, T. Bhandari, P. Sahbaie, S. Friedland, R. Soetikno and C. H. Contag, "Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy," Proc Natl Acad Sci U S A 104(40), 15864-15869 (2007)

15. X. Zhang, Y. Xu, Y. Zhang, L. Wang, C. Hou, X. Zhou, X. Ling and Z. Xu, "Intraoperative Detection of Thyroid Carcinoma by Fourier Transform Infrared Spectrometry," J Surg Res (2010)

16. C. Krafft, S. B. Sobottka, K. D. Geiger, G. Schackert and R. Salzer, "Classification of malignant gliomas by infrared spectroscopic imaging and linear discriminant analysis," Anal Bioanal Chem 387(5), 1669-1677 (2007)

17. J. Kononen et al. (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nature Medicine 4, 844 – 847

18. R. Wolthuis, A. Travo, C. Nicolet, A. Neuville, M.-P. Gaub, D. Guenot, E. Ly, M. Manfait, P. Jeannesson, O. Piot. Anal. Chem. 80, 8461, 2008

19. M. Khanmohammadi, A. B. Garmarudi, K. Ghasemi, H. K. Jaliseh and A. Kaviani, "Diagnosis of colon cancer by attenuated total reflectance-Fourier transform infrared microspectroscopy and soft independent modeling of class analogy," Med Oncol 26(3), 292-297 (2009)

20. M. Khanmohammadi, A. Bagheri Garmarudi, S. Samani, K. Ghasemi and A. Ashuri, "Application of linear discriminant analysis and Attenuated Total Reflectance Fourier Transform Infrared microspectroscopy for diagnosis of colon cancer," Pathol Oncol Res 17(2), 435-441 (2010)

21. M. Khanmohammadi, M. A. Ansari, A. B. Garmarudi, G. Hassanzadeh and G. Garoosi.

Cancer diagnosis by discrimination between normal and malignant human blood samples using attenuated total reflectance-Fourier transform infrared spectroscopy.

Cancer Invest, 2007, 25, 397–404.

22. E. Gazi, M. Baker, J. Dwyer, N. P. Lockyer, P. Gardner et al. A Correlation of FTIR Spectra Derived from Prostate Cancer Biopsies with Gleason Grade and Tumour Stage.. european urology 50, 2006; 750–761

23. P. Lasch, W. Haensch, D. Naumann and M. Diem, "Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis," Biochim Biophys Acta 1688(2), 176-186 (2004)

24. H. Martens, J. P. Nielsen and S. B. Engelsen, "Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures," Anal Chem 75(3), 394-404 (2003)

25. A. Kohler, C. Kirschner, A. Oust and H. Martens, "Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform infrared microscopy images of cryo-sections of beef loin," Appl Spectrosc 59(6), 707-716 (2005)

26. E. Ly, O. Piot, R. Wolthuis, A. Durlach, P. Bernard and M. Manfait, "Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies," Analyst 133(2), 197-205 (2008)

27. D. Sebiskveradze, V. Vrabie, C. Gobinet, A. Durlach, P. Bernard, E. Ly, M. Manfait, P. Jeannesson and O. Piot, "Automation of an algorithm based on fuzzy clustering for analyzing tumoral heterogeneity in human skin carcinoma tissue sections," Lab Invest 91(5), 799-811 (2011)

28. D. C. Fernandez, R. Bhargava, S. M. Hewitt, et al. Infrared spectroscopic imaging for histopathologic recognition. Nature biotechnology 2005; Volume 23, number 4

29. L.M. Wang,D. Kevans, H. Mulcahy, J. O'Sullivan, D. Fennelly, et al. Tumor Budding is a Strong and Reproducible Prognostic Marker in T3N0 Colorectal Cancer. Am J Surg Pathol 2009; Volume 33, 1

30. H. Kanazawa, H. Mitomi, Y. Nishiyama, I. Kishimoto, N. Fukui, et al. Tumour budding at invasive margins and outcome in colorectal cancer. Colorectal Disease, 2008; 10, 41–47

31. F. Prall. Tumour budding in colorectal carcinoma. Histopathology 2007; 50, 151–162.

32. H. Gabbert. Mechanisms of tumor invasion: evidence from in vivo observations. Cancer Metastasis Rev. 1985; 4:293–309.

33. R. K. Sahu, S. Argov, S. Walfisch, E. Bogomolny, R. Moreh et al. Prediction potential of IR-micro spectroscopy for colon cancer relapse Analyst, 2010; 135, 538–544

34. V. K. Katukuri, J. Hargrove, S. J. Miller, K. Rahal, J. Y. Kao, et al. Detection of colonic inflammation with Fourier transform infrared spectroscopy using a flexible silver halide fiber. Biomedical Optics Express, 2010; 1, 3, 1014

35. B. Rigas, S. Morgello, I. S. Goldman, P. T. Wong. Human colorectal cancers display abnormal Fourier-transform infrared spectra. Proc Natl Acad Sci U S A, 1990; 87(20), 8140-8144.

36. S. Argov, J. Ramesh, A. Salman, I. Sinelnikov, J. Goldstein, et al. Diagnostic potential of Fourier-transform infrared microspectroscopy and advanced computational methods in colon cancer patients. Journal of Biomedical Optics 2002; 7(2)

37. J. Einenkel, U. D. Braumann, W. Steller, H. Binder, L. C. Horn. Suitability of infrared microspectroscopic imaging for histopathology of the uterine cervix. Histopathology 2012; 60, 1084–1098

38. S. Argov, R. K. Sahu, E. Bernshtain, A. Salman, G. Shohat, et al. Inflammatory Bowel Diseases as an Intermediate Stage between Normal and Cancer: A FTIR-Microspectroscopy Approach. Biopolymers, 2004; 75, 384–392

39. P. Wong, H. M. Yazdi. Normal and Malignant Human Colonic Tissues Investigated by Pressure-Tuning FT-IR Spectroscopy. Applied Spectroscopy, 1993; 44, 10

III.6: Supplementary work to spectral histopathology of tissue arrays

III.6. 1: Identification of early biochemical changes in adenomatous tissues; towards tumor grading:

This part constitutes a preliminary study on the prediction outcome of few unknown adenomatous tissue samples. In the previous part of the work, the prediction model based on LDA was able to accurately identify unknown samples as tumoral or non-tumoral accurately. Along with tumor identification, some of the features associated with tumor such as tumor budding, tumor-stroma association were also revealed. The prediction was performed on the tumoral samples which were all moderately differentiated adenocarcinomas.

In order to assess the efficiency of this model for other tumor grades, adenomatous tissue samples showing low-grade and high-grade dysplasia were tested. Interestingly, these samples were identified as tumoral samples as shown in a figure 1.

In general, an adenoma is characterized by different degrees of cell dysplasia, presence of irregular cells with hyperchromatic nuclei, decreased mucosecretion, while the basement membrane and the muscularis mucosa are intact. Thus the IR spectral histopathology indicated that the early molecular changes in the adenomatous samples were picked up by the prediction model that classified them into the tumoral group. This high-sensitive discriminating potential of the prediction model shows good prospects in studying various other tumor grades.

It has to be noted that the prediction model did not contain a separate class for adenomas. However, based on the spectral proximity, the adenomatous signatures which are considered as early molecular changes in progression towards cancer were identified as belonging to the tumor group. In extension to this preliminary work, an LDA model with inclusion of a separate class for adenomas needs to be carried out in order to discriminate between normal, adenoma, and carcinoma tissues automatically.

The identification of the small abnormal characteristics by the model could be a good predictive marker for early diagnosis of colorectal cancers. Large sample population needs to be tested to validate this capability of IR spectral histopathology for early diagnosis, with higher certitude.

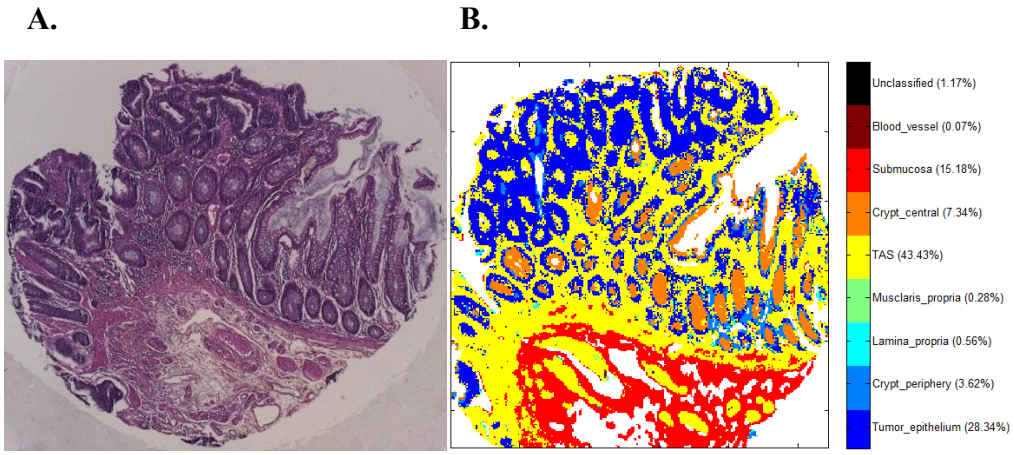


Figure 1: Early detection of tumoral signatures.

A. HPS stained image of an adenomatous colon tissues which shows high grade dysplasia, along with some normal glands, and mucus. B. A LDA predicted image in which these early changes associated in the form of an adenoma are identified. The high-grade dysplastic regions are completely identified as tumoral, while in the seemingly normal glands, few pixels are indicated as tumor. The HPS image is at 5X magnification.

III.6.2: A special note on peri-cryptal fibroblastic sheath (PCFS):

The k-means clustered IR images were highly correlated with the reference HPS stained images (refer to figure 2 of article 3 as an example). The gross histological features were clearly demarcated based on the intrinsic biochemical signatures. With regard to the crypts, each time they were assigned to two clusters: one constituting the crypt central part and the other to the crypt outer part. The central lumen which is filled with mucus was often detected as a third cluster. In the reference histological images the crypt outer part was deeply stained corresponding to the localization of peripheral nuclei in the colonic crypts. In comparison to the reference images, the crypt outer part was clearly demarcated the in the k-means cluster images as well as in the LDA predicted images.

The layer of PCFS corresponding to the basement membrane of the normal colonic epithelium is usually seen surrounding the crypts, specifically in conjunction with the outer nuclear part. The PCFS of the colonic glands is in the scale of 2-5 microns thickness. Hence with a pixel size of 6.25 microns utilized in this study, this region may not be isolated as a separate cluster. However, it could be presumed that some of its signatures could contribute to the class crypt outer part.

The degradation of the PCFS is one of the important signs of tumor invasiveness. Although, the degradation of PCFS was neither detected in the k-means cluster images nor in the LDA predicted images owing limitation of the spatial resolution, the loss of differentiation of the crypt outer part was observed in both the k-means and the predicted images. Since the PCFS is geographically closely associated with the crypt-OP as shown in the IHC image stained for smooth muscle actin (figure 2), the loss of differentiation of the crypt OP can be presumed to be a sign of invasive tumors.

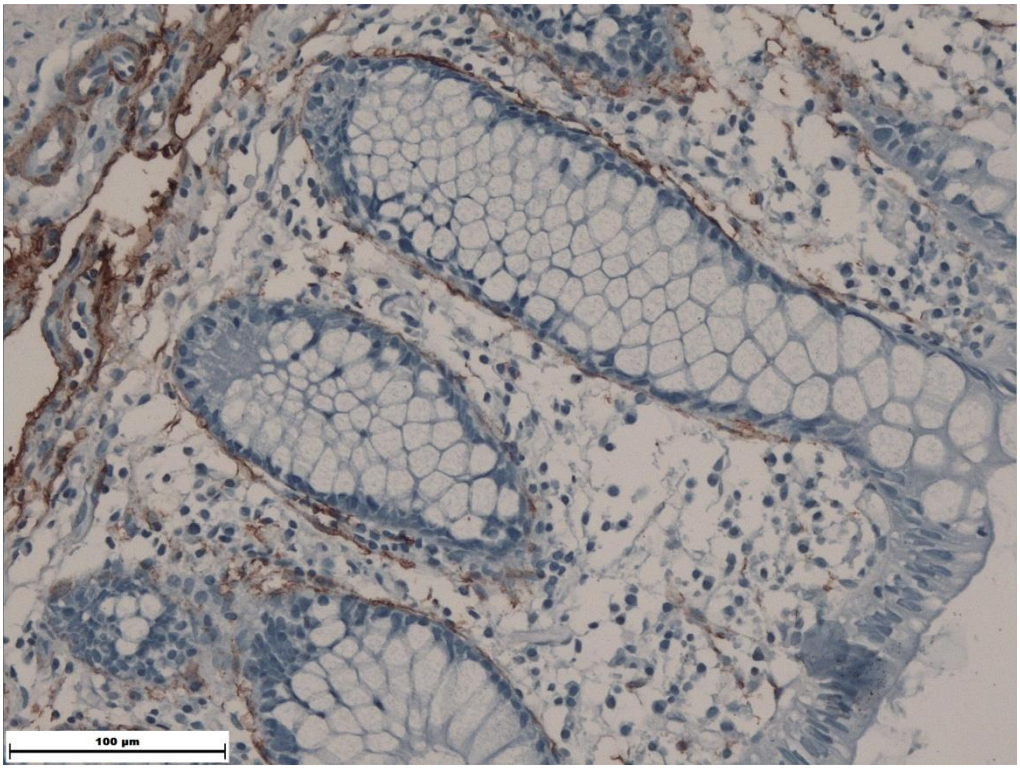


Figure 2: Stable peri-cryptal fibroblastic sheath around the normal colonic glands.

The smooth muscle actin positive staining of PCFS corresponding to the basement membrane around the crypts supporting the epithelial cells. The muscularis mucosa on the top left is also positively stained (positive control). The IHC image is at 20 X magnification).

However, increasing the spatial resolution in the order of few microns would permit to detect individually the PCFS, which could permit access to an important diagnostic marker. Other vibrational spectroscopic approaches such as IR-ATR and Raman imaging which give higher spatially resolved biochemical information could shed more light on this point.

III.7: Conclusions

The biophotonic approach of IR spectral imaging has been applied to colon tissues that permitted characterization of histopathological features, and for automated cancer diagnosis paving the way for the development of spectral histopathology.

In the initial part of the work, a novel concept of spectral barcodes was implemented using IR spectral imaging approach. The barcodes constituted a way to visualize the spectral biomarkers, based on the most discriminant wavenumbers between normal and malignant colonic mucosa. Several discriminant spectral zones were identified which were correlated to different biochemical features of the analyzed tissues in an easy-to-interpret manner.

In the following work, demonstration of the IR spectral imaging methodology applicable to paraffinized tissue arrays has been performed. An important aspect in this work is that paraffinized tissue array stabilized in an agarose matrix could be directly analyzed without any chemical dewaxing thus simplifying the experimental protocol. Since the spectral images from the tissue arrays consisted of interferences originating from paraffin and agarose, a modified EMSC algorithm was developed. This is the first time that EMSC has been implemented to correct the spectral interferences from both paraffin and agarose together. Additionally, using multivariate analysis, complementary information on the changes associated with the biochemical properties between normal and malignant tissues were recovered, in a single measurement and in a label-free manner. Spectral analysis revealed specific profile for mucinous adenocarcinoma that differentiated it from its normal counterpart and other non-mucinous cancer types. Spectral features associated with nucleotides, carbohydrates and proteins were also identified as discriminant. The PCA analysis showed a clear separation between the normal and the tumoral groups while a close association was observed for the tumor and its associated stroma.

Finally, large scale application of potentials of IR spectral imaging was carried out in the following study in order to validate the concept of spectral histopathology. This methodology based on the conjunction of IR imaging and multivariate statistical analysis enabled label-free classification of the colonic tissues into their histological features, and the construction of digitally color-coded spectral images. A prediction model developed from these images revealed the inter-class spectral heterogeneity and proximity of several histological features. The model when applied on unknown tissue samples not only identified the normal and the

tumoral features of colonic tissues, but also revealed other tumor associated features without a priori knowledge of the sample. Malignancy associated features such as tumor budding which are difficult to discern by conventional histopathology were identified. In addition, the tumor-stroma association was also delineated which appeared in spectrally close proximity. In this study, the IR spectral range of 1080-1300 cm^{-1} showed the best prediction outcome. This region was associated with some of the important biomolecules implicated in colon cancers such as nucleotides, mucin, and carbohydrates.

This original approach has permitted to differentiate and detect normal and tumoral tissues of colon based on their intrinsic biochemical characteristics in a non-destructive manner. This novel imaging approach which necessitates no staining or chemical treatment opens a new way for spectral histopathology for automated and objective diagnosis of colon cancers independent of the operator-inherent variability.

At the same time, several applications of this methodology are envisaged in the future.This methodology of IR spectral imaging applied to paraffinized tissue microarrays enables high-throughput, molecular level analysis of large tissue archives. This could permit to construct libraries of spectral biomarkers which could be used in complement to conventional histopathology and also for prognostic and predictive purposes in cancer therapy.

Chapter IV

Complementary modalities for spectral histopathology

Manuscript in preparation

IV.1: Résumé:

L'imagerie spectrale IR constitue une méthode diagnostique prometteuse capable de fournir des informations complémentaires à l'histopathologie conventionnelle. Dans ce chapitre nous abordons des analyses complémentaires pour (1) évaluer l'influence de la résolution spatiale sur les données spectrales tissulaires et (2) évaluer la potentiel de l'imagerie IR conventionnelle au niveau d'un autre pathologie. Les différentes potentialités de cette méthode ont été exploitées dans diverses études. En raison de l'hétérogénéité des tissus biologiques, la taille du pixel représente un facteur important puisqu'elle définit la précision de l'information spectrale acquise.

Dans l'étude des tissu arrays, nous avons utilisé un imageur équipés d'une barrette de 16 détecteurs MCT, chacun donnant une taille de pixel de 6, 25x6, 25 μm^2. Avec ce paramètre, les caractéristiques majeures des tissus sont clairement définies, mais des détails plus fins ne sont pas assez résolus. Nous avons vu dans l'étude précédente que c'était un facteur dans la détection des PCFS.

C'est cette raison que d'autres modalités de spectroscopie vibrationnelles pouvant fournir une meilleure résolution spatiale ont été évaluées. Les imageurs type matrice à plan focal (FPA à 64x64 pixels) permettent d'acquérir des images à une résolution de 4,2 μm/pixel et d'améliorer sensiblement la résolution. L'imagerie infrarouge peut aussi être améliorée en utilisant le mode ATR (Attenuated Total Reflection ou Réflexion Totale Atténuée) qui utilise un cristal ayant un indice de réfraction élevé tel que le germanium. De cette façon, la résolution spatiale peut être améliorée d'un facteur 4 au niveau de chaque pixel. D'autre part, nous avons aussi appliqué l'imagerie Raman qui permet d'accéder à une résolution spatiale de l'ordre du micron et de fournir des images spectrales hautement résolues.

Bien que l'imagerie tissulaire soit possible en utilisant des approches par imageries ATR et Raman, relativement très peu d'études ont été réalisées sur des tissus en comparaison avec l'imagerie IR. Il est à noter que ces deux approches ne se prêtent pas aux tissu arrays de grande taille (3 mm de diamètre) comme nous l'avons utilisé dans l'étude précédente. Nous avons donc effectué une étude comparative entre l'imagerie IR classique, l'imagerie IR-ATR et l'imagerie Raman, sur des zones réduites de coupes de tissus coliques congelés, afin de mettre en évidence les avantages et les limites de ces différentes méthodes. Des facteurs tels

que la résolution de l'image, la finesse des détails tissulaires qui peuvent être résolus et les temps d'acquisition ont été comparés.

Dans un deuxième temps, la méthodologie d'imagerie IR spectrale appliquée aux tissu arrays paraffinés de côlon a été testée sur des tissus mammaires, afin de mettre en évidence des marqueurs spectraux pour l'identification des cancers du sein. Ce travail préliminaire a été effectué directement sur des tissus mammaires paraffinés. Enfin, l'application potentielle de ces approches au sein de laboratoires cliniques et les aspects à améliorer sont discutés.

IV.2: Summary:

The following chapter constitutes a complementary work to histopathology undertaken using different imaging modalities with the aim (1) to evaluate the influence of spatial resolution on the spectral data obtained from the tissues and, (2) to evaluate the potential of conventional IR imaging applied to another pathology.

Due to the inherent heterogeneity of the biological tissues, the measured pixel size is an important factor that defines the precision of the acquired information. In the IR imaging approach of the tissue arrays, a 16 element MCT detector with the possible pixel size of 6.25x6.25 μm^2 was used. At this size, although the major tissue features are demarcated, finer details are not clearly resolved. In the preceding work, this was one of the factors that limited the detection of PCFS.

In this regard, other modalities of vibrational spectroscopy that can provide improved spatial resolution have been tested. The imagers with Focal Plane Array (FPA with 64x64 pixels) detectors can provide a resolution of 4.2 μm/pixel. The IR imaging can also be improved using an ATR (Attenuated Total Reflection) mode which uses a high refractive index crystal such as germanium which can provide a four-fold higher resolution at each pixel. Complementarily, Raman imaging was also applied which can also provide highly resolved spectral images in the order of microns.

Although imaging is possible using ATR and Raman approaches, there have been relatively very few studies that employed them for tissue imaging in comparison to IR imaging. It has to be noted that the applicability of these two approaches is limited by the large size of the tissue arrays (3mm diameter) as was used in the preceding work. Therefore in order to look at the feasibility of other imaging methods, we performed a comparative study using conventional IR imaging, ATR-IR imaging and Raman imaging on limited zones of frozen colonic tissue sections, in order to evaluate the various parameters involved, their advantages and limitations. Factors such as image resolution, the finer tissue details that can be resolved, the time constraints have been compared to have a clear view of the different imaging modalities available in the common laboratory equipments. In an independent study, the IR spectral imaging methodology that has been tested on paraffinized tissue arrays has been tested on breast tissues in order to develop spectral markers for the identification of breast cancers.

This preliminary work was performed directly on paraffinized breast tissues. Finally, the potential applicability of these approaches in a clinical scenario and the aspects to be improved are discussed.

IV.3: Introduction:

Vibrational spectroscopic approaches comprising IR absorption and Raman scattering have been regarded as one of the important potential candidates for diagnosis of various cancers since they provide biochemical information within cells and tissues (Bhargava, 2007). The IR spectroscopy works on the principal of absorption of light by IR active biomolecules. The absorptions are measured in the form of a spectrum which provides the biochemical fingerprint of cells and tissues in a label-free manner, and provide insight into the structural organization of biological systems (Ellis, 2006). Complementary to this and in another physical phenomenon, when a beam of monochromatic light is incident on a molecule, most photons are elastically scattered. These elastically scattered photons have the same energy and therefore same wavelength as that of the initial photons in a phenomenon known as Rayleigh scatter. A small fraction of photons from these are scattered at different energies from that of the incident, resulting in inelastic scattering known as the Raman effect. This shift in the energy of the inelastically scattered photons due to Raman active molecules plotted against the intensity of scattered light gives the Raman spectrum that is routinely used to gain insight into the sample molecular composition.

Histopathology is the current gold standard method of cancer diagnosis which is based on microscopic examination of tissue morphological features (Fernandez, 2005). If combined with vibrational spectroscopic approaches, important information based on the tissue biochemistry under different conditions can be obtained. As such the capabilities of these vibrational spectroscopic techniques have been exploited to study normal and cancerous tissue states with diagnostic significance (Wang, 2007; The, 2008).

IR and Raman spectroscopic imaging methods enable label-free visualization of the tissue structural features with a spatial distribution of the molecular contents where each pixel harbors the full spectral information (Lasch, 2002). IR imaging can be performed in the conventional transmission mode (IR-T) in which the IR light is absorbed by a thin tissue section in its path, transflection mode (IR-TF) in which the IR light is transflected by a reflecting surface on which the tissue is placed or the less conventional ATR-IR mode. Each of these techniques poses advantages in certain aspects and limitations in the other. Table I gives a comparative overview of various parameters involved in IR and Raman spectroscopies. In ATR-IR, an IR beam is internally reflected, producing an evanescent wave, onto a high refractive index internal reflection element (like Germanium or Diamond

crystals) that is in contact with the sample. Compared to conventional IR imaging, this feature gives fourfold higher spatial resolution by reducing the diameter of the light focused onto the sample.

Table I: A comparative overview of some important parameters involved in IR and Raman spectroscopies.

	IR	ATR-IR	Raman
Principle	Absorption	Total internal reflection	Inelastic scattering
Nature of histological samples	Frozen, paraffinized	Fresh, frozen, paraffinized	Fresh, frozen
Measurement in water	-	+	+++
Sample preparation	+++	++	+++
Spatial resolution	+	++	+++
Measurement time	+++	++	+
Confocality	-	-	++
Singnal-noise ratio	+++	+++	++
Depth analysis			++
Mode of acquisition	Transmission, Transflection	Attenuated reflection	Inelastic scattering
Qualitative	++	++	+++
Quantitative	+++	++	+
Degree of expertise	+++	++	++

+++ - Highest grading
+ - Least grading
- - Not applicable

The choice to exploit any of these approaches to obtain the requisite information from a sample is based on the compromise between spatial resolution and the spectral quality, together with the time constraints. A comparative analysis of the spectral images acquired using these approaches, can provide an overview of the applicability of these modalities for prospective studies related to tissue imaging. In this perspective, imaging based on IR-T, ATR-IR and Raman has been performed on colon tissue sections. The spectral images acquired were partitioned using k-means clustering method into their constituent structural features and compared to the adjacent HE stained images that served as the morphological reference. Various parameters involved in these imaging methods, their advantages, and their limitations are discussed.

IV.4: Materials and methods:

Sample preparation:

Two 10-micron thick, unstained non-tumoral frozen colon tissue sections were mounted onto a calcium fluoride (CaF_2) window for imaging measurements using the different imaging approaches. The window quality was chosen so that it is compatible for all the three imaging approaches. An adjacent 10-micron thick tissue section placed on glass was HE stained and used for histopathological recognition and comparison. While the same region of the same tissue section was used for IR-T and ATR-IR measurements, the Raman imaging was performed on the second sample. Each time, a normal colonic crypt was analyzed for comparing the three imaging approaches.

IR-transmission imaging: The IR imaging system (Spotlight 300, Perkin Elmer, France) was equipped with nitrogen-cooled 16-element MCT detector calibrated for imaging in transmission mode. The image acquisition was carried in the mid-IR range from 750 cm^{-1} to 4000 cm^{-1} at 4 cm^{-1} spectral resolution averaged to 16 accumulations using a 6.25x6.25 μm^2 pixel size. Each time, prior to image acquisition, energy was optimized and a reference spectrum from the bare CaF_2 window was recorded keeping the same parameters constant as those of the tissue image. This served as a background spectrum which was subsequently subtracted from the dataset.

IR-ATR imaging: The same imaging set up was calibrated into ATR mode into which the sample mounted on the CaF_2 window was placed on the ATR sample holder. The ATR set up consisted of a Germanium crystal (600 μm diameter) for the internal reflection of the IR beam. The crystal is put into contact with the sample and a maximum area of 500x500 μm^2 can be scanned with the movement of the XY stage. Due to the high refractive index of the Germanium crystal (n=4.0), ATR further made it possible to use 4 times higher spatial resolution at 1.56x1.56 μm^2 pixel size. Prior to image acquisition, the crystal background was acquired each time on the bare window. The same acquisition conditions and the same multi-element detector as with the IR-T technique were used. The imaging system was continuously purged with dry air during the image acquisition. A lay out of the ATR-IR imaging principal and its set up in an bench top instrument is shown in figure 1.

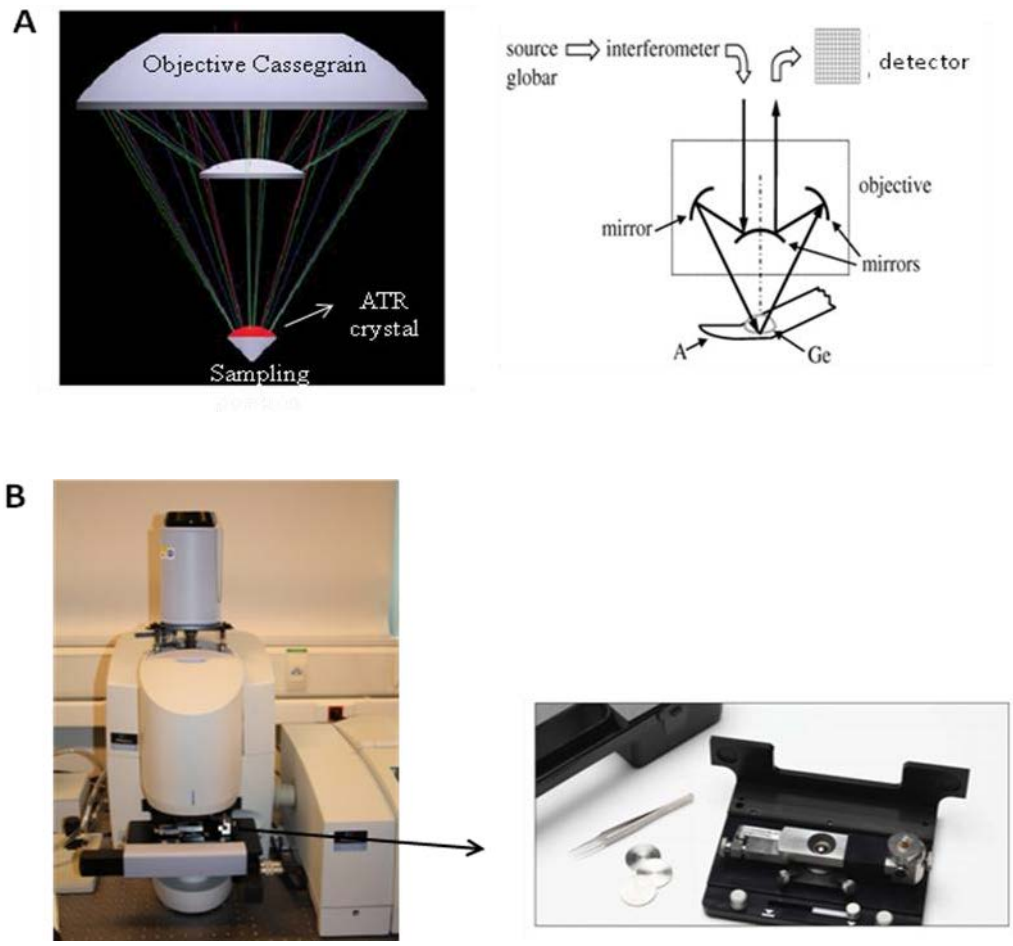
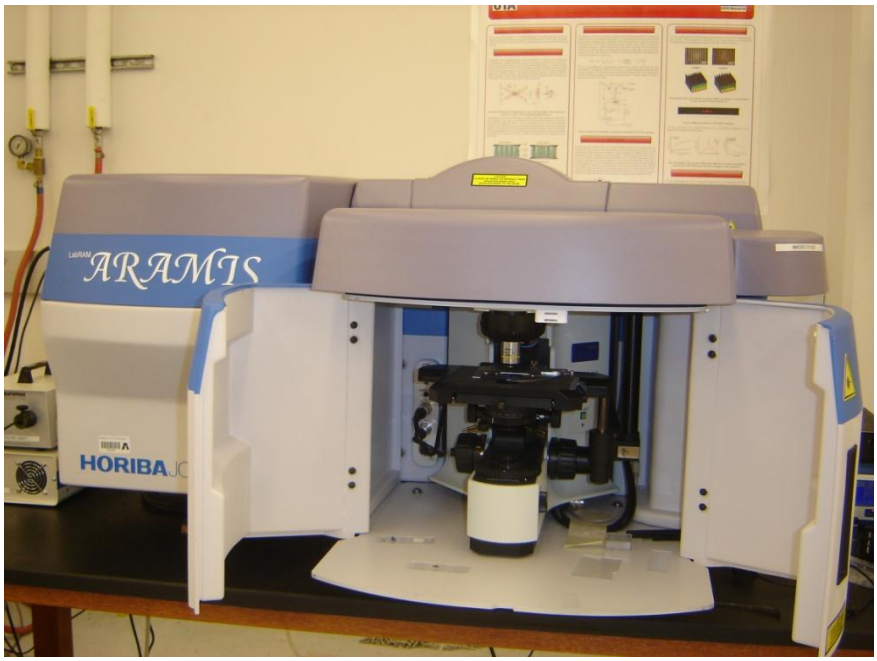


Figure 1: ATR-IR imaging.

(A) Layout of the ATR crystal under an FTIR imaging system's reflective Cassegrain objective, and (B) The ATR-IR setup on a bench-top spectrometer.

Raman imaging: Raman spectral images were recorded with Raman micro-spectrometer (Aramis, Horiba Jobin Yvon, France) coupled to a microscope (Olympus, BX 41, France) shown in figure 2. An excitation source of 532 nm laser (Type Solid, Quantum Ventus, France) was used and the laser was focussed onto the sample using a 100X long focal objective (Olympus, France) with a numerical aperture of 0.9 and the power on the sample was kept around 30 mW. The sample was localized using a white light image captured by the screen image recorder camera attached to the microscope. A step size of 4 µm and a double acquisition time of 10 seconds/spectrum were used for imaging. Prior to image acquisition, Raman shift calibration was performed using silicon standards. Scattered light was collected by the same objective which was then analyzed by the spectrometer equipped with Pelletier-

cooled charge-coupled device (CCD) detector at -70 C. The spectral window ranged from 400 cm^{-1} to 1750 cm^{-1} and was collected in a single step.



.

Figure 2: Horiba Jobin Yvon LabRAM Aramis Raman Spectrometer

Pre-processing:

The IR spectral images (transmission and ATR mode) were corrected for atmospheric absorptions of water vapor and carbon-dioxide by the built-in software of Perkin Elmer Spotlight. Further pre-processing and analysis of all the spectral images were performed using in-hose algorithms written in Matlab 7.2 (The Mathworks, Natick, MA). Extended multiplicative signal correction (EMSC) was used to eliminate the spectra with low signal to noise ratio from the data set. EMSC was employed by performing a fitting on the average spectrum of the data set. To preserve the most eligible spectra for the analysis as many outliers as possible were removed that did not fit well to the average spectrum. EMSC also corrected the spectra for baseline followed by normalization. Pre-processing, processing and analysis of the IR spectra were carried out on spectral images in the absorption range of 900-1800 cm^{-1}. The Raman spectral image was initially corrected for baseline. Further, EMSC

was employed to eliminate the spectra with low signal to noise ratio, followed by normalization in the spectral range of 400 cm^{-1} - 1750 cm^{-1}.

Processing of the spectral data:

Each spectral image of the colonic tissue acquired via IR-T, IR-ATR and Raman imaging was subjected to k-means clustering for classification into their respective histological classes. K-means clustering is an unsupervised non-hierarchical clustering method (Lasch, 2004) that partitions the spectra into pre-assigned number of clusters using randomly selected cluster centres as detailed in the chapter 3. This method enabled to spectrally identify the histological features of the analyzed tissues in comparison to adjacent HE stained section that served as the morphological reference.

IV.5: Results:

The application of EMSC to the spectral data eliminated the outliers with low fitting to the average spectrum. K-means clustering of the pre-processed spectral images of the colonic tissues segregated the spectra into clusters representative of the biochemical and hence their structural components. The k-means images showed different degree of correlation with the HE stained images for each of the imaging method considered and represented the colonic histology that could be correlated to the tissue biochemical composition. Figure 3 show the k-means cluster images of the normal colonic tissues imaged using the three different imaging modalities and compared to adjacent HE stained image (figure 3A).

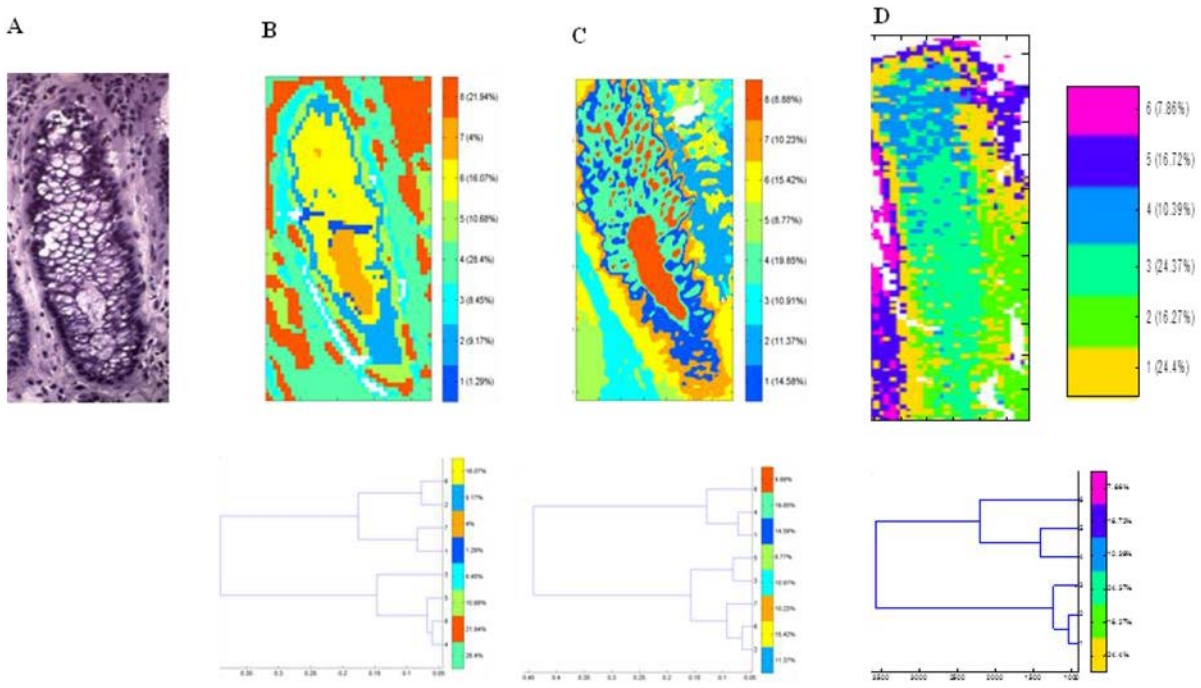


Figure 3: K-means clustering results.

K-means cluster images (with respective dendrogram calculated from their cluster centroids) of the normal colonic crypt imaged using IR-transmission (B), ATR-IR (C) and Raman (D) compared to the HE stained image (A). The HE image is at 20X magnification.

Image B is classified using 8 clusters. The informative clusters are represented as follows: Cluster 2 and 6 - cellular region, cluster 7 - mucus filled lumen of crypt, cluster 5 – probably the peri-cryptal fibroblastic sheath in conjunction with the outer region of the crypt. Clusters 3, 4 and 8 are the surrounding connective tissue

Image C is classified using 8 clusters and are represented as follows: Cluster 1 and 4 - cellular region, cluster 8 - mucus filled lumen of crypt, cluster 7 – probably the peri-cryptal fibroblastic sheath in conjunction with the outer region of the crypt. Clusters 2, 3, 5 and 6 are surrounding connective tissue plus adjacent crypts

Image D is classified using 6 clusters and are represented as follows: Cluster 1, 2, 3 and 4 - cryptal region, cluster 5 - probably the peri-cryptal fibroblastic sheath in conjunction with the outer region of the crypt.

The IR-transmission spectral image of normal tissue (figure 3B) clustered using 8 clusters partitioned into 4 clusters representing cryptal regions and 4 clusters representing the regions external to crypts. The informative clusters within the crypts were clusters 2 and 6 that could be attributed to the outer and the inner regions of the crypts respectively, containing different kinds of epithelial cells (mucus producing goblet cells, paneth cells, endocrinal cells, stem cells., etc), and cluster 7 represented the mucus filled central lumen of the crypts. In extra-cryptal regions, cluster 5 could be probably attributed to the peri-cryptal fibroblastic sheath (PCFS) corresponding to the basement membrane, in conjunction with the outer region of the crypt, while clusters 3, 4 and 8 are attributed to the surrounding connective tissue (lamina propria) in which the crypts are embedded.

Similarly classification of the ATR-IR spectral image of normal tissue (figure 3C) attributed clusters 1, 4 and 8 to the inner and the outer regions of the crypt out of which 1 and 4 could be attributed to the epithelial cells and 8 to the mucus filled central lumen of the crypts. Attribution of certain cryptal regions, exterior to the lumen, to cluster 8 could be from the mucus filled goblet cells which could be detected owing to the higher spatial resolution achieved by ATR-IR imaging. Extra cryptal regions were represented by clusters 2, 3, 5, 6, 7 out of which cluster 7 could be attributed to PCFS in conjunction with the outer region of the crypt which appears to be more resolved, and the rest to the surrounding connective tissue together with small proportions of the adjacent crypts. For the Raman spectral image of the normal tissue (figure 3D) that was classified using 6 clusters (increasing the number of clusters to 8, did not provide any further histological information), clusters 1, 2, 3 and 4 could be attributed to the cryptal regions while the cluster 5 probably to the to the PCFS in conjunction with the outer region of the crypt. It must be noted that images B and C were acquired on the same crypt while image D was from another tissue sample. Imaging using different modalities and multivariate analysis provided digitally stained maps of normal colonic crypts which were classified into their different intrinsic constituents. A comparative evaluation of the time constraints in this imaging study is presented in table II.

Table II: An interpolated comparison of time constraints involved in IR and Raman spectroscopies.

	IR-transmission	ATR-IR	Raman
Pixel size (µm)	6.25	1.56	4
Image size measured (X*Y)	275 x 444 µm	214 x 398 µm	135 x 458 µm
No. of points (X*Y)	45 x 72	138 x 256	34 x 115
Acquisition time / image	40 minutes	6 hours	22 hrs
Interpolated time for 100 x 100 µm image	3.16 minutes	42 minutes	213 minutes

IV.6: Discussion:

The choice to implement an imaging approach for obtaining the requisite information from a tissue sample is importantly based on the compromise between spatial resolution and the spectral resolution, together with the time constraints. Imaging of tissues based on conventional IR is well-documented (Ly, 2008; Fabian, 2006; Steller, 2006; Travo, 2010; Nasse, 2011; Yano, 2000; Krafft, 2007), and although imaging is possible with Raman (Larraona-Puy, 2009; Krafft, 2007, Beljebbar, 2009) and ATR-IR (Colley, 2004, Heather, 2010) modes, very few attempts have been made to exploit this possibility especially ATR-IR, in comparison to conventional IR imaging. Most studies on tissues using Raman (Kanter, 2009, Ly, 2010) and ATR-IR (Khanmohammadi, 2009; Zhang, 2010; Khanmohammadi, 2010) have been employed in point by point mapping mode on various tissues like cervix, skin, colon, thyroid, etc. A comparative imaging study was carried out on colonic tissues in order to look into various parameters involved, their conveniences and limitations. The acquired spectral images were clustered into their respective classes representing the histological organization of the colonic tissue. The normal colonic tissue consisted of crypts which constitute different types of epithelial cell populations. They include the

undifferentiated stem cells that continuously replace other intestinal epithelial cells; the goblet cells specialized for mucus secretion; paneth cells which secrete anti-bacterial proteins like lysozymal enzymes. Absorptive cells or enterocytes are the other predominant epithelial cell types. Surrounding the crypt is the peri-cryptal fibroblastic sheath (PCFS) that acts as an interface epithelia from the underlying connective tissue.

Out of the spectral images acquired using the three imaging methods, the ATR-IR appeared to have the advantage of achieving higher spatial resolution (due to the high refractive index of the crystal) at feasible measurement times. IR-T images showed complementary information in lesser time than ATR-IR, but are less resolved. Although, there was similar histological information from certain classes of crypts in both IR-transmission and ATR-IR imaging modes, ATR-IR provided much more specific information in regard to certain classes. As an example, cluster 8 of figure 3C is attributed to mucus filled central lumen together with certain pixels in the cellular regions of the crypt in the ATR image. The common signature for this class probably indicates the localization of mucinous goblet cells that are known to have basal nucleus and remaining majority of the cytoplasm filled with mucinogen granules. This feature was not observed in IR-T and Raman imaging. It must be emphasized that the Raman image was constructed with a step size of 4 microns and increasing to 1 μm would largely improve the image quality but with longer acquisition times. In another scenario, the outer region of the crypt probably in conjunction with the PCFS (cluster 5 of fig 1B and cluster 7 of figure 3C) although was visible in all the imaging modes, however showed spatially sharp transition from one structure to another in the ATR images.

For Raman imaging, a step size of 4 μm was used in this study. Although Raman spectroscopy enables higher spatial resolution than IR, it takes longer time in achieving this. As can be seen in the table 2, even at 4 μm step size, the time taken for the acquisition of a similar area of the tissue, took considerably longer time. With ATR-IR, similar resolution closer to that of conventional Raman imaging can be achieved in less time due to the multi-element imaging detector. This was evident from the difference in the classification and localization ability of goblet cells in ATR-IR imaging.

The ATR-IR also provides advantages in terms of the refractive index and the optical path length. The change in optics when light path crosses different media with a difference in the refractive index, may give rise to scattering, diffraction, reflection, and dispersion. Such

spectral aberrations that are encountered using IR-T or IR-transflection methods due to long optical path lengths could be eliminated using ATR imaging. In ATR-IR, the penetration depth is reduced to few microns achieved using the germanium internal reflection element, which decreases the optical path lengths, hence only the information from the superficial layers of the tissue can be obtained. Although, ATR provided high-resolution images, the total sample area that can be measured is limited to 500x500 μm^2 per each image acquisition, and takes longer time than conventional IR-T measurements and may leave inevitable tissue damage due to the pressure applied rendering it unexploitable for further analysis. Since ATR-IR imaging involves contact between the crystal and the tissue, its handling and maintaining the sensitive tip of the crystal requires higher technical expertise than conventional IR-T imaging or Raman micro-imaging.

Complementarily, this comparative study of different imaging techniques also puts forward a notion of multimodal approach wherein the capacity of each of the technique can be exploited in combination with the other techniques. As an example, biological tissue sections can be imaged using IR-T which globally provides a rapid scan of the whole tissue region. With the information obtained from this approach, finer details can be investigated using ATR-IR (such as specific regions of the crypts in a colonic tissue). Further, Raman imaging can also be employed on specific regions of the tissue to have complementary information that cannot be accessed using IR imaging methodologies.

Conclusion: Vibrational spectroscopic imaging approaches constitute label free method for spectral histopathology of colon tissue sections. The biochemical information obtained is well-correlated to the histology of the colonic crypts. This study provides a comparative overview of the imaging methodologies that could be adapted for tissue imaging. In comparison, ATR-IR imaging appeared to provide the optimal compromise between the spatial resolution and the time constraints for the images analyzed and would be well adapted for imaging tissue microarrays that are 500 μm in diameter.

IV.7: Characterization of breast tissues using infrared spectral imaging: a preliminary study:

The potentials of IR spectral imaging as a diagnostic tool have been examined with optimistic results on various tissue types (Ly, 2008, Travo 2010, Stellar 2006). Different types of cancers have been studied by this non-destructive method that allowed characterizing various pathological conditions of tissues. The initial part of the present study involved development of a working methodology that could be used to study different characteristic features of colon cancers, and the changes in their spectral profile to provide diagnostically relevant information in cancerous conditions. The methodology was implemented and tested on several paraffinized and frozen colonic tissue samples for automated diagnosis of unknown colonic tissues as presented in the preceding chapters.

In a similar approach, a preliminary study has been undertaken to test the feasibility of IR spectral imaging to other pathology namely the breast. For this paraffinized breast tissues were procured from the Pathology department of the Institute Jean Godinot, Reims. Similar to colon tissues, both normal and cancerous breast tissues were procured, and it was attempted to see if the same methodology could be adapted to other tissue types as well.

The main objectives of the study were therefore to establish a methodology adapted to breast tissues and to characterize the spectral markers associated with breast cancers for rapid and molecular based identification. For this, 32 paraffinized breast tissue samples (16 normal and 16 tumoral) from 16 patients that were examined using conventional histopathology (HE staining) were imaged using IR spectral imaging directly on 10 µm thick sections. The IR imaging system (Spotlight 300, Perkin Elmer, Les Ulys, France) equipped with nitrogen-cooled 16-element MCT detector was used for imaging at 4 cm^{-1} spectral resolutions averaged to 16 accumulations and a spatial resolution of 6.25 µm per pixel.

Initially, the imaging was performed on the paraffinized tissue sections deposited on CaF_2 support. Although the quality of spectra collected was good enough, the need for certain technical adaptations was observed at this stage. Unlike the colon tissue arrays for which the tissue core dimensions were already put in place using standardized protocols, the paraffinized breast tissues obtained were large compared to the tissue array samples. It was more complicated to place such large tissue sections on the CaF_2 support. Alternatively, the tissue sections were placed on much larger IR compatible Kevely supports. The Kevely slides

are similar to standard laboratory glass slides, which are one side coated with a reflecting surface that permits IR imaging in the reflection mode. In principle, the Perkin Elmer Spotlight 300 imaging system can be adapted to transmission (on IR transparent windows such as CaF_2) or transflection measurements (on low e-slides such as Kevley slides, Kevley Technologies, USA). In reflection mode, the IR radiation entering the thin tissue section placed on the reflecting slide is reflected back after striking it. Hence, the light travels twice through the tissue, once while entering and then after reflecting as such this mode is also called as double transmission or simply transflection. Advantages of the Kevley slides are that they are transparent to visible light, they are cheaper than transmission substrates, and they can be stained for histological examination after the IR measurement.

Here after, the IR imaging acquisitions were performed on breast tissue sections placed on Kevley slides in transflection mode. Each time prior to image acquisition, energy was verified and a reference spectrum from the bare substrate was recorded keeping the parameters constant as that of the tissue images. This served as a background which was subsequently subtracted from the dataset. The spectra were initially corrected for atmospheric corrections of water vapor and CO_2. Since no prior chemical deparaffinization was employed, an alternative method of mathematical deparaffinization was employed using a modified EMSC algorithm. In order to identify the biochemical distribution of various histological features of the breast tissue, the preprocessed spectral images were subjected to clustering algorithm using k-means (performed on 8 sample pairs) consisting of normal and tumoral breast tissues. The tumoral breast tissues consisted of both carcinoma *in situ* and invasive carcinoma. A representative k-means clustered image of an invasive carcinoma is presented in figure 4.

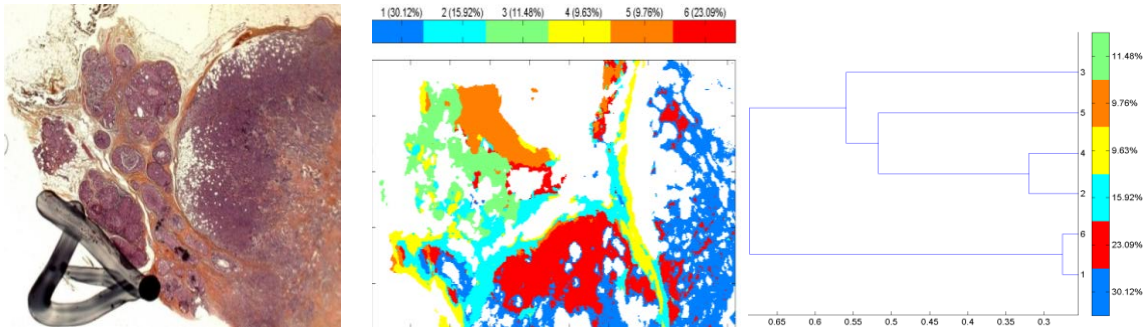


Figure 4: K-means clustering results.

K-means clustered images along with the respective dendrogram of a breast tissue with an invasive and *in situ* carcinoma in comparison to the adjacent HE stained images. The cluster representation is: Cluster 1 - invasive carcinoma, Cluster 6 - *in situ* carcinoma, Cluster 4 -connective tissue, clusters 2, 3, 5 - unattributed. The HE image is at 5X magnification.

In this preliminary work, it was attempted to initially classify the biochemical information of the breast tissues using clustering algorithms. Although it appears that the IR imaging with clustering algorithm was able to differentiate invasive and *in situ* carcinoma, at this point histopathological and spectral attribution was not clearly achieved due to various factors such as lesser morphological correlation between the reference image and the clustered images. It also appears that the dominating adipose tissue in the breast tissue leaves empty regions (white pixels) due to numeric deparaffinization by EMSC. However this needs to be verified. The perspective of this work would be to achieve a better correlation between the HE reference images and the spectral images and then use the spectral information to develop spectral markers which could differentiate different types of breast cancers. Also, since, the IR spectra were collected in transflection mode the spectral signatures are liable to certain light scattering effects. Hence, it is envisaged to characterize these spectral effects using various spectral treatments such as EMSC.

CHAPTER V

Conclusions and perspectives

V.1: Résumé :

L'imagerie spectrale IR constitue une méthode de diagnostic prometteuse capable de fournir des informations complémentaires à l'histopathologie conventionnelle.

L'un des objectifs de ce travail nous a permis de développer une nouvelle approche d'imagerie spectrale IR, ne nécessitant aucune étape de déparaffinage chimique, et ainsi de mettre au point une méthode directe d'analyse de tissu arrays paraffinés. Cette approche, combinant l'imagerie IR sur des coupes des tissus côloniques et l'analyse statistique multivariée a permis de donner lieu à un nouveau concept d'histopathologie spectrale permettant d'une part d'identifier les caractéristiques biochimiques et structurales intrinsèques des tissus, et d'autre part de différencier les tissus normaux et tumoraux du côlon.

Nous avons exploité les informations spectrales obtenues pour construire un modèle de prédiction automatisé, à partir de la gamme spectrale IR 1100-1300 cm^{-1}, qui est la plus discriminante. Cette région spectrale correspond à des vibrations moléculaires spécifiques telles que les vibrations des liaisons phosphate au sein des nucléotides et les vibrations des liaisons des polysaccharides ; ces biomolécules sont impliquées dans la cancérogenèse du côlon.

La validation sur des échantillons tissulaires totalement inconnus a permis de les identifier avec une sensibilité de 100%. Les images spectrales reconstruites selon un code couleur par le modèle de prédiction, ont révélé non seulement les caractéristiques biochimiques spécifiques des tissus coliques tumoraux, mais également des caractéristiques particulières comme le phénomène de tumor budding et l'interaction de la tumeur et du stroma.

D'autres méthodes d'imagerie vibrationnelle (ATR-IR et Raman) permettant d'améliorer la résolution spatiale ont été évaluées et leurs performances, avantages et inconvénients comparés. Nous avons par ailleurs mis en évidence les différents marqueurs spectraux des tissus normaux et tumoraux de manière simplifiée sous la forme d'un code-barres. Ce nouveau concept a pour but de faciliter l'interprétation de ces marqueurs et doit être testé sur un plus grand nombre d'échantillons.

V.2: Conclusions:

IR imaging is a biophotonic, non-destructive approach based on the interaction of light with matter. When applied to cells and tissues, the interaction can reveal the vibrational modes of chemical bonds and can provide a biochemical fingerprint that can be correlated to a diseased tissue state.

In this perspective, and with the aim to characterize colon cancers at the molecular level, IR spectral imaging in combination with multivariate statistical analysis has been implemented on colonic tissues. The colonic tissues included in the study consisted of both frozen tissues, and paraffinized tissue arrays. Several insights into the potentials of IR spectral imaging when applied on tissues for diagnostic purposes have been observed.

The initial feasibility work performed on the frozen tissues enabled to establish IR spectral imaging methodology to colonic tissues. The multivariate statistical analyses applied on the IR spectral signatures of the normal and the tumoral epithelial components constituted a novel concept of IR spectral barcodes which present an easy-to-interpret representation of discriminant features associated with normal and malignant colonic tissues.

In the following work which describes a novel approach of IR spectral histopathology on colon tissues for an automated colon cancer diagnosis, the paraffinized colonic tissue arrays were stabilized in an agarose matrix. To neutralize the spectral interferences arising from paraffin and agarose, a modified EMSC algorithm was implemented. The realization of the EMSC correction was observed as white pixels in the IR spectral images, and also in the retained spectra that were baseline corrected and normalized.

The EMSC algorithm not only permitted to neutralize the spectral interferences arising from paraffin and agarose, but also permitted comparison between the IR spectral images (after data processing) and the conventional histopathological images. The neutralization of the agarose interference adds another dimension to the EMSC algorithm projecting it as a custom-made correction method that can be employed to treat a variety of spectral interferences in an automated manner.

The EMSC corrected IR spectral images were subjected to an unsupervised k-means clustering which constituted a rapid and robust method well-adapted for huge spectral data sets. This method enabled to construct digitally stained spectral images representing the

overall colonic tissue organization. Comparison of the clustered images to the conventional histopathological images permitted to identify the normal and the tumoral colonic histological features based on their intrinsic biochemical composition. At the same time, the clustering method was also efficient in retrieving the IR spectral signatures specific to different histological components of normal and tumoral colonic tissues.

On the basis of these spectral signatures, biomolecular attribution of discriminant IR vibrations was made between the normal and the tumoral tissues using statistical tests such as Mann-Whitney *U* test, and PCA. These tests showed significant differences between the normal and the tumoral epithelium associated mainly to mucin, features along with alterations in nucleotides, carbohydrates, and proteins.

Further using these spectral signatures the diagnostic potential of IR spectral imaging was evaluated using a LDA based automated prediction model. Before implementation of the model for unknown sample identification in an external validation, the prediction model was tested in an internal validation set. The sensitivity of the prediction model observed in the form of a confusion matrix, from the confrontation between the histopathological class annotation (real class) and the IR spectral prediction (predicted class) showed several characteristics in correlation to the histological features. The confusion matrix showed no confusion between the normal and the tumoral epithelial components indicating a good separation between these two groups. Other features such as tumor associated stroma were observed to have some pixels grouped in the tumor, indicating the close biomolecular features associated with these groups.

Application of the model on unknown normal and tumoral samples permitted to recover the histopathology of the tissues in an automated manner. The color code for each histological class showed the percentage of pixels attributed to each class thus giving a quantitative dimension to the prediction. Importantly all the tumoral tissues were identified with 100% sensitivity. While the normal tissues were dominated by features such as mucosa, and connective tissue, the predominant features in the tumoral tissue were the tumor epithelium and its associated stroma. In addition, this approach also provided insights into certain important tumor-associated features like tumor budding, tumor-stroma association, and tissue inflammation.

The phenomenon of tumor budding, that is associated with tumor aggressiveness and lymph node metastasis, although can be identified by conventional histopathology at high

microscopic magnifications, was identified automatically without any chemical stains by the IR imaging approach. Furthermore, the intrinsic biochemistry based specific color code specifically differentiated the epithelial counter parts in the same tissue.

Interestingly, some of the normal tissues that were spectrally classed as tumor, revealed the presence of inflammation up on further analysis using IHC studies. Since, there was no group for inflammation in the model the spectra corresponding to this group were classed into tumor group. At this point, there is a need for amelioration of the prediction model where in additional classes specifically corresponding to spectral signatures from inflammation can be incorporated into the model and tested on unknown samples.

A complementary study that was carried out in order to compare the feasibilities of different imaging approaches put forth a general view of the advantages and the limitations of IR-conventional, IR-T, IR-ATR and Raman imaging. Comparison of the spectral images from these techniques showed that the ATR was able to identify some finer details of the colonic tissues that were not discernible in the conventional imaging. However, the time taken to achieve this was longer and the analyzed tissues are rendered inutile for further analysis due to the contact with the ATR crystal. While Raman imaging can further provide finer details the time taken increased considerably. However, this could render useful in providing complementary biochemical information from small and specific regions of interest.

Finally, the potential of IR spectral imaging methodology in combination with multivariate statistical analyses can be foreseen on the large scale validation of the prediction model. Moreover, it eliminated the need for sample staining and a priori knowledge of the unknown samples to be analyzed for biomolecular understanding. This study applied directly to paraffinized tissue arrays opens a new avenue for high-throughput retrospective studies for assimilating large spectral database. Based on the optimistic results, it appears as a complementary histological tool for diagnosis of colon cancers at this stage.

V.3: Perspectives:

The detection of the inflammatory signatures in the colonic tissues is an encouraging step for scrutinizing different inflammatory conditions associated with colon tissues, which may pose risk for development of cancer. A prediction model with a defined class for various inflammatory signatures considering the polymorphic nature of the inflammatory infiltrates,

could throw light into the direct diagnostic capabilities of this approach for inflammatory conditions.

Furthermore, the identification of adenomatous tissues by the prediction model gives optimistic diagnostic indications for the applicability of this methodology for early detection of biochemical changes. This need to be validated on a large scale studies before they can be employed in routine.

V. 3. 1: Clinical applications of infrared imaging:

Cancer is a disease that has its roots in the perturbations of molecular homeostasis of a living system (Stratton, 2011). Hence, early molecular level diagnosis is a critical step as it influences the outcome and overall survival rate of cancer patients.

In colorectal cancers, although several screening methods exist, a molecular level analysis of the symptomatic tissue cannot be accessed. IHC can be used to identify specific proteins of interest, which gives a molecular level understanding of the malignant condition (Fernandez, 2005). Presently, the demands of cancer diagnosis are met by gold standard histopathology by examining the morphological aberrations in tissues in a diseased condition (Kendall, 2009). The biophotonic technique of IR imaging which has been exploited in several studies appears as a good candidate complementary technique to access the more important molecular features of tissues, together with the morphological alterations.

Compared to the current general pathological protocols where, obtained tissue biopsies are fixed, paraffin embedded, microtome sectioned, stained, and examined under microscope, the advancements in the IR imaging capabilities over recent years show optimistic potentials to complement and simplify some of these steps. Firstly, regions of interest of the microtome sections can be directly imaged by IR imaging, and using multivariate analysis, spectral color-coded maps can be generated in a label-free manner, based on the intrinsic biochemical signatures. Secondly, with the advent of numerical deparaffinization, it is now possible to analyze directly the paraffinized tissues. This reduces the use of toxic chemical deparaffinization agents such as xylene (Travo, 2010; Ly, 2008) and gives access to large set of archived tissues with known patient history.

Recent technological advancements have spurred a new generation of IR spectrometers that now provide high acquisition speed with better spectral sensitivity. With advancements in the detector capabilities such as focal plane arrays, the acquisition of IR images of tissues can be rapidly performed facilitating analysis of large tissues in lesser time constraints (Bellisola, 2012). This when applied to tissue microarrays can provide huge spectral databanks in a high-throughput manner that can be used to develop and identify spectral markers for cancer diagnosis. Thus, a rapid diagnosis can be obtained thereby considerably reducing the amount of time taken. Technological advancements generate huge data sets, the exploitation of which have necessitated powerful chemometric algorithms and user-friendly software packages which have emerged in parallel.

One of the important aspects of IR imaging for clinical applications is the possibility to automate the diagnosis procedure. It has been shown that IR imaging combined with multivariate statistical methods can be automated by using computer driven programmes (Fernandez, 2005). This not only facilitates the operator utility, but also reduces the manual involvement, and in a longer run, the cost associated with it.

Most of the IR imaging studies have been restricted to analyzing *ex vivo* tissue samples. In regard to the use for *in vivo* applications, the main limitation of IR studies is its sensitivity to water considering the fact that it is the most abundant species confronted, which dominates the IR spectra (Bellisola, 2012). In this respect, the advent of IR probes (silver halide and ATR) has been tested directly on patients and has shown promising results in discriminating between normal and malignant tissues (Katukuri, 2010; Li, 2005).

Such technological improvements will progressively increase the number of potential applications of IR imaging to cancer research and clinical diagnosis, and represent new reasons of hope to introduce IR imaging into clinics. Although the potentials of IR imaging in clinical context are enormous based on their capabilities to provide biochemical information in a non-destructive, objective and label-free manner, with minimal sample preparations, at present this powerful analytical technique remains as a complementary tool to the conventional histopathology. This leaves an important demand for several interdisciplinary scientific research works for identification and application of appropriate approaches in cancer research, as well as in the diagnosis and follow-up of cancer diseases.

REFERENCES

Alberici P. The Adenoma-Carcinoma Sequence in Colorectal Cancer: Scratching the surface

Colorectal Surgery 2007.

Alberici P, Fodde R. The role of the APC tumor suppressor in chromosomal instability.

Genome Dyn. 2006;1:149-70.

Argov S, Sahu R, Bernshtain E, Salman A, Shohat G, et al. Inflamatory Bowel Diseases as an Intermediate Stage between Normal and Cancer: A FTIR-Microspectroscopy Approach. Biopolymers, 2004; Vol. 75, 384–392.

Barer R, Cole AR, Thompson HW, Infra-red spectroscopy with the reflecting microscope in physics, chemistry and biology. Nature 1949; 163: 198-201.

Beljebbar A, Bouché O, Diébold MD, et al. Identification of Raman spectroscopic markers for the characterization of normal and adenocarcinomatous colonic tissues. Critical Reviews in Oncology/Hematology 2009; 72, 255-264.

Bellisola G, Sorio C. Infrared spectroscopy and microscopy in cancer research and diagnosis. Am J Cancer Res 2012; 2(1):1-21.

Bhargava R. Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology. Anal Bioanal Chem 2007; 389:1155–1169.

Blout ER, Mellors RC. Infrared Spectra of Tissues. Science 1949; 110: 137-138.

Burt RW, Leppert MF, Slattery ML, et al. Genetic testing and phenotype in a large kindred with attenuated familial adenomatous polyposis. Gastroenterology 2004; 127 (2): 444-51.

Colley CS, Kazarian SG, Weinberg PD, et al. Spectroscopic Imaging of Arteries and Atherosclerotic Plaques. Biopolymers, 2004; Vol. 74, 328-335.

Ellis DI, Goodacre R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst 2006, 131, 875–885.

Fabian H, Thi NA, Eiden M, et al. Diagnosing benign and malignant lesions in breast tissue sections by using IR-microspectroscopy. Biochim Biophys Acta 2006; 1758(7), 874-882.

Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990; 61(5): 759-767.

Ferlay J, Shin HR, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010; 127(12), 2893-2917.

Fernandez DC, Bhargava R, Hewitt SM, et al. Infrared spectroscopic imaging for histopathologic recognition. Nature biotechnology 2005; Volume 23, number 4.

German MJ, Hammiche A, Ragavan N, Tobin MJ, Cooper LJ et al. Infrared spectroscopy with multivariate analysis potentially facilitates the segregation of different types of prostate cell. Biophys J 2006; 90(10), 3783-3795.

Giovannucci E, Stampfer MJ, Colditz GA, et al. Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study. Ann Intern Med 1998; 129 (7): 517-24.

Gremlich HU, Yan B, Infrared and Raman spectroscopy of biological materials (practical spectroscopy); Marcel Dekker Inc.: New York, 2000

Heather J, Stahl G, Bledsoe SB, et al. The Advantages of an Attenuated Total Internal Reflection Infrared Microspectroscopic Imaging Approach for Kidney Biopsy Analysis. Applied Spectroscopy 2010; Volume 64, Number 1.

Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; 411(6837):599-603.

International Agency for Research on Cancer (IARC). www.iarc.fr/en/publications/pdfs-online/pat-gen/bb2/bb2-chap6.pdf. Tumours of the Colon and Rectum, Chapter 6, 2004.

Ionov Y, Peinado MA, Malkhosyan S, et al. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 1993; 363: 558–561.

Janssen KP, Alberici P, Fsihi H, et al. APC and Oncogenic KRAS Are Synergistic in Enhancing Wnt Signaling in Intestinal Tumor Formation and Progression. Gastroenterology 2006; 131:1096-1109.

Kanter EM, Majumder S, Kanter GJ, et al. Effect of hormonal variation on Raman spectra for cervical disease detection. Am J Obstet Gynecol. 2009; 200(5).

Katukuri VK, Hargrove J, Miller SJ, Rahal K, Kao JY, et al. Detection of colonic inflammation with Fourier transform infrared spectroscopy using a flexible silver halide fiber. Biomedical Optics Express 2010; 1, 1014.

Kendall C, Isabelle M, Hegemark FB, et al. Vibrational spectroscopy: a clinical tool for cancer diagnostics. Analyst 2009; 134(6):1029-45.

Kendall C, Stone N, Shepherd N, et al. Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus. J Pathol 2003; 200: 602–609.

Khanmohammadi M, Ansari MA, Garmarudi AB, et al. Cancer diagnosis by discrimination between normal and malignant human blood samples using attenuated total reflectance-Fourier transform infrared spectroscopy. Cancer Invest, 2007, 25, 397–404.

Khanmohammadi M, Garmarudi AB, Ghasemi K, et al. Diagnosis of colon cancer by attenuated total reflectance-Fourier transform infrared microspectroscopy and soft independent modeling of class analogy. Med Oncol 2009; 26(3), 292-297.

Khanmohammadi M, Garmarudi AB, Samani S. Application of linear discriminant analysis and Attenuated Total Reflectance Fourier Transform Infrared microspectroscopy for diagnosis of colon cancer. Pathol Oncol Res 2010; 17(2), 435-441.

Kinzler KW, Vogelstein B. Landscaping the cancer terrain. Science 1998; 280 (5366): 1036-7.

Kohler A, Kirschner C, Oust A. Extended multiplicative signal correction as a tool for separation and characterization of physical and chemical information in Fourier transform

infrared microscopy images of cryo-sections of beef loin. Appl Spectrosc 2005; 59(6), 707-716.

Krafft C, Codrich D, Pelizzo G, et al. Raman and FTIR microscopic imaging of colon tissue: a comparative study. J. Biophoton. 2008; 1, No. 2, 154–169.

Krafft C, Kirsch M, Beleites C, et al. Methodology for fiber-optic Raman mapping and FTIR imaging of metastases in mouse brains. Anal Bioanal Chem 2007; 389:1133–1142.

Krafft C, Sobottka SB, Geiger KD, et al. Classification of malignant gliomas by infrared spectroscopic imaging and linear discriminant analysis. Anal Bioanal Chem 2007; 387(5), 1669-1677.

Kwiatkoski JM, Reffner JA, FT-IR microspectrometry advances. Nature 1987; 328, 837.

Laken SJ, Papadopoulos N, Petersen GM, et al. Analysis of masked mutations in familial adenomatous polyposis. Proc. Natl. Acad. Sci. USA 1999; Vol. 96, 2322–2326.

Larraona-Puy M, Ghita A, Zoladek A, et al. Development of Raman microspectroscopy for automated detection and imaging of basal cell carcinoma. Journal of Biomedical Optics 2009; 14_5.

Lasch P, Haensch W, Lewis N, et al. Characterization of colorectal adenocarcinoma sections by spatially resolved FT-IR microspectroscopy. Applied Spectroscopy 2002; 56, 1.

Lasch P, Haensch W, Naumann D, et al. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. Biochim Biophys Acta 2004; 1688(2), 176-186.

Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 1998; 396: 643-649.

Li QB, Xu Z, Zhang NW, Zhang L, Wang F, et al. In vivo and in situ detection of colorectal cancer using Fourier transform infrared spectroscopy World J Gastroenterol 2005;11(3):327-330.

Liu KZ, Xu M, Scott DA. Biomolecular characterisation of leucocytes by infrared spectroscopy. Br J Haematol 2007; 136: 713-722.

Ly E, Durlach A, Antonicelli F, et al. Probing tumor and peritumoral tissues in superficial and nodular basal cell carcinoma using polarized Raman microspectroscopy. Experimental Dermatology 2010; 19: 68–73.

Ly E, Piot O, Wolthuis R, et al. Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies. Analyst 2008; 133(2), 197-205.

Martens H, Nielsen JP, Engelsen SB. Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures. Anal Chem 2003; 75(3), 394-404.

Martin FL, Kelly JG, Llabjani V, et al. Distinguishing cell types or populations based on the computational analysis of their infrared spectra. Nat Protoc 2010; 5(11), 1748-1760.

Miyoshi H, Oka M, Sugi K, et al. Accuracy of detection of colorectal neoplasia using an immunochemical occult blood test in symptomatic referred patients: comparison of retrospective and prospective studies. Intern Med 2000; 39(9), 701-706.

Mourant JR, Yamada YR, Carpenter S, Dominique LR, Freyer JP. FTIR Spectroscopy Demonstrates Biochemical Differences in Mammalian Cell Cultures at Different Growth Stages. Biophysical Journal, 2003 ; 85, 1938–1947.

Nasse MJ, Walsh MJ, Mattson EC, et al. High-resolution Fourier-transform infrared chemical imaging with multiple synchrotron beams. Nat Methods 2011; 8(5), 413-416.

Pijanka J, Sockalingum GD, Kohler A, Yang Y, Draux F, et al. Synchrotron-based FTIR spectra of stained single cells towards a clinical application in pathology. Laboratory Investigation 2010, 1–11.

Pijanka JK, Kumar D, Dale T, Yousef I, Parkes G, et al. Vibrational spectroscopy differentiates between multipotent and pluripotent stem cells. Analyst, 2010 ; 135, 3126-3132.

Rex DK. Colon tumors and colonoscopy. Endoscopy 2000; 32(11), 874-883.

Rigas B, Morgello S, Goldman IS, Wong PT. Human colorectal cancers display abnormal Fourier-transform infrared spectra. Proc Natl Acad Sci U S A, 1990; 87(20), 8140-8144.

Rigas B, Wong PTT. Human colon adenocarcinoma cell lines display infrared spectroscopic features of malignant colon tissues. Cancer Research, 1992; 52, 84-88.

Rose DP, Boyar AP, Wynder EL. International comparisons of mortality rates for cancer of the breast, ovary, prostate, and colon, and per capita food consumption. Cancer 1986; 58 (11); 2363-71.

Salman A, Argov S, Ramesh J, Goldstein J, Sinelnikov I, et al. FT-IR microscopic characterization of normal and malignant human colinic tissues. Cellular and Molecular Biology, 2001; 47, 22, 159-166

Steller W, Einenkel J, Horn LC, et al. Delimitation of squamous cell cervical carcinoma using infrared microspectroscopic imaging. Anal Bioanal Chem 2006; 384(1), 145-154.

Stratton MR. Exploring the genomes of cancer cells: progress and promise. Science 2011; 331: 1553-1558.

Tfayli A, Piot O, Durlach A, Bernard P, Manfait M. Discriminating nevus and melanoma on paraffin-embedded skin biopsies using FTIR microspectroscopy. Biochim Biophys Acta 2005; 1724(3), 262-269.

The SK, Zheng W, Ho KY. Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue. British Journal of Cancer 2008; 98, 457-465.

Thiagalingam S, Laken S, Willson JK, et al. Mechanisms underlying losses of heterozygosity in human colorectal cancers. Proc Natl Acad Sci. 2001; 98:2698–2702.

Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon.

Science 1993; 260:816–819.

Travo A, Piot O, Wolthuis R, et al. IR spectral imaging of secreted mucus: a promising new tool for the histopathological recognition of human colonic adenocarcinomas. Histopathology 2010, 56, 921–931.

Triantafillidis JK, Nasioulas G, Kosmidis PA. Colorectal Cancer and Inflammatory Bowel Disease: Epidemiology, Risk Factors, Mechanisms of Carcinogenesis and Prevention Strategies. Anticancer Research 2009 ; 29: 2727-2738.

Vogelstein B, Kinzler KW. The multistep nature of cancer. Trends Genet 1993; 9 (4): 138-41.

Vogelstein B, Kinzler KW. The Genetic Basis of Human Cancer. 2nd ed. New York, NY: McGraw-Hill 2002; 583-612.

Wang TD, Triadafilopoulos G, Crawford JM, et al. Detection of endogenous biomolecules in Barrett's esophagus by Fourier transform infrared spectroscopy. PNAS 2007; vol. 104, 40, 15864-15869.

Weisenberger DJ, Siegmund KD, Campan M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet 2006; 38 (7): 787-93.

Wetzel DA. LeVine SM, Imaging Molecular Chemistry with Infrared Microscopy. Science 1999, 285, 1224-1225.

White SB, Romeo M, Chernenko T, Regina A, Miljković M, et al. Cell-cycle-dependent variations in FTIR micro-spectra of single proliferating HeLa cells: Principal component and artificial neural network analysis. Biochimica et Biophysica Acta, 2006 ; 1758, 908–914.

Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med. 1993; 329(27):1977-81.

Woernley DL. Infrared absorption curves for normal and neoplastic tissues and related biological substances. Cancer Res 1952; 12: 516-523.

Wood BR, Quinn MA, Tait B, Ashdown M, Hislop T, et al. FTIR Microspectroscopic Study of Cell Types and Potential Confounding Variables in Screening for Cervical Malignancies. Biospectroscopy, 1998 ; Vol. 4, 75–91.

Yano K, Ohoshima D, Gotou Y, et al. Direct Measurement of Human Lung Cancerous and Noncancerous Tissues by Fourier Transform Infrared Microscopy: Can an Infrared Microscope Be Used as a Clinical Tool? Analytical Biochemistry 2000; 287, 218–225.

Zhang X, Xu Y, Zhang Y, et al. Intraoperative Detection of Thyroid Carcinoma by Fourier Transform Infrared Spectrometry. Journal of Surgical Research 2010; 1-7.

Zuber TJ. Flexible sigmoidoscopy. Am Fam Physician 2001; 63(7), 1375-1380, 1383-1378.

PUBLICATIONS AND COMMUNICATIONS

International Publications

1. Infrared imaging as a cancer diagnostic tool: introducing a new concept of spectral barcodes for identifying molecular changes in colon cancers

Jayakrupakar Nallala, Olivier Piot, Marie-Danièle Diebold, Cyril Gobinet, Olivier Bouché, Michel Manfait, Ganesh D. Sockalingum

Submitted to the Journal of Cytometry Part A, July 2012

2. Infrared spectral imaging as a novel approach for histopathological recognition in colon cancer diagnosis

Jayakrupakar Nallala, Cyril Gobinet, Marie-Danièle Diebold, Valérie Untereiner, Olivier Bouché, Michel Manfait, Ganesh Dhruvananda Sockalingum, Olivier Piot

Submitted to the Journal of Biomedical Optics, April 2012

3. Infrared spectral histopathology for cancer diagnosis; a novel approach for automated pattern recognition of colon adenocarcinoma

Jayakrupakar Nallala, Marie-Danièle Diebold, Cyril Gobinet, Olivier Bouché, Ganesh D. Sockalingum, Olivier Piot, Michel Manfait

Submitted to the Journal of Cancer Research, July 2012.

Book Chapter

Diagnostic du cancer du côlon par histologie spectrale infrarouge

Jayakrupakar Nallala, Marie-Danielle Diebold, Cyril Gobinet, Valérie Untereiner, Olivier Bouché, Michel Manfait, Ganesh-Dhruvananda Sockalingum, Olivier Piot

LIVRE Biophotonique opt-diag, published by Publications MRCT CNRS (in press).

Communications orales

International

Spectral histopathology of paraffinized colon tissue microarrays: a new approach by infrared spectral imaging for colon cancer diagnosis
J. K. Nallala, C. Gobinet, O. Piot, V. Untereiner, M-D. Diebold, O. Bouché, M. Manfait, G. D. Sockalingum.
SPEC 2010, 26 June - 1 July, Manchester, United Kingdom

National

Infrared spectral imaging as a novel approach for automated histopathological recognition in colon cancer diagnosis
Nallala Jayakrupakar, Diebold Marie-Danielle, Gobinet Cyril, Untereiner Valérie, Bouché Olivier, Manfait Michel, Sockalingum Ganesh-Dhruvananda, Piot Olivier
Journée des Jeunes Chercheurs de la SFR CAP-Santé, 7th June, Amiens, France.

Communications par voie d'affiche

International:

Diagnosing colon cancer by infrared imaging; a new insight using spectral histopathology of paraffinized tissue microarrays

J. K. Nallala, C. Gobinet, M-D. Diebold, O. Bouché, V. Untereiner, G. D. Sockalingum, M. Manfait, O. Piot

OPT-DIAG, Diagnostic et imagerie optique en médicine, 9-11 Mai, 2012, ENSTA, Paris, France

Diagnosing colon cancer by infrared imaging: a new insight using spectral histopathology of paraffinized tissue microarrays

J. K. Nallala, M-D. Diebold, C. Gobinet, O. Piot, O. Bouché, V. Untereiner, G. D. Sockalingum, M. Manfait

CRP Santé Luxembourg - SFR CAP Santé Reims, November 28, 2011, Reims

Diagnosing colon cancer by infrared imaging: a new insight using spectral histopathology of paraffinized tissue microarrays

J. K. Nallala, M-D. Diebold, C. Gobinet, O. Piot, O. Bouché, V. Untereiner, G. D. Sockalingum, M. Manfait

European Conference on the Spectroscopy of Biological Molecules (ECSBM) 2011, 29 August - 3 September, Coimbra University, Portug

Diagnosing colon cancer by infrared imaging: a new insight using spectral histopathology of paraffinized tissue microarrays

J. K. Nallala, M-D. Diebold, C. Gobinet, O. Piot, O. Bouché, V. Untereiner, G. D. Sockalingum, M. Manfait

American Association for Cancer Research (AACR) 2011, 2 - 6 April, Orlando, Florida, USA

Spectral histopathology of paraffinised colon tissue microarrays: a new approach by infrared spectral imaging for colon cancer diagnosis

J. K. Nallala, C. Gobinet, O. Piot, V. Untereiner, M-D. Diebold, O. Bouché, G. D. Sockalingum, M. Manfait

American Association for Cancer Research (AACR) 2010, 17 - 21 April, Washington, DC, USA

Application of FT-IR imaging and data processing for spectral histopathology: a preliminary study on colorectal cancers

J. K. Nallala, C. Gobinet, O. Piot, V. Untereiner, M-D. Diebold, O. Bouché, M. Manfait, G. D. Sockalingum

Indo-French workshop on **Vibrational Spectroscopy and imaging 2010**, 13-15 February, Mumbai, India

National

Diagnosing colon cancer by infrared imaging: a new insight using spectral histopathology of paraffinized tissue microarrays

J. K. Nallala, M-D. Diebold, C. Gobinet, O. Piot, O. Bouché, V. Untereiner, G. D. Sockalingum, M. Manfait

5ème Forum de Cancéropôle Grand-Est, 2-3 november 2011, Strasbourg, France

Spectral histopathology of paraffinized colon tissue microarrays: a new approach by infrared spectral imaging for colon cancer diagnosis

J. K. Nallala, C. Gobinet, O. Piot, V. Untereiner, O. Bouché, G. D. Sockalingum, M. Manfait, M-D. Diebold

4ème Forum, Cancéropôle Grand-Est, 28 - 29 October, 2010, Strasbourg, France

Spectral histopathology of paraffinised colon tissue microarrays: a new approach by infrared spectral imaging for colon cancer diagnosis

Jayakrupakar Nallala, Cyril Gobinet, Olivier Piot, Valérie Untereiner, Marie-Danielle Diebold, Olivier Bouché, Michel Manfait, Ganesh D. Sockalingum

International symposium of the Federative Research Institute N° 53, Cell-Microenvironment interactions, June 7th to 9th, 2010, Reims, France

Jayakrupakar NALLALA

MOLECULAR CHARACTERIZATION OF TUMORAL LESIOSN BY INFRARED SPECTRAL IMAGING: IMPLEMENTATION OF A NEW CONCEPT BASED ON SPECTRAL HISTOPATHOLOGY FOR COLON CACNER DIAGNOSIS

Th. Pharm. Univ. : Reims: 2012

Abstract:

Innovative cancer diagnostic methods complementary to the gold standard histopathology are the need of the hour. In this perspective, the biophotonic approach of infrared spectral micro-imaging is one of the candidate methods capable of providing a biochemical fingerprint of cells and tissues in a label-free manner. Hence, a novel concept of infrared spectral histopathology of colonic tissues has been implemented in order to identify spectral signatures specific of colon histological structures, and to exploit these signatures to develop a prediction model comprising potential diagnostic markers for rapid and automated colon cancer diagnosis. For this, infrared images of colonic samples (moderately differentiated adenocarcinoma and non-tumoral) were acquired using an infrared imaging system. A mathematical deparaffinization was carried out on the spectral images using a modified Extended Multiplicative Signal Correction (EMSC) algorithm. The spectral data was subjected to clustering analysis in order to identify spectral signatures specific of colonic tissues. These signatures were used to develop a robust prediction model which was applied on unknown colonic tissue samples for histopathological identification. The prediction model not only identified the unknown tumoral tissues with 100 % sensitivity, but also some important tumor associated features such as tumor budding and tumor stroma association. Infrared spectral micro-imaging in conjunction with multivariate statistical analysis constituting a non-destructive and label-free approach, demonstrates the potential as a novel complementary tool to conventional histopathology for an automated and objective cancer diagnosis.

Key Words: Infrared spectral imaging, colon cancer, paraffinized tissue arrays, spectral histopathology

JURY

Rapporteurs:	Dr. Dominique GUENOT (Strasbourg)
	Dr. François LE-NAOUR (Paris)
Examinateurs:	Dr. Nicholas STONE (Exeter, GB)
	Dr. Jacques KLOSSA (Paris)
	Pr. Marie-Danièle DIEBOLD (Reims)
	Pr. Michel MANFAIT (Reims)
Directeurs de thèse:	Pr. Ganesh D. SOCKALINGUM (Reims)
	Dr. Olivier PIOT (Reims)

Author Address:

Jayakrupakar NALLALA

Apt 3B, 8 Place Maurice Utrillo, 51100, Reims.

Jayakrupakar NALLALA

CARACTERISATION MOLECULAIRE DE LESIONS TUMORALES PAR IMAGERIE SPECTRALE INFRAROUGE : IMPLEMENTATION D'UN NOUVEAU CONCEPT BASE SUR L'HISTOPATHOLOGIE SPECTRALE POUR LE DIAGNOSTIC DU CANCER DU COLON

Th. Pharm. Univ. : Reims: 2012

Résumé:

A l'heure actuelle, des méthodes innovatrices complémentaires à l'histopathologie pour le diagnostic de cancer sont en voie de développement. Dans cette perspective, une approche biophotonique telle la micro-imagerie spectrale infrarouge représente une méthode candidate capable de fournir une empreinte biochimique des cellules et des tissus sans étape de marquage. Par conséquent, un nouveau concept d'histopathologie spectrale infrarouge des tissus du côlon a été mis en œuvre afin d'identifier les signatures spectrales spécifiques des structures histologiques du côlon, et d'exploiter ces signatures afin de développer un modèle de prédiction comprenant des marqueurs potentiels pour le diagnostic du cancer du côlon de manière rapide et automatisée. Pour cela, les images infrarouges de différents échantillons coliques (adénocarcinome modérément différencié et non-tumorale) ont été acquises en utilisant un système d'imagerie infrarouge. Un déparaffinage mathématique a été réalisé sur les images spectrales en utilisant l'algorithme « extended multiplicative signal correction » (EMSC). Les données spectrales ont été soumises à une analyse de clustering, afin d'identifier les signatures spectrales spécifiques des tissus du côlon. Ces signatures ont été utilisées pour développer un modèle de prédiction robuste qui a été appliqué sur des échantillons des tissus du côlon inconnus pour l'identification histopathologique. Le modèle de prédiction, a non seulement identifié d'une part les tissus tumoraux inconnus avec une sensibilité de 100%, mais aussi d'autre part des caractéristiques importantes associées à la tumeur telles que le tumor budding et l'association de la tumeur et du stroma. La micro-imagerie spectrale infrarouge en conjonction avec l'analyse statistique multivariée, constituant une approche non destructive et ne nécessitant aucun marquage, démontre le potentiel de cette méthode comme outil complémentaire à l'histopathologie classique pour un diagnostic de cancer automatisé et objectif.

Mots clés: Imagerie spectrale infrarouge, cancer du côlon, histopathologie spectrale.

JURY

Rapporteurs:	Dr. Dominique GUENOT (Strasbourg)
	Dr. François LE-NAOUR (Paris)
Examinateurs:	Dr. Nicholas STONE (Exeter, GB)
	Dr. Jacques KLOSSA (Paris)
	Pr. Marie-Danièle DIEBOLD (Reims)
	Pr. Michel MANFAIT (Reims)
Directeurs de thèse:	Pr. Ganesh D. SOCKALINGUM (Reims)
	Dr. Olivier PIOT (Reims)

Adresse de l'auteur:

Jayakrupakar NALLALA
Apt 3B, 8 Place Maurice Utrillo, 51100, Reims.