Final Degree Project

MULTIDISCIPLINARY APPROACH FOR THE STUDY OF THE LIPID METABOLISM IN COLON CANCER

Simona Ivanova Ivanova

Biotechnology degree at the University of Vic-UCC

Tutors: Lara Nonell, Gwendolyn Barceló

Vic, September 2021

Acknowledgements

I am extremely grateful to the research team of Lipids in Human Pathology, who have tutored and mentored me along the way. They have truly inspired me and made an impact on my soon to begin career, taught me every possible knowledge the have and helped me with every issue I had along the project.

I would like to extend my gratitude to Dr. Gwendolyn Barceló for letting me be part of them during my internship and throughout my final degree project.

Special thanks to Albert Maimó for teaching me to act in a scientific manner and arouse my curiosity towards organoids and bioinformatics. Besides his key role in the redaction of this project and his patience. Also, to Maria Barceló for showing me the importance of being constant and hard-working in science. Last but not least, to Karim Pérez who taught me to elaborate proper hypothesis and to pursue them. This remarkable team has made and continues to make such a significant impact with their work, so I was fortunate to have learned from them.

Furthermore, special appreciation to Lara Nonell, my UVic tutor that played a significant role in this project and always has been available for my doubts.

Acronyms

CRC	Colorectal cancer.
FAP	Familiar adenomatous polyposis.
APC	Adenomatous polyposis coli gene.
HNPCC	Hereditary non-polyposis colorectal cancer.
MLH1	DNA mismatch repair gene related to colorectal cancer.
MSH2	DNA mismatch repair gene related to colorectal cancer.
SCs	Stem cells.
TAs	Transit amplifying cells.
FDs	Fully differentiated cells.
Lgr5	Leucine-rich-repeat containing G-protein-coupled receptor.
EphB2	Ephrin type B receptor 2.
FACS	Fluorescence activated cell sorting.
CIN	Chromosomal instability.
MSI	Microsatellite instability.
CIMP	CpG island methylator phenotype.
SP	Serrated pathway.
SCNA	Somatic copy number alterations.
TSA	Traditional serrated adenomas.
SSA/P	Sessile serrated adenomas or polyps.
TCGA	The Cancer Genome Atlas project.
CRCSC	Colorectal cancer subtyping consortium.
CMS	Consensus molecular subtypes.
TNM	Tumor-Node-Metastasis system.
AJCC	American Joint Committee on Cancer.
gFOBT	Guaiaic fecal occult blood test.
FIT	Fecal immune-chemical test.
MS	Mass spectrometry.
IMS	Imaging mass spectrometry
MALDI	Matrix assisted laser desorption/ionization.
WGCNA	Weighted gene co-expression network analysis.
GS	Gene significance.
MM	Module membership.
KIM	Intramodular connectivity.
PBS	Phosphate-buffered saline.
DTT	Ditiotreitol.
SB	Staining buffer.
FBS	Fetal bovine serum
DMEM	Dulbeccos Modifyed Eagle Medium.
TOM	Topological overlap matrix.

Summary

Title: Multidisciplinary approach for the study of the lipid metabolism in colon cancer.
Author: Simona Ivanova Ivanova.
Date: September 2021
Keywords: colorectal cancer, colonocytes, lipidomics, transcriptomics, WGCNA analysis, gene mutations, organoids.

Colorectal cancer is a leading cause of cancer death worldwide, its late detection contributes for become one of the deadliest. Research team of Lipids in Human Pathology has proven a relationship among membrane lipids and tumorigenic processes in CRC. Thus, the importance of studying the impact of colorectal cancer in the lipid composition during the differentiation process of colonocytes. Given that lipidome of colonocytes is sensible to changes occurring during differentiation from stem cell to mature colonocyte, EphB2 is used to as a marker to separate stem cells from fully differentiated colonocytes.

Main objectives of this project are to get more knowledge on the regulatory mechanisms involving lipid composition throughout bioinformatics and organoid culture. Methodology followed consists of several approaches such as transcriptomics, lipidomics and as already said, organoid culture.

Weighted gene co-expression network analysis (WGCNA) was performed from microarray data of CRC patients to explore complex relationships between genes and phenotypes in this case lipid traits by transforming gene expression data into co-expression modules. Specially focusing on the relationship with sphingolipid SM d34:1. Results indicated higher presence of SM in the bottom of the crypt containing stem cells. Whilst through the differentiation into mature colonocytes levels of SM where lower. This observation was similar in both healthy and tumor condition. Furthermore, results suggested that the transcriptomic profile (ie. expression profile) between healthy and tumor condition is different. Consequently, regulation mechanisms of SM d34:1 among the two conditions are also different.

On the other hand, to evaluate the incorporation of deuterated fatty acids in healthy colon and tumor organoids particularly between differentiated and non-differentiated regions of the crypt, organoid culture experiment was performed. Unfortunately given the fact that organoid culture is a 6-8 weeklong process, it wasn’t possible to extract lipidomic results.

Index of contents

Index of tables

Index of figures

1. Introduction

This final degree project has been carried out at the Lipids in Human Pathology Laboratory located at the Health Research Institute of the Balearic Islands (IdISBa, Institut d'Investigació Sanitària Illes Balears) situated at the University Hospital of Son Espases. Research group led by Dr. Gwendolyn Barceló-Coblijin focuses its interests on the role of membrane lipids in cell pathophysiology to apply this knowledge to develop new early diagnosis and treatment tools for conditions such as inflammatory bowel disease and colorectal cancer.

1.1 Colorectal cancer

1.1.1 Demographics

Colorectal cancer (CRC) is the fourth leading cause of cancer death and the third most diagnosed malignancy worldwide (Bray et al., 2018). This type of cancer is preventable although its late detection contributes to become one of the deadliest. Even though, implementation of screening procedures and early detection programs in the population over fifty years old has reduced the CRC incidence in high income countries, there is evidence suggesting an increase of incidence on the population under fifty years old (Araghi et al., 2019). For this reason, it is necessary to develop proper diagnosis and classification methods gathering phenotypic and genetic traits of the tumor. I will discuss them after introducing some of the most important pathology concepts needed for a proper understanding.

1.1.2 Risk and preventive factors

Several studies from the American Institute for Cancer Research, suggest diet to be one of the most significant exogenous factors in colorectal cancer. Hence, processed and red meat combined with excessive alcohol consumption increases chances of developing CRC (Labianca et al., 2010).

In addition, the importance of proper physical activity plays a key role for colorectal cancer prevention even though evidence is stronger for colon than for rectum cancer (Wiseman, 2008). Thus, body and abdominal fatness along with overweight are also causes of CRC cancer.

Furthermore, there are non-dietary risk factors like age, smoking tobacco (Giovannucci, 2001), chronic consumption of non-steroidal anti-inflammatory drugs (Asano & McLeod, 2004), combination with other colorectal diseases (i.e Chron's) (von Roon et al., 2007) and ulcerative colitis (Eaden et al., 2001), genetic predispositions, and metabolic syndrome (Ahmed et al., 2006).

Most cases of CRC are developed sporadically although in around 5% of the cases it can be hereditary (Kwak & Chung, 2007). Genetic factors related to CRC propensity are polyposis or non-polyposis syndromes. Most common polyposis syndrome is familial adenomatous polyposis (FAP) (Waller et al., 2016), associated with mutations or losses of APC gene called adenomatous polyposis coli. It is an autosomal dominant inherited disorder causing early appearances of thousands of adenomatous polyps around the colon. Without treatment this contributes to colon cancer appearance by the age of 35-40 years. The other syndrome, hereditary non-polyposis colorectal cancer (HNPCC) or Lynch syndrome is an autosomal dominant condition. It is associated with six DNA mismatch repair genes, most common of which are MLH1 and MSH2 (Lynch & Lynch, 2000).

1.1.3 Pathology

The colon and rectum are part of the large intestine, being colon a major part of the large bowel. Its functions are absorption of water, electrolytes, and vitamins (Azzouz & Sharma, 2021). Besides, forming, and propelling stool in direction to the rectum for posterior elimination. In some literature, the word colon is used to designate the large intestine creating a confusion. As well as the terms colon and colorectal cancer where both affect the large intestine but start in different places, first one located in the colon and the second either in the colon or rectum. Colon and rectal cancer are grouped together because of similar features like incidence, symptoms, risk factors and genetic conditions.

Colon and rectum contain four layers which are mucosa, submucosa, muscularis propia and serosa. Starting from the exterior part, serosa is the outer surface and contains sheets of connective tissue. The main function of serosa is protection from the spread of any inflammatory or malignant process. Next layer is muscularis propia including the inner circular layer, the intermuscular space and outer longitudinal layer. The major function of muscularis propia is propelling food through the gut. Adjacently, there is the submucosa layer with diversity of inflammatory cells, lymphatics, nerves and ganglion cells. Here is where the arteries and venous channels are found. Its function is to allow the junction of mucosa and muscular layer. Finally, the mucosa, which is the inner layer with structural and functional complexity. It is the most important layer of the gastrointestinal tract because of its absorptive function. Mucosa divides into three layers: epithelium, lamina propia and muscular muscosae. Epithelium is in direct contact with the intestinal lumen, it is composed of a single monolayer of colonocytes invaginating towards the stroma creating the crypts. This project is focused on the colon crypts at the epithelium (Jaladanki & Wang, 2011).

Colon and intestine crypts contain stem cells (SCs) at the bottom, transit-amplifying cells (TAs) in the middle and fully differentiated cells (FDs) at

the top of the crypt. SCs divide and differentiate while ascending the crypt to form mature colonocytes (figure 1). Moreover, in this case, they are responsible for tissue regeneration of the intestine. There are two types of SCs: central stem cells (CeSCs) located at the very bottom and border stem cells (BSCs) located between CeSCs and TA with high differential potential (Mahdipour-Shirayeh & Shahriyari, 2018). Transit-amplifying cells give place to three differentiated cell types which are enterocytes, goblet and enteroendocrine cells (Barker et al., 2007).

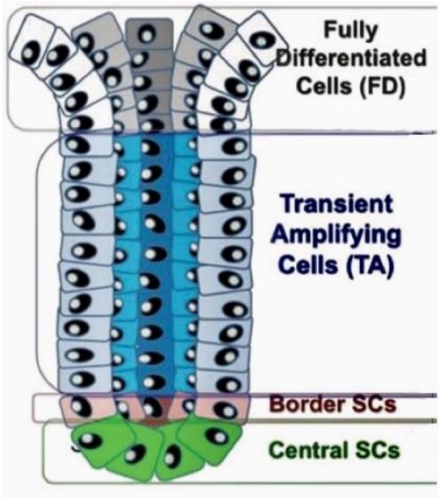

Figure 1: Types of cells forming the crypts. Stem cells at the bottom, central and border SCs. Following a significant part of transient amplifying cells. Finally, the fully differentiated cells at the top of the crypt. (Barker et al., 2007).

The most used stem cell marker in crypts is leucine-rich-repeat containing G-protein-coupled receptor 5 (Lgr5 or Gpr49). It is a Wnt target gene that allows to mark stem cells in various tissues and cancers. Thus, when there is expression of Lgr5+ it means that there are stem cells present (Barker et al., 2007). Another possible marker for intestinal SCs is Eph receptor B2, a tyrosine kinase transmembrane molecule (EphB2). It is also a Wnt target gene as the activation of Wnt pathway that it will be explained later contributes to expansion of the crypt (Merlos-Suárez & Batlle, 2008). Intestinal stem cells have high expression of Lgr5 and EphB2 and this expression decreases to become inexistent as the cells differentiate.

The purpose of using stem cell markers is to isolate the intestinal epithelial stem cells using fluorescence-activated cell sorting (FACS) and to visualize their gradient of expression.

Molecular pathogenesis of CRC is heterogeneous and contains multiple-step processes, usually develops after 10 years. Most common type of colorectal cancer are adenocarcinomas, whereas the others can be carcinoid tumors, gastrointestinal stromal tumors, lymphomas, Turcot

syndrome, Peutz-Jeghers syndrome and familiar CRC. Hence, this project is about adenocarcinomas.

As said, high percentage of colorectal cancers occur after mutations at the abnormal cells of the mucosa layer that keep dividing forming a polyp usually benign, slowly growing, encapsulated and non-invasive. There are different types of polyps, classified histologically into two categories neoplastic and non-neoplastic, being adenomas or adenomatous polyps the most common (Colucci et al., 2003; Mansoor et al., 2013) (table 1).

Histological classification	Polyp type	Malignant potential
Neoplastic	Tubular adenomas Tubulovillous adenomas Villous adenomas Serrated adenomas	Yes
Non-neoplastic	Hyperplastic Hamartomatous Lymphoid aggregates Inflammatory polyps	No

Table 1:Classification of colorectal polyps. This table contains a classification of the different type of polyps with information about histological classification and potential malignancy.

Since the last decade, researchers have identified four pathways of colorectal carcinogenesis. These are chromosomal instability (CIN), microsatellite instability (MSI), CpG island methylator phenotype (CIMP) and serrated pathway (SP) (figure 2).

Chromosomal instability is the most common path adenoma-carcinoma by 84% of sporadic CRC. It is an alteration of the structure and number of chromosomes including translocations, deletions, gains, and similar chromosomal changes. Resulting in high number of somatic copy number alterations (SCNA), mutations in oncogenes and tumor suppressor genes such as APC, TP53, KRAS and BRAF. In addition, mutations on APC genes that are negative regulators for WNT signaling pathway cause hyperactivation of WNT pathway. And this event creates uncontrolled proliferation and differentiation of crypts (Müller et al., 2016).

Microsatellite instability is accounting for 16% of CRC and it's based on mismatch repair gene inactivation, as a result produces high frequency of replication errors. Three-quarters of MSI tumors are sporadic and present hypermethylation of mismatch repair gene MLH1 triggered by CpG island methylation phenotype, thus silencing the MMR gene. Furthermore, sporadic hypermutated cancers have $BRAF^{V600E}$ or similar mutations that are used to distinguish sporadic MSI from hereditary MSI tumors. The remaining one-quarter corresponds to hereditary MSI tumors such as Lynch syndrome and FAP, containing somatic mutations on mismatch

repair genes (MMR) like MSH2, MSH6, MLH1 and PMS2 (Harada & Morlote, 2020).

CIMP pathway is responsible for the hypermethylation of promoters CpG island loci causing the inactivation of tumor suppressor genes. CIMP is not exclusive like CIN and MSI, it overlaps with MSI in the cases of sporadic MSI tumors that present CIMP methylation of MLH1 gene (Harada & Morlote, 2020).

A small minority of CRC can develop throughout the serrated pathway. Even though most serrated polyps are benign, there are some that can progress into carcinoma. The two premalignant precursors in SP are traditional serrated adenomas (TSA) and sessile serrated adenomas or polyps (SSA/P). In this case, characterization is not made by mutations, but by morphological changes because SP pathway overlaps with CIN and MSI. TSAs are microsatellite stable present KRAS mutations and sometimes BRAF. Morphologically cytoplasm is eosinophilic and central elongated hyperchromatic nuclei. On the other hand, SSA/P present BRAF mutations and CIMP as an early feature with MLH1 promoter hypermethylation. Morphologically crypts have abnormal form of inverted-anchor and horizontal growth (O'Brien et al., 2015).

Hence, there are two molecular pathological classification systems that use these four principles. First the Cancer Genome Atlas project (TCGA) and then the CRC Subtyping Consortium (CRCSC). TCGA project is found on genomic and transcriptomic analysis for CRC characterization using sequencing technologies and arrays (Müller et al., 2016). And CRCSC, is based on gene expression profiling into four Consensus Molecular Subtypes (CMS) (Guinney et al., 2015).

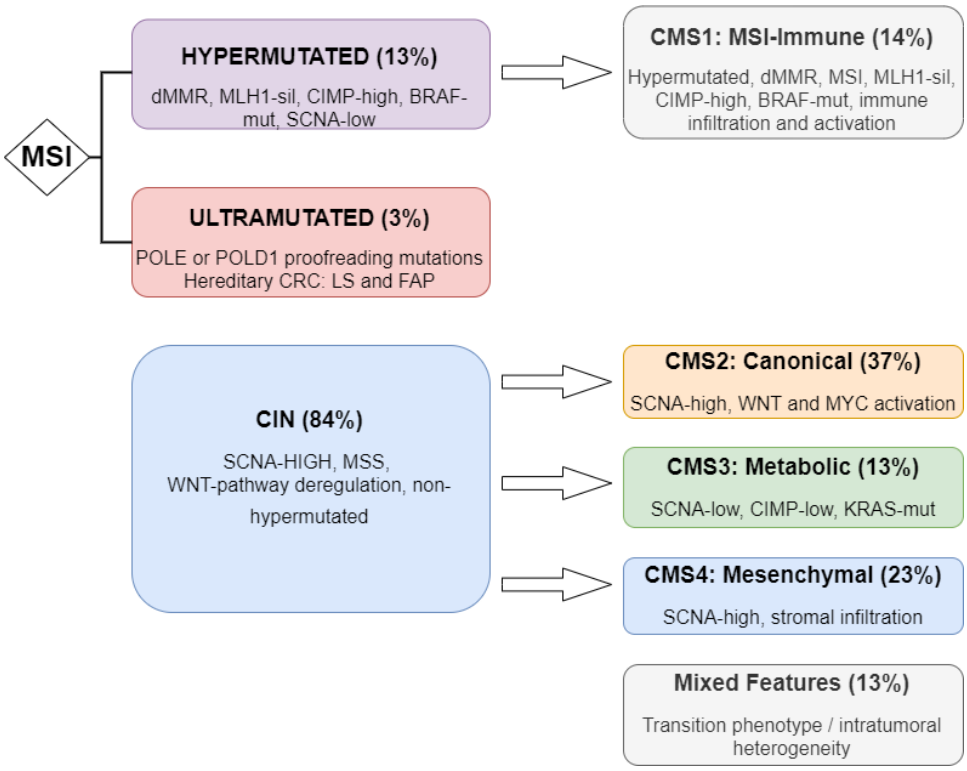

Figure 2: Molecular classification systems of CRC. TCGA classification on the left side and CMS on the right side. TCGA classification is divided into two major groups according to the mutation rate: (1) 84% CIN, non-hypermutated (low mutation rate) with high frequency of somatic copy number alterations (SCNA), microsatellite stability (MSS) and deregulation of WNT pathway most of the times by APC mutation. (2) 13% Hypermutated tumors with microsatellite instability (MSI) due to defective mismatch repair (dMMR) majority of cases triggered by MLH1 silencing via CpG island methylation phenotype (CIMP). Approximately 80-90% of sporadic hypermutated cancers present BRAF mutations. (3) 3% Ultramutated tumors with proofreading mutations leading to defective mismatch repair and MSI, this group corresponds to hereditary CRC as Lynch syndrome (LS) and Familial adenomatous polyposis (FAP). On the right side, CMS classification based on the expression signatures, four CMS groups: (1) 14% MSI-Immune, (2) 37% Canonical, (3) 13% Metabolic, (4) 23% Mesenchymal and a residual unclassified group with mixed traits. The molecular features of each subtype are indicated above. Author's figure.

1.1.4 Diagnosis methods

Diagnosis of CRC is made histologically by extracted biopsy sample during endoscopic tests like colonoscopy. Nowadays, the most common histological tool used worldwide by clinicians to stage and classify CRC is the Tumor-Node-Metastasis system (TNM). Created by Pierre Denoix in 1940 it has evolved to this day incorporating the latest updates through the American Joint Committee on Cancer (AJCC). This system name stands for primary tumor aspects (T) like size, local growth, and bordering structures. Followed by the extent of the lymphatic node metastasis (N) like number and location of the nodes involved and the distant metastasis (M). Additionally, each letter of the TNM system can be combined with another letter or number from zero to four allowing a more accurate staging of the cancer and facilitating a proper treatment plan (Telloni, 2017).

Another approach for diagnosis is using molecular markers such as key mutations happening during CRC. Since the knowledge on colorectal molecular pathogenesis is properly described, it can be used as a reference on diagnosis. Hence, as explained before, each group of CRC has its characteristic mutations and traits that can be used for the purpose of an accurate diagnosis. Likewise, there are types of screening like fecal test (guaiaic fecal occult blood test, gFOBT; fecal immune-chemical test, FIT and fecal DNA test), radiologic test (computed tomographic colonography) and blood test (Septin9 gene test)(Maida et al., 2017).

1.2 Lipids

In the past, shortage of techniques allowing massive lipid analysis has made it difficult to progress research on these topics. Lipidomics is a discipline that studies cellular lipids structures, functions, pathways, networks, and interaction with other cellular components on a large scale based on analytical chemistry principles and technological tools (Yang & Han, 2016). To control the abundance as well as the distribution of the molecular variants of lipids regarding different lipid species, the cells spend significant amount of energy in terms of enzymes to accomplish it. Thus, changes in the lipid composition at the level of affecting the pathophysiological state of cells increases the interest in the field of lipidomics allowing a desire to improve and dedicate more attention to identify novel lipid functions and interactions. That is why there is a growing interest on lipids in biomedical research combining multi-omics approaches (Kopczynski et al., 2017).

Techniques like mass spectrometry (MS), an analytical tool that identifies chemical species based on their mass-to-charge ratio (m/z) made a significant positive impact allowing to untangle lipid heterogeneity as well as roles of lipid metabolism in biological processes. Likewise, imaging mass spectrometry (IMS) improves the knowledge on lipids, proving a regulation of their distribution and presenting the visualization of hundreds of spatially organized lipid phenotypes in their context. In addition, it demonstrates that cell lipidome is dependent on cell type and changes like differentiation and tumorigenesis. In IMS the ionization source can be an ion beam, a laser on a solvent of charged droplets. In this case, the ionization source is a laser called MALDI (Matrix Assisted Laser Desorption/Ionization) where the lipid identification is based on comparing the m/z ratio values and the lipids in the software lipid database. The result is the capability of visualizing pixel by pixel the changes that happen in the lipidome through the crypt following a mathematical equation of a first-degree equation (Barceló-Coblijn & Fernández, 2015) (figure 3).

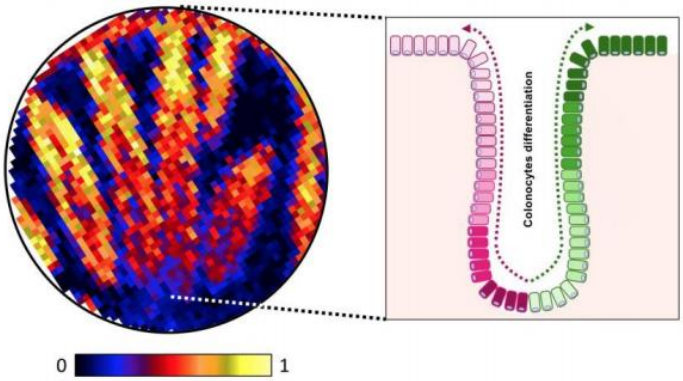

Figure 3: Analysis of human colon section by MALDI-IMS. Showing that a subset of lipids: phosphatidylinositol and phosphatidylethanolamine plasmogens change their signal intensity through the crypt and differentiation process. On the right, the diagram of the crypt indicating the direction of the differentiation. There is an increasing gradient of lipids containing mono-unsaturated fatty-acids from bottom to top (following a first-degree equation (in green colors), while a decreasing gradient of lipids containing arachidonic acid (in pink colors) (Bestard-Escalas et al., 2019).

Lipids are relatively small compared to other biological macromolecules and changes in their composition most of cases are due to the reposition of a single or double bond. Moreover, lipids are key components of cellular membranes and lipid particles for instance lipoproteins. Even though they are complex regarding species, they have essential roles in cellular barriers, membrane matrices, signaling pathways and energy storage. In addition, lipids are very dynamic, continuously changing because of physiological, pathological, and environmental conditions (Yang & Han, 2016). Furthermore, several studies validate that membrane lipids are very good candidates for biomarkers because of the high specificity of the lipidome used to categorize cells in line with its origin and pathophysiological state (Bestard-Escalas et al., 2019).

The classification of mammalian lipid categories according to the Lipid Maps Structure Database includes eight categories: fatty acids, glycerophospholipids (or phospholipids), sterols, sphingolipids, glycerolipids, prenol lipids, saccharolipids and polyketides. Acetyl-coA is their common precursor and the major lipid components in mammalian membranes are phospholipids and sphingolipids. There is a challenge regarding lipid analysis due to lipid complexity, many membrane lipid species are similar as to chemical structure. Specially, phospholipids and sphingolipids that contain numerous different molecules depending on the fatty acids they have present in each case, some cases with differences only in a double bound but advances in mass spectrometry methods for lipid analysis has improved over the last years (Bestard-Escalas et al., 2019).

Cell membrane, also known as plasma membrane, is a phospholipid bilayer with amphipathic behavior. The hydrophobic domain contains fatty-acid tails while the hydrophilic domain contains the head groups. This way polar head groups are in contact with the water whilst hydrophobic core is not and allows the activity for transmembrane proteins. Lipid membrane composition varies depending on the tissue and subcellular organelle. At first, scientists believed that lipids were randomly allocated through the membrane leaflet. On reality, there are two organizational levels one between the two leaflets (inner and

outer leaflet, transversal asymmetry) and the other within each leaflet (lateral asymmetry). Lipid composition of the membranes determine biophysical properties of such membrane. Interestingly it is the transversal asymmetry which includes differences in the composition among the two leaflets with functions modulating properties of the bilayer such as surface charge, membrane potential, permeability, shape, and stability. Inner leaflet contains phosphatidylethanolamine, phosphatidylserine, and phosphatidylinositol, meanwhile the outer leaflet contains phosphatidylcholine, sphingomyelin, and glycolipids. The phospholipids are asymmetrically placed among the two halves of the bilayer as well as the cholesterol (van Meer et al., 2008) (figure 4).

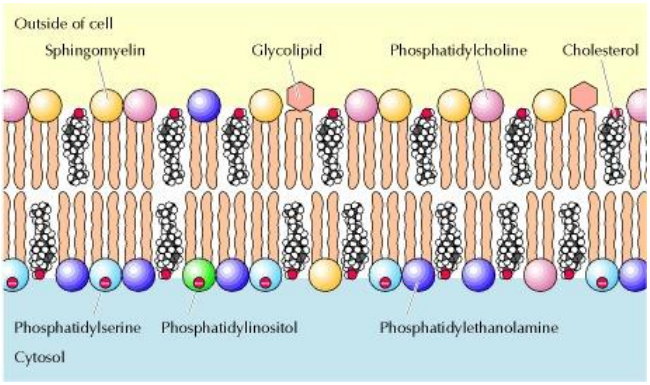

Figure 4: Plasma membrane structure. Inner leaflet contains phospholipids such as phosphatidylethanolamine, phosphatidylserine, and phosphatidylinositol. Outer leaflet contains phosphatidylcholine, sphingomyelin, and glycolipids. In addition, there are phospholipids and cholesterol placed between in both leaflets of the membrane bilayer. Cooper, G. M., & Hausman, R. E. (2009). The cell: A molecular approach. Washington, D.C: ASM Press.

Phospholipids have amphipathic nature, containing hydrophilic polar region and hydrophobic non-polar region. They are the major component of plasma membranes. Furthermore, phospholipid synthesis happens at the endoplasmic reticulum. The structure consists of hydrophilic head (phosphate group attached to glycerol) and two hydrophobic tails of fatty acids. Moreover, the phosphate group can be replaced by ethanolamine, choline, or serine. In addition, the classification of main phospholipids is phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, phosphatidylinositol and phosphatidate (Quinn, 2002).

Sphingolipids have amphipathic nature, containing hydrophobic and hydrophilic properties. Starting with the hydrophobic region that comprises sphingoid base with a fatty acid chain attached. On the other hand, the hydrophilic region has phosphate groups, sugar residues and hydroxyl groups. As said, sphingolipids are crucial components of plasma membranes. They start synthesizing at the endoplasmic reticulum and ends in the Golgi apparatus where takes place the addition of carbohydrates and other components on the ceramide. Different molecules are generated depending on the polar head that they attach. Hence, depending on the modifying structures they are classified as sphingomyelins (phospocholine or phosphoethanolmine and ceramide) or glycosphingolipids (cerebrosides and gangliosides) (Futerman & Riezman, 2005).

1.3 Weighted gene co-expression network analysis

Before getting deep with weighted gene co-expression network analysis, I am going to contextualize some important bioinformatic aspects. The on growing omics sciences are a wide field dedicated to study biological processes like genomics, transcriptomics, lipidomics, proteomics and so on (Li et al., 2017). This project is focused on transcriptomics and lipidomics, being the study of cell RNAm transcripts and the study of the lipidome of cells respectively. Improvements in transcriptomics have taken the identification of differentially expressed genes on to the next level thanks to technologies like microarrays (Olivier et al., 2019). Microarray technology allows to study the expression of multiple genes at once. Consists of placing thousands of gene sequences in known locations on a glass slide which is known as the gene chip. Where the sample of either DNA or RNA is placed in contact with the gene chip. When complementary base pairing among sample and gene sequence happens, the chip emits light which is measured (Loewe & Nelson, 2011). The vast majority of data obtained from experiments are analyzed using the R and Rstudio platform presenting many advantages such as the ability to process higher amounts of samples data and obtain conclusions.

In addition, after Sanger sequencing there has been an improvement because of the development of high-throughput technologies. Since then, they have become crucial in biology and medicine research specially genomics, epigenomics and transcriptomics. Hence, using high-throughput technologies allows to sequence multiple DNA molecules in parallel and creating large data sets for later analysis. Here is when network biology became popular, providing deeper understandings on complex biology systems and modularity in tissue of cellular networks. These correlation networks used in bioinformatic applications give place to the development of the weighted gene co-expression network analysis (WGCNA) by Steve Horvath and his colleagues Peter Health and Bin Zhang (Churko et al., 2013).

WGCNA analysis is very helpful in system biology, used to describe correlation patterns between genes from microarray samples. It is useful for finding clusters which are the modules of highly correlated genes, also allows to summarize clusters through module eigengene. Besides, using eigengene network method it can relate modules among them and to external traits such as in this case, relative expression of lipid species. Moreover, it can calculate module memberships and assist network-based gene screening techniques allowing an easier identification of possible biomarker candidates or therapeutic targets. The network construction that takes place in WGCNA analysis contains the nodes that represent genes, this nodes are connected depending on their co-expression in the chosen tissue samples, in this case colon epithelium. (Langfelder & Horvath, 2008).

All this can be achieved by R studio using WGCNA software package containing numerous functions that allow the weighted gene co-expression network analysis. For instance, functions for network construction, module detection, module and gene selection, topological properties calculations, data simulation and visualization.

To facilitate the comprehension of key terms explained and yet to be explained in the methodology, I provided a table of definitions.

Term	Definition
Correlation network	Constructed based on correlations between quantitative measurements with a matrix structure where the row corresponds to the nodes of the network and the column to sample measurements. This correlation network can be used to identify clusters (known as modules) that contain the nodes.
Node	Contains the genes in co-expression networks.
Node profile	Is the gene expression network
Node significance	Is the measurement of the gene significance (GS).
Co-expression network	Is an undirected, weighted gene network. The nodes of such a network correspond to gene expression profiles, and edges between genes are determined by the pairwise correlations between gene expressions. Through raising the absolute value of the correlation to a power $\beta \geq 1$ (soft thresholding), the weighted gene co-expression network construction emphasizes high correlations at the expense of low correlations.
Modules	Are clusters of highly interconnected genes. In an unsigned co-expression network, modules correspond to clusters of genes with high absolute correlations. In a signed network, modules correspond to positively correlated genes.
Connectivity	Regarding each gene is the sum of connection strengths with the other network genes. At co-expression networks it measures the correlation of a gene with all the other genes in the network.
Intramodular connectivity (KIM)	Allows to measure the degree of connectivity or co-expression of a given gene regarding the other genes of a particular module. It is understood as a measure of module membership.
Module eigengene (E)	Considered as the representative of gene expression profiles in a module. Is the main component of a module.
Eigengene significance	When a microarray sample trait is available it can be correlated with the module eigengenes. Eigengene significance is the correlation coefficient.
Module membership (MM)	Also known as eigengene-based connectivity KME. It is a measure for each gene. And this estimate measurement of

	module membership comes from the correlation of its gene expression profile with the module eigengene of a given module. When the MM value is close to 0 it means that this gene is not part of this particular module. Whereas if the value is 1 o -1 this gene is highly connected to that module. The sign of the value translates into if the gene has positive or negative relationship with the module. In addition, MM is very related to intramodular connectivity KIM.
Gene significance (GS)	Use of GS measures allows to incorporate external information into the co-expression network. The gene is more biologically significant when its value is very high. On the contrary, when its value is close to 0 indicates that this gene is not significant.
Module significance	Average absolute gene significance measure for all genes forming a module
Hub gene	Abbreviation of highly connected gene.

Table 2: Glossary of most used WGCNA terminology. Autor's table.

2. Objectives and hypothesis

The team of Lipids in Human Pathology has proven a relationship between membrane lipids and the tumorigenic process in CRC. Hence, the lipidome of colonocytes is sensible to changes happening during differentiation from stem cell to fully matured colonocyte and during the malignancy process. Besides, stem cells have higher levels of phospholipids containing arachidonic acid (molecule with key role on cell signaling). These levels tend to decrease regularly throughout cell differentiation. Moreover, the levels of these lipids increase in the tumoral colonocytes. In order to get more knowledge on the regulatory mechanisms involving lipid composition two objectives are proposed.

Objective 1: Performing a system biology approach by weighted gene co-expression network analysis using sphingolipid SM d34:1.

Objective 2: Evaluate the incorporation of deuterated fatty acids in healthy colon and tumor organoids and between the differentiated and non-differentiated regions of the crypt.

3. Methodology

3.1 Weighted gene co-expression network analysis

In order to carry out the WGCNA analysis there are several steps included that can be divided in two categories. First category before the WGCNA analysis contains steps such as sample collection, isolation of crypts, fluorescence activated cell sorting sample preparation, FACS gating and sample collection (for lipidomics analysis by MALDI-IMS and for transcriptomic analysis by microarray technology), analysis by MALDI-IMS, transcriptome profiling microarray. Second category is the WGCNA analysis.

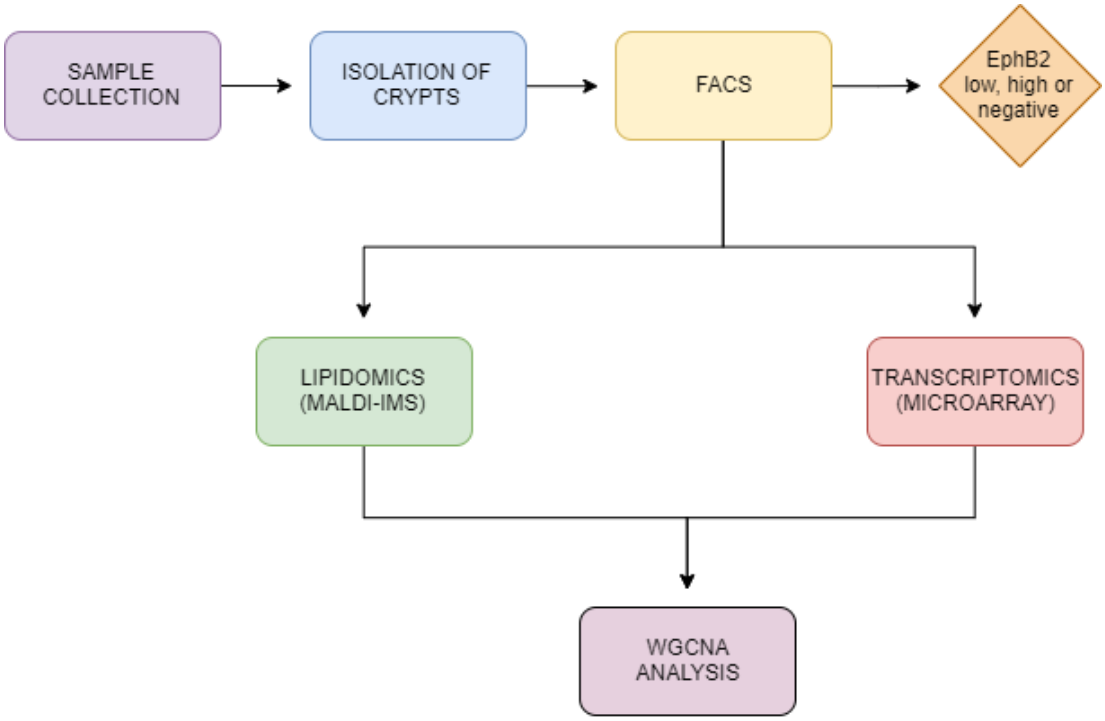

Figure 5: Scheme of the general methodology. Autor's figure.

3.1.1 Sample collection

The human sample collection is from colon endoscopic biopsies, surgical colon fragments or glioblastoma surgical biopsies. All the sample collection for this study was specifically approved by the Ethics Research Committee of the Balearic Islands (IB 2118/13 PI, IB 2291/14 PI, IB 3350/16 PI, IB3626/18 PI). The informed consent was obtained from each patient before sample collection. Surgical colon specimens were obtained by the Department of General and Digestive Surgery at the HUSE (Palma, Spain) and were dissected by the Department of Pathological Anatomy unit.
Fresh primary tumor samples and distant non-tumoral tissue were collected from the same patient.

3.1.2 Isolation of crypts

The protocol was adapted from (Merlos-Suárez et al., 2011). Healthy and tumor colon crypts are isolated from CRC surgical patients derived biopsies. First, samples were washed twice with standard PBS, cut into 1-2 cm fragments using surgical scissors, and incubated with 10 µM DTT then twice dissolved during 5 min in PBS at room temperature. After removing the PBS, fragments were washed and then incubated during 45-60 min in 8 mM EDTA-HBSS at 4 °C. Finally, the supernatant was replaced with 5% FBS-HBSS (staining buffer, SB) and crypts were isolated by vigorously shaking of the colon fractions for 4 min. The supernatant containing the detached crypts was recovered. This step was repeated at least 3 times to obtain a fraction highly enriched in crypts. To sum up, crypts were washed with SB and centrifuged (100×g, 4 °C, 10 min) three times.

3.1.3 Fluorescence activated cell sorting sample preparation

Healthy crypts or epithelial enriched fractions were centrifuged and incubated with TrypLE™ Express Enzyme. Incubations were for 10-15 min at 37 °C with gentle agitation. Mechanical disaggregation by pipetting was performed to ensure high cell disaggregation. Digestion buffer was neutralized by adding FBS and filtered through 100 and 40 µm cell strainer. Single-cell fraction was recovered at 100×g, 10 min, 4 °C. Pellet was resuspended in 250 µl of SB and incubated 45 min at 4 °C with the following mix of antibodies: CD31, CD45, CD11b, CD117, EpCAM, EPHB2. Samples were washed 3 times with SB (100×g, 10 min, 4 °C) and DAPI solution was added before cytometry analysis, DAPI serves as a viability marker.

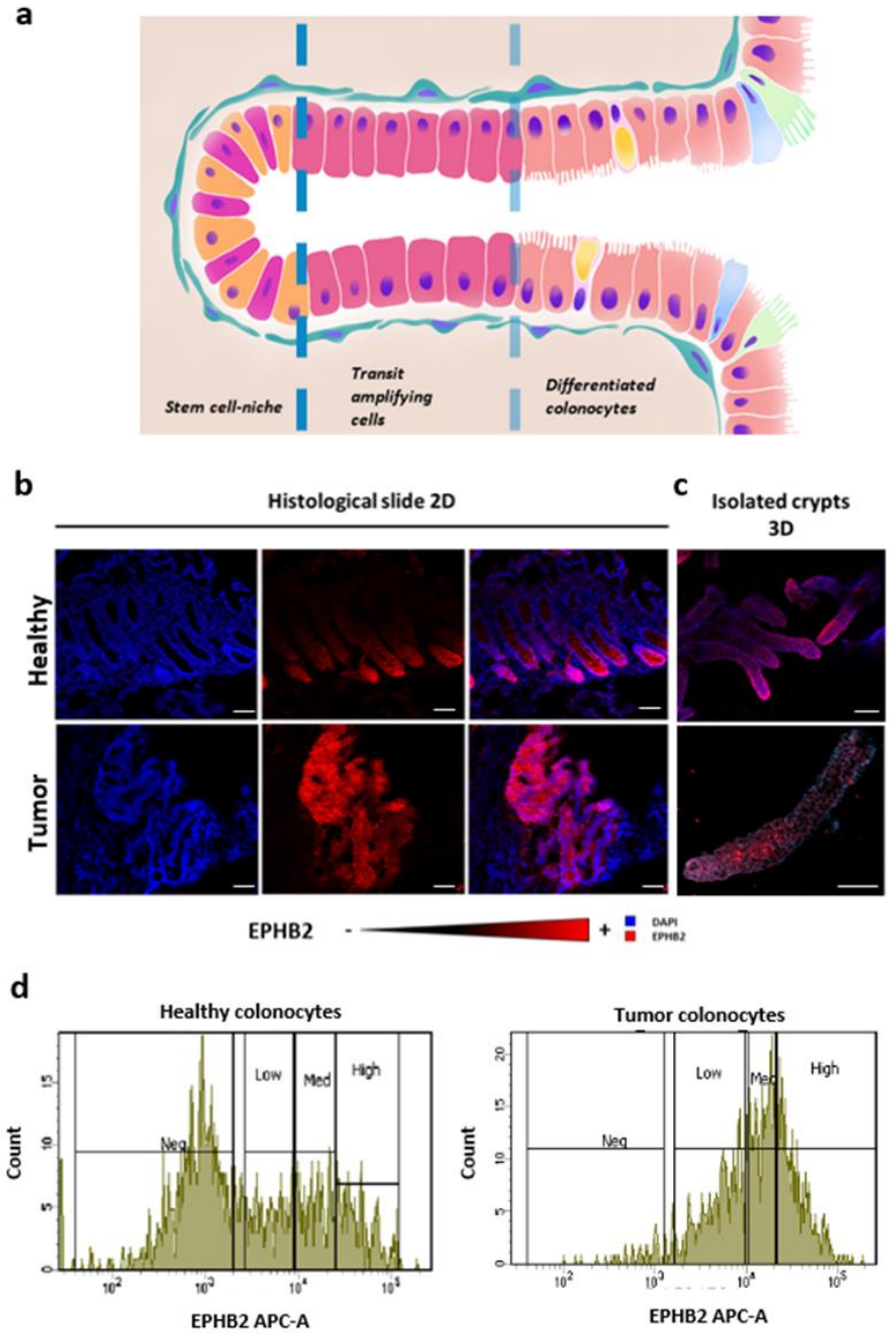

Figure 6: Sorting of colonocytes according to their EphB2 expression used to classify them depending on their differentiation degree. As previously said, levels of EphB2 expression tend to decrease as cells differentiate along the crypt. ***Subfigure a*** *show again the structure of the crypt with the types of cells in each region. Whereas stem cells are located in the bottom, they divide and give place to different colonocyte signatures that will keep differentiating along their way up.* ***Subfigure b*** *are immunofluorescent images of histological cuts from colon tissue samples both healthy and malignant, showing regions of EphB2 expression. It's gradually expressed along the crypt (red tonality) being most abundant at the base of healthy crypts, while the gradual expression is altered in the malignant tissue.* ***Subfigure c*** *is an immunofluorescent image as well, but the difference is that contains isolated crypts; besides that it can be observed the same characteristics as in (b).* ***Subfigure d*** *is a plot from the FACS, where after colonocytes are disaggregated they can be separated into different subpopulations according to EphB2 expression. This case they separated into four groups: high, medium, low and negative but for this project it will be taken into account only the groups of high, low and negative. . (Jung et al., 2011; Merlos-Suárez et al., 2011; Pasquale, 2008)*

3.1.4 FACS gating and sample collection

The combination of antibodies used was to ensure the isolation of highly purified colon epithelial cells, according to their expression of EPHB2.
CD31, CD11b, CD45 were used to mark endothelial, myeloid, and leukocytic populations, respectively. EpCAM was used to mark the colon epithelial cell population, CD117 to mark colon epithelial cell subpopulation, and EPHB2 was used to subdivide the EpCAM+ into different subpopulations based on EPHB2 expression. For each sample, 3 different tubes were prepared for cytometry analysis and gating adjustment: 1. Negative control, 2. The Fluorescence minus one control (FMO) (CD31, 45, 11b, 117 and EpCAM) and 3. Full staining tube (all antibodies including EPHB2). Negative control was used to adjust background and sample fluorescence intensity. FMO to adjust EPHB2 + gating for FACS. Full tube contained the complete sample to process by FACS. Samples were collected in PBS (1.5-ml microtubes), vortex for 1 min, pelleted at 100×g, 10 min, 4 °C. Pellets were stored at -80 °C.
The protocol used in this case is an adaptation of Merlos-Suárez (Merlos-Suárez et al., 2011).
FACS cells processing and sampling preparation for MALDI-IMS analysis contained few additional steps. The EphB2 subpopulations obtained were thawed by heat-shock (1 min, 37 °C) and immediately pelleted by centrifugation (300×g, 15 min, 4 °C). Pellets were kept on ice and homogenized by pipetting in 5-10 µl of PBS prior to apply the sample by dripping 0.5-1 µl drops, one on top of each other on a standard glass slide. Samples from same patient and condition were distributed on the same glass slide, separating into single drops each EPHB2 subpopulation. A total of 48 EphB2 subpopulations were analyzed from healthy and tumor samples.

3.1.5 Lipidomics analysis by MALDI-IMS

Samples were analyzed in the SGiker Lipidomic Service (UPV/EHU, Bilbao, Spain). Briefly, histological sections obtained from patient-derived biopsies or surgical sections were prepared and analyzed by MALDI-IMS as described in Garate *et al* (Garate et al., 2015). DAN (1,5-diaminonaphtalene) was used as matrix for negative detection mode. Matrix was sublimated with the aid of a commercial glass sublimator. Scans were performed in both negative- and positive-ion mode, using a MALDI-LTQ-Orbitrap XL analyzer.
The resulting lipid spectra were aligned using the Xiong method (Xiong et al., 2012) and was normalized to the total ion current. Lipid assignment was based on the comparison between the experimental m/z and the species in the software's database and in the lipid maps database (www. lipidmaps.org).

3.1.6 Transcriptomic analysis by profiling microarray of EphB2+ cell population sorted by FACS

For the transcriptomic analysis, RNA was extracted and then a purity analysis was made. RNA extraction was made from the FACS isolated cells of the CRC patients using RNeasy Micro Kit (Qiagen). Approximately 1-1.5·10^5 FACS cells were collected in 2 ml Rnase-free Eppendorf containing 350 µl of Buffer RLT and vortexed 2 min to ensure cell disruption. RLT volume was adjusted to keep the proportion of 350 µl RLT for every 100 µl of sorted cells. At this point, cells were stored at 80 °C until processed. Then, frozen lysates were incubated at 37 °C in a water bath, until completely thawed and salts are dissolved. Samples were centrifuged at 5000×g, 5 min, 20 °C, to eliminate undissolved materials. The supernatant was transferred to a gDNA Eliminator spin column and manufacturer instructions were followed. Total RNA was eluted of the RNeasy MinElute spin column by adding 14 µl RNase-free water directly to the center of the spin column membrane and centrifuged for 1 min at full speed, 20 °C to elute the RNA. A final volume of 12 µl of concentrated RNA was obtained.

For assessing the purity of the RNA, it was used the nanodrop (multi-well spectrophotometer) using 1 or 2 µl of solution. In any case, the 260/280 ratio was between 1.7 and 2.0 to be acceptable for analysis.

Then for the transcriptome profiling microarray of EphB2+ cell subpopulations sorted by FACS was used Clariom S Pico Assay, human (Affymetrix). This made possible the characterization of gene expression levels of approx. 15,000 viable cells of each of the EphB2 subpopulations isolated (high, low intensity and the negative control). The array was prepared following step by step the manufacturer instructions. The quality of the array data was assessed and normalized using dedicated software.

3.1.7 WGCNA analysis

In this case, WGCNA analysis was performed from microarray data containing expression profiles according to the table I at the annex. In table I is found the clinical information such as age, sex, localization of the tumor, histological type (all being adenocarcinomas), TNM classification and specification of the analysis performed in each case. Given the limited amount of sample and limitations for instance the considerable amount of sample required to perform lipidomics analysis, it wasn't possible to perform all the types of analysis for each patient. That is why table I specifies the analysis performed in each case. Moreover, to compensate the methodological limitation that one patient sample

wasn't enough for doing the two types of analysis (transcriptomics and lipidomics) we used samples from different patients but with same condition tending to maximally homogenize using TNM classification. On the other hand, lipidomic data was normalized to the total ion current (TIC) the transcriptomic data used was given to me already normalized by the standard protocol.

Weighted correlation network analysis contains five key steps that are indicated in the figure 7 below. An overall purpose is to transform genomic data into a scale-free network and use co-expression and others as a measure of connectivity between genes.

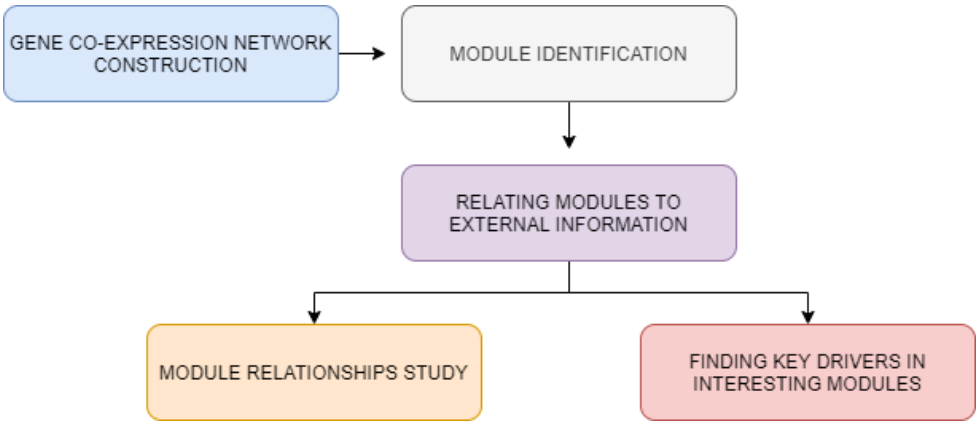

Figure 7: General scheme of weighted gene co-expression network analysis. This methodology consists of several steps. First the construction of the gene co-expression network using interaction patterns among genes, measuring correlations. Then the identification of modules containing the highly interconnected genes, this step allows hierarchical clustering and dynamic tree cuts. Next, there is the possibility of relating modules with external information such as in this case lipid traits (relative normalized expression of lipidic species at the lipid class). Therefore, one can study module relationships thanks to the eigengene networks (that contain eigengene modules which are summary profiles of each module). In addition, allows to find possible biomarkers by assessing intramodular connectivity. Autor's figure.

First was made the gene co-expression network using interaction patterns between genes. Hierarchical clustering was performed using function hclust based on the average method. A similarity matrix among genes was made measured by correlation with values from -1 to 1 using spearman method. Resulting matrix is symmetric with number of rows and the number of columns equal to the number of genes included in the network.

Then this matrix is transformed into an adjacency matrix to be more similar to a scale free network used in systems biology to later assess the number of nodes (genes in the co-expression networks) and their connectivity. For this step it was important to choose the type of network and decide the soft-threshold beta. Types of networks can be unsigned, signed, and signed hybrid. In unsigned networks, genes are similar if they are strongly correlated, whereas in signed network genes are similar only if they are positively

correlated and don't consider genes with opposite correlation. In signed hybrid happens that negative associations are treated as zeros in the similarity matrix. For this study, unsigned network was chosen to be the more appropriate thus it is not omitting negative correlations such as the signed.

Then to decide soft-threshold beta (value used to power the correlation of the genes meant to emphasize high correlations at the expense of low correlations and in some way to reduce the noise of the adjacency matrix) plots were made through WGCNA package called pickSoftThreshold where scale independence and mean connectivity plots (figure 8 and 9) were made and decided that the best values are twelve for healthy condition and twenty-two for tumor condition given the similarity between samples. So, we choose the power which produced higher similarity with a scale-free network.

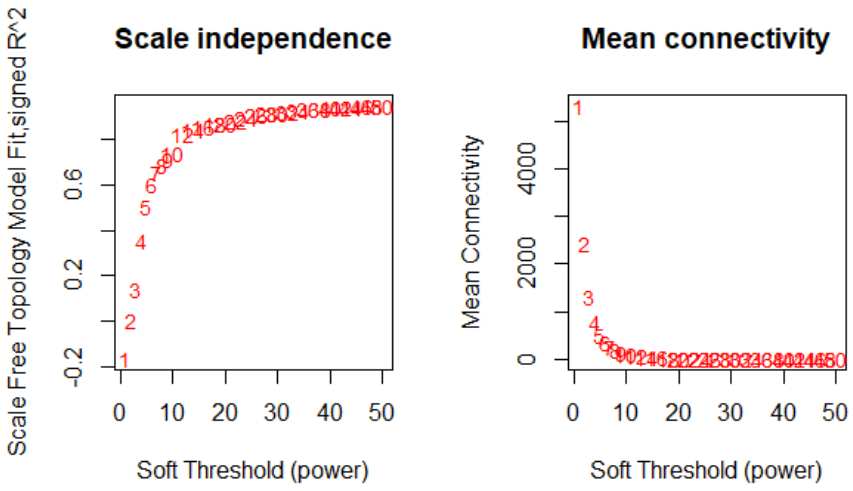

Figure 8: Determination of soft-thresholding power for the healthy condition in weighted gene co-expression network analysis (WGCNA) using scale independence and mean connectivity plots. Autor's figure.

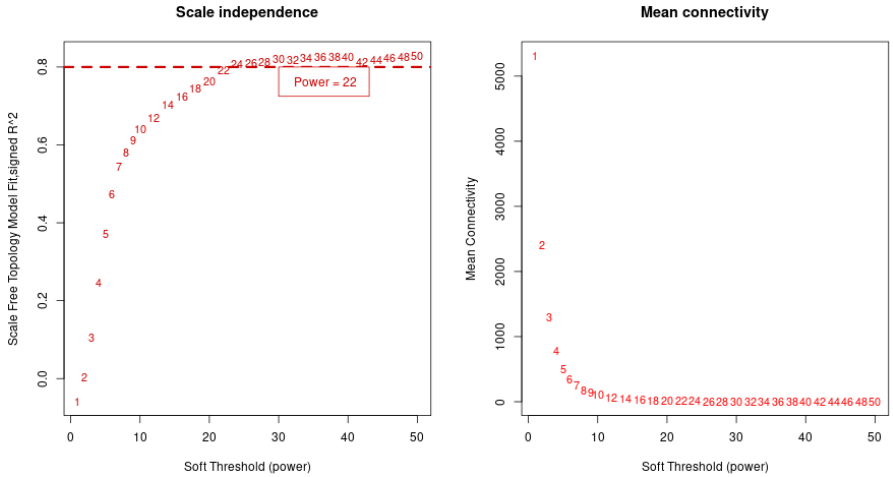

Figure 9: Determination of soft-thresholding power for the tumor condition in weighted gene co-expression network analysis (WGCNA) using scale independence and mean connectivity plots. Autor's figure.

Second big step was module identification to create the clusters of highly interconnected genes. Topological overlap matrix (TOM) was created from the adjacency matrix incorporating both direct and indirect relationships into TOM matrix. Additionally, TOM was transformed to represent dissimilarity among genes. TOM is a measure of similarity where larger values indicate more similarity between two genes, but for hierarchical clustering a measure of dissimilarity is needed so it was transformed to dissimilarity.

Third step allowed to relate modules to lipidomics information by assigning connection to external traits and find biologically interesting modules. All this association analysis can be possible through module eigengenes, that summarize the expression patterns of the module's genes across the samples. Thus, this module-trait associations were made by the correlation among module eigengene and the phenotype, which is the lipid traits, allowing the identification of modules highly correlated to the phenotype. Additionally, for each expression profile was calculated the gene significance (GS) as a value of correlation among expression profile and trait. Module membership (MM) was also defined as the correlation of expression profile and each module eigengene.

Moreover, was made a functional enrichment analysis of the co-expression modules. Where the modules were ranged by number of genes, p-value, Gene Ontology ID code, functionality of each gene as well as their term. Final two steps allowed the study of module relationships using eigengene network concepts and finding key drivers in interesting modules by intramodular connectivity for example, these steps were more interpretative rather than conceptual and will be commented on the results and conclusions.

The scripts made for this type of study followed the tutorial of Peter Langfelder and Steve Horvath (Langfelder & Horvath, 2008). The tutorial is detailed and contains the standard scripts used for this analysis. A total of 16 scripts were applied and modified for this project, they can be found in the complementary folder. Starting with the general script containing the libraries and packages needed such as BiocManager, affycoretools, WGCNA, clusterProfiler, imager, stats, dplyr and DOSE. And then for each condition (healthy "S", or tumoral "T") there is a total of 8 scripts numerated in order and separated by their purpose.

First, *01_prep_data.R* the normalized expression set data was loaded and the two conditions were defined (high and low EphB2 expression), then standard deviation and variance. Data was adjusted so that can be grouped by condition. On the other hand, csv file was loaded containing the lipid traits where rows corresponded to the samples defined by condition and position whilst columns represented the different lipids.

Later, on *02_data_input_cleaning.R* the expression data was transposed because it will be needed for the network constructure. With the function goodSamplesGenes was checked for missing values on genes. Next, samples were clustered to check on outliers using hclust function. Also, trait data was matched with the expression data. Last part of this script saved the expression data under an R.data file that was used for the following steps of the analysis.

Network construction and module detection was made in the *03_network_construction and module detection.R* this step is the most important part on of the analysis. That's why soft thresholding power was chosen from the pickSoftThreshold function. Then co-expression similarity matrix was transformed into an adjacency matrix. To lower the background noise the adjacency matrix was transformed into a Topological Overlap Matrix. Additionally, hierarchical clustering was made to form the dendrogram with hclust function. Next, modules with similar profiles were merged and eigengenes were set to quantify co-expression similarity of entire modules. Finally, the network data was saved.

Scripts *04_relating the consensus modules* and *04_PART_B_load_trait* have on particularity and is that last one is directly ran from *04_relating the consensus modules*. In those scripts was important to relate the modules with lipid traits. First were quantified module-trait associations by correlating eigengenes with lipid traits. Module trait relationship heatmap plot was made. Then MM vs. GS scatterplots were made for each module to quantify similarities of all genes in each module from

04_PART_B_load_trait. Furthermore, data frame was created containing information like probe names, gene ID, module color, GS, MM and p-Value of each module and saved on an excel file.

05_interfacing network_GO was made the enrichment by Gene Ontology relating genes from each module to the genes that appear in GO database. Information like p-Value, GO ID's or term name was saved on a file to later be used in tables 3 and 4.

Finally remaining scripts *06_network visualization* and *07_cytoscape.* First one was made to create plots of the network like network heatmap plot of all genes or from a selection of genes, eigengene dendrogram or eigengene adjacency heatmap. Last one, was made to create files for posterior export of the network for visualization on Cytoscape software.

3.2 Organoid culture

The protocol I used for organoid culture contains several important steps to follow. First and foremost, important is the sample collection, then the isolation of crypts from the sample. Followed by culture of crypts, organoid differentiation, and organoid passing. Then depending on the purpose, it can be followed by incubation with deuterated fatty acids and fluorescence-activated cell sorting. This part of organoid culture follows some same steps as on the WGCNA analysis methodology, and they are sample collection, isolation of crypts, fluorescence activated cell sorting preparation and FACS gating and sample collection.

3.2.1 Sample collection

As well as in the previous case, sample collection follows the same protocol. Sample collection is from colon endoscopic biopsies, surgical colon fragments or glioblastoma surgical biopsies. Surgical colon specimens were obtained by the Department of General and Digestive Surgery at the HUSE (Palma, Spain) and were dissected by the Department of Pathological Anatomy unit.

Fresh primary tumor samples and distant non-tumoral tissue were collected from the same patient in order to proceed with the organoid culture steps.

3.2.2 Isolation of crypts

The protocol was adapted from (Merlos-Suárez et al., 2011). Healthy and tumor colon crypts are isolated from CRC surgical patients derived biopsies.

Samples were washed twice with standard PBS, cut into 1-2 cm fragments using surgical scissors, and incubated with DTT then twice

dissolved during 5 minutes in PBS at room temperature. After removing the PBS, fragments were washed and then incubated during 45-60 minutes in EDTA-HBSS at 4 °C. Finally, the supernatant was replaced with 5% FBS-HBSS (staining buffer, SB) and crypts were isolated by vigorously shaking of the colon fractions for 4 minutes. The supernatant containing the detached crypts was recovered. This step was repeated at least 3 times to obtain a fraction highly enriched in crypts. Finally, crypts were washed with SB and centrifuged (100×g, 4 °C, 10 min) three times.

3.2.3 Obtention of patient derived organoids from colon crypts

This part of the methodology was made in laminar flow cabinet. First supernatant was removed and resuspended with Advanced DMEM/F12 medium supplemented with Primocin, GlutaMax, HEPES and FBS. Then mixed the isolated crypts with Matrigel Matrix at a rate of 10 crypts per µl of Matrigel at 4ºC using a pipette. Next, using a P-100 pipette, 25 µl of mixture drops (5 drops approx./ well) were seed into a 24-well culture plate avoiding bubble formation. After an incubation of 15 min at 37ºC upside-down to achieve Matrigel solidification, 300 µl of crypt medium was added per well. Finally, the plate was left in a humidified cell culture incubator (37ºC and 5% CO_2) and only taken out when photos are made periodically at the cell observer microscope. Crypt-culture medium was replaced every 3 days (*containing Wnt3a-conditioned medium, Primocin, GlutaMax, HEPES, B-27, Nicotinamide, N-Acetyl-L-cysteine, human EGF, human Noggin, human Big Gastrin, LY2157299 (TGFBR1 inhibitor), SB202190 (p38 MAP kinases inhibitor), IGF-1, FGF2*).

3.2.4 Organoid differentiation

To achieve organoid differentiation after 4-6 days in culture the growth factors such as PGE2, Nicotinamide, SB202190 were omitted. Wnt3a-conditioned medium was reduced to 5% and then omitted 48 hours prior organoid harvesting. This protocol is an adaptation from T. Sato et al. 2011 and P. Jung et al. 2011 527,528.

3.2.5 Organoid passing

Performed once the organoids were stable and not growing anymore. Culture medium was replaced by 300 µl of Cell Recovery solution (that degrades the Martigel Matrix) and incubated one hour at 4ºC. Content of the wells was collected in 15 ml tube and pipetted several times gently. To remove the debris the solution in the 15 ml tube was centrifuged 40xg, 5-10 min at 4ºC. Later the supernatant was discarded. Organoid's pellet was resuspended with TrypLE™ Express and incubated 10 min at 37ºC. Then it was inactivated by adding Advanced DMEM/F12 medium supplemented with GlutaMax, HEPES and FBS. In addition, centrifugation at 1.200 rpm, 10 min at 4ºC was made to remove the supernatant. The remaining pellet was

resuspended in Advanced DMEM/F12 supplemented medium with Primocin, GlutaMax, HEPES and FBS.
Depending on the experimental purpose, organoids were seeded back following instructions in 3.2.3 paragraph, or they were frozen.

3.2.6 Incubation with deuterated fatty acids

Depending on the experimental purpose in each case, colon organoids were incubated with 10 µM of arachidonic acid -d8 (AA-d8) before organoid harvest by the Matrigel Matrix disruption using 300 µl per well of Cell Recovery solution.

3.2.7 Fluorescence activated cell sorting preparation

Healthy crypts or epithelial enriched fractions were centrifuged and incubated with TrypLE™ Express Enzyme. Incubations were for 10-15 min at 37 °C with gentle agitation. Mechanical disaggregation by pipetting was performed to ensure high cell disaggregation. Digestion buffer was neutralized by adding FBS and filtered through 100 and 40 µm cell strainer. Single-cell fraction was recovered at 100×g, 10 min, 4 °C. Pellet was resuspended in 250 µl of SB and incubated 45 min at 4 °C with the following mix of antibodies: CD31, CD45, CD11b, CD117, EpCAM, EPHB2. Samples were washed 3 times with SB (100×g, 10 min, 4 °C) and DAPI Solution was added before cytometry analysis.

3.2.8 FACS gating and sample collection

The combination of antibodies used was to ensure the isolation of highly purified colon epithelial cells, according to their expression of EPHB2.
CD31, CD11b, CD45 were used to mark endothelial, myeloid, and leukocytic populations, respectively. EpCAM was used to mark the colon epithelial cell population, CD117 to mark colon epithelial cell subpopulation, and EPHB2 was used to subdivide the EpCAM+ into different subpopulations based on EPHB2 expression. For each sample, 3 different tubes were prepared for cytometry analysis and gating adjustment: 1. Negative control, 2. The Fluorescence minus one control (FMO) (CD31, 45, 11b, 117 and EpCAM) and 3. Full staining tube (all antibodies including EPHB2). Negative control was used to adjust background and sample fluorescence intensity. FMO to adjust EPHB2 + gating for FACS. Full tube contained the complete sample to process by FACS. Samples were collected in PBS (1.5-ml microtubes), vortex for 1 min, pelleted at 100×g, 10 min, 4 °C. Pellets were stored at -80 °C.
The protocol used in this case is an adaptation of Merlos-Suárez (Merlos-Suárez et al., 2011).

4. Results and discussion

4.2 WGCNA analysis

Colorectal cancer is a leading cause of cancer death worldwide, its late detection contributes for becoming one of the deadliest. Hence the importance of studying the impact of colorectal cancer in the lipid composition during the differentiation process of colonocytes. Given that lipidome of colonocytes is sensible to changes occurring during differentiation from stem cell to matured colonocyte. With the purpose to get more knowledge on the regulatory mechanisms involving lipid composition, WGCNA analysis is the best suitable approach. Since it has already been used in cancer and genetic studies to explore complex relationships between genes and phenotypes by transforming gene expression data into co-expression modules.

Before getting deep with WGCNA analysis, it is important to justify why both healthy and tumor conditions where studied. As said, the purpose is studying how lipid composition changes though differentiation. Thus, changes in the amount of sphingolipid type sphingomyelin where observed (figure 10). Overall composition between healthy and tumor condition remained similar but there are changes among high and low conditions which stand for stem cells and mature colonocytes (FD) respectively. Being higher levels of SM both in stem cells for healthy and tumoral condition. Quite reverse are the levels of SM in mature colonocytes being lower.

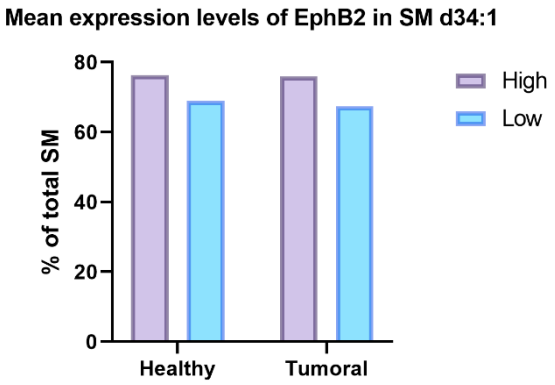

Figure 10: Histogram of mean EphB2 expression levels regarding sphingomyelin total composition. High stands for high expression of EphB2 while low stands for low expression of EphB2. As described in the introduction, regions with high EphB2 expression correspond to stem cells like. Whereas regions with low EphB2 expression correspond to fully diferentiated or mature colonocytes. Autor's figure.

4.2.1 Healthy condition

As already said, this type of analysis allows to describe correlation patterns between genes from microarray samples, finding clusters and relating modules with external traits such as lipid expression levels.

For the data aggregation, hierarchical clusters were created and representing them into dendrograms containing the samples. This was useful to check for outliers and the possible linkages between clustering. As we can see in figure 11 samples containing high levels of EphB2 expression are clustered all together and have more similarity than samples of low EphB2 expression.

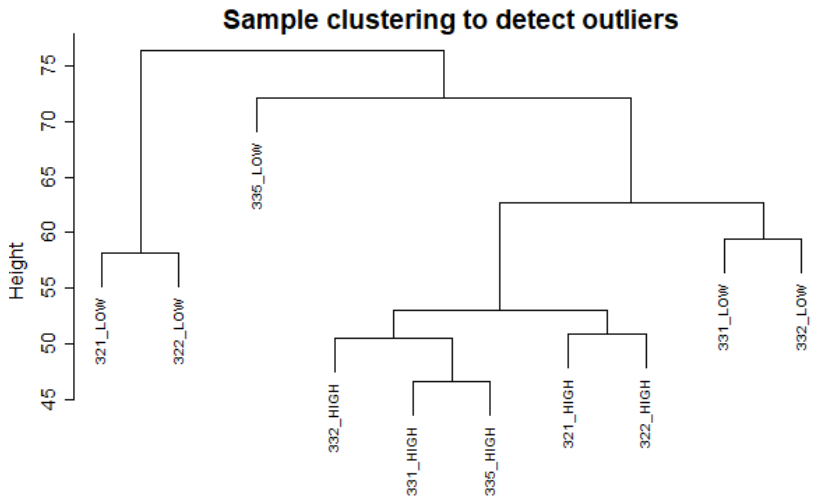

Figure 11: Hierarchical cluster represented by dendrograms based on the "average" method. This dendrogram created by hclust function, indicates the similarity between samples merging them into clusters. Autor's figure.

Then the first set of modules obtained by the Dynamic Tree Cut algorithm. Correlated modules were merged, and each colored row represents a color-coded module containing a group of highly connected genes as seen in figure 12.

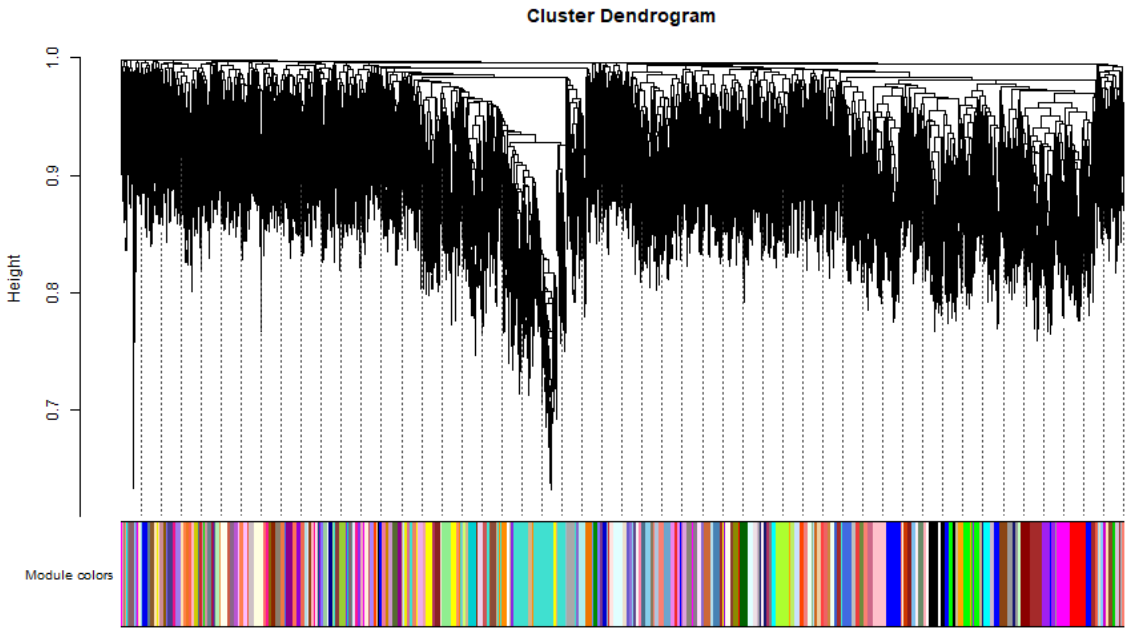

Figure 12: Identification of gene co-expression modules via hierarchical average linkage clustering (Dynamic tree cut algorithm was used to identify modules, and genes in the same branch could be assigned to different modules). At the top is the hierarchical clustering plot whilst at the bottom are represented the modules with different colors respectively. Autor's figure.

Heatmap was made to see the interactions between modules and lipid traits (relative expression of lipid species normalized at the lipid class) (figure 13) a result of summarizing each module's gene co-expression by eigengene and then calculated the correlations of each eigengene with the lipid traits. It allows an easy identification of expression sets (modules) that are highly correlated to the phenotype.

In this case, modules correlating with high expression of EphB2 are strong positively correlated with our lipid of interest that is SM d34:1. Whereas, modules correlating with low expression of EphB2 are negatively correlated with this type of sphingolipid. Moreover, there are two visible patterns where the top modules have negative correlation with the first half of the lipid traits and positive correlation with the second half. On the contrary the central and lower modules present positive correlation with the first half of lipid traits and negative correlation with a minor portion of the second half.

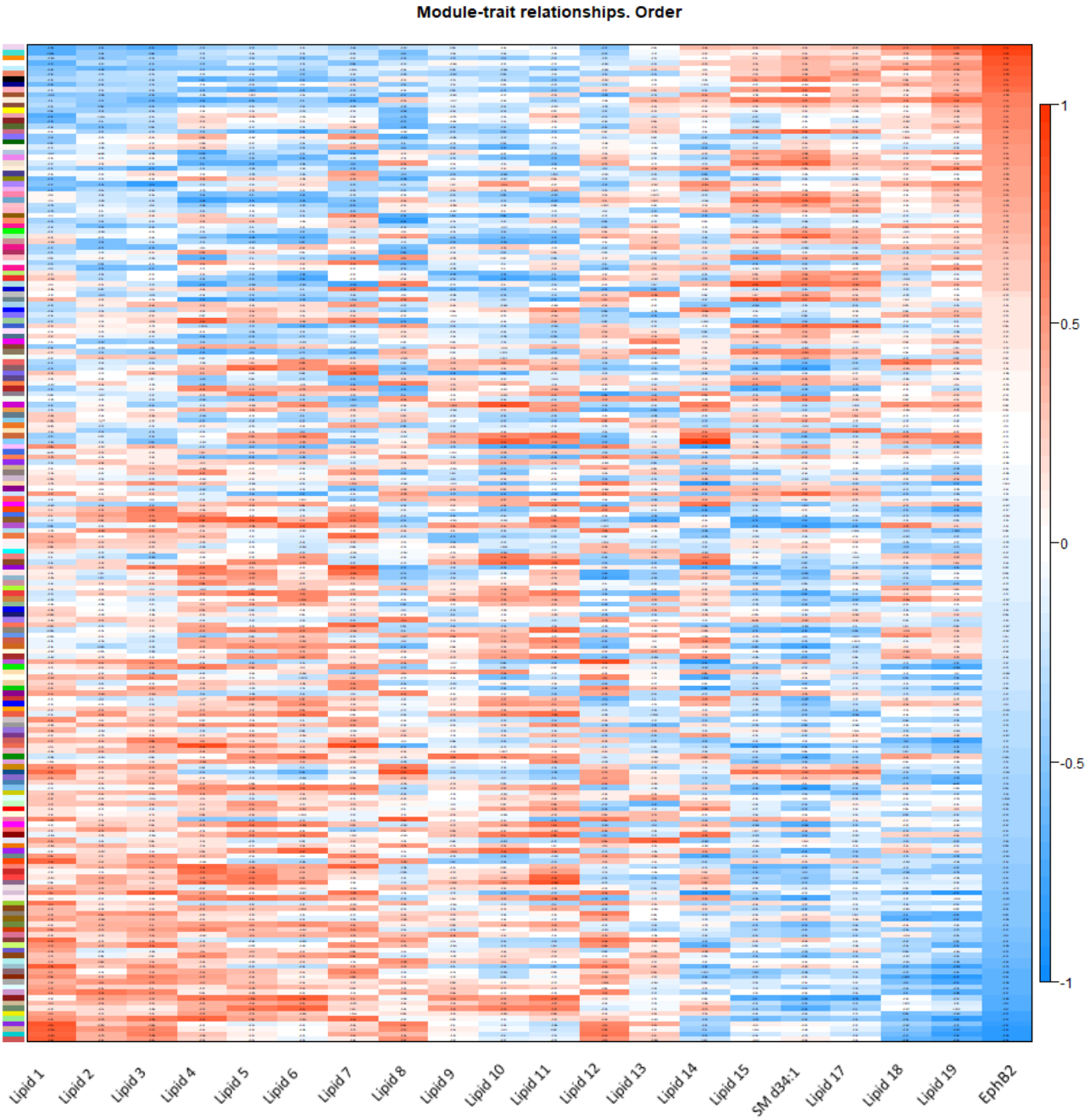

Figure 13: Heatmap of the correlation between eigengene modules and lipid traits (normalized relative expression of lipid species to lipid class). Modules are present in the y axis whereas x axis contains the lipid traits. Thus, each row corresponds to a module eigengene, and each column corresponds to a lipid species. Every cell contains the corresponding correlation and p-value. Values of correlation range from -1 to 1 using a color scale blue for negatively correlated and red for positively correlated. Last column on the right named EphB2 stands for the levels of EphB2 expression red for high expression and blue for low expression. Autor's figure.

Selection of the best modules correlated to SM d34:1 was made according to correlation and p-value, assessed by module membership vs. gene significance plots (figure 14). Top four

modules significatively correlated were lightpink4 with correlation value of 0.69, mediumorchid3 with correlation value of 0.64, brown3 with correlation value of 0.59 and magenta with correlation value of 0.39. Each plot has the module genes distributed according to MM and GS, thus genes located in the upper left part are the ones more significant and are usually called module regulator genes.

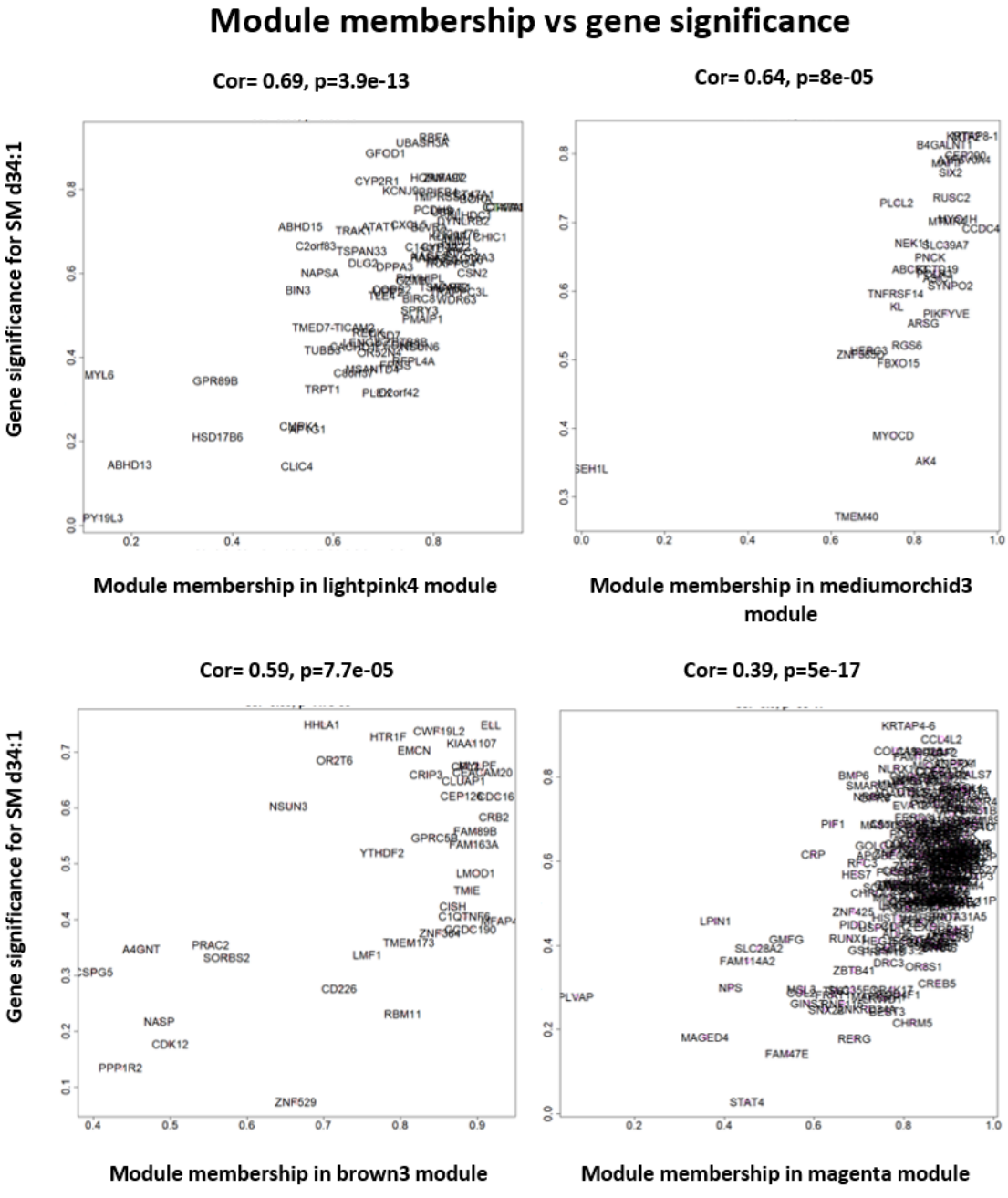

Figure 14: Plot of module membership (MM) vs gene significance (GS) for each of the top four modules. Y axis represents gene significance for SM d34:1 whereas X axis represents module membership of the genes in the specific module. Values of correlation and p-value can be found below the title of each plot. Autor's figure.

Enrichment analysis by Gene Ontology (GO) was made as seen in table 3. Its purpose is to extract conclusions from expressed genes in modules. By determining in which biological processes (BP), cellular components (CC) or molecular function (MF) are participating the genes.

In this case, module lightpink4 contains genes involved mostly in biological processes, such as negative regulation of gene expression, fatty acid elongase complex, transmembrane-ephrin receptor activity, ephrin activity, inactivation of MAPK activity and negative regulation of calcium-dependent cell-cell adhesion.

Whereas mediumorchid3 contains genes involved mostly in molecular functions regarding lipid metabolism regulation, for example inositol tetrakisphosphate 1-kinase activity, inositol phosphorylation and fatty acid elongase complex.

Module brown3 genes are involved in biological processes and molecular functions such as inactivation of MAPK activity and negative regulation of calcium-dependent cell-cell adhesion.

Genes from magenta module are majority involved in biological processes regarding cell division like interkinetic nuclear migration, nucleosome and chromatin assembly and negative regulation of centriole replication.

Table 3: Significantly enriched functional terms of genes in most significant modules selected by correlation and p-values. This table contains information regarding four modules which are lightpink4, mediumorchid3, brown3 and magenta. For each there is information about p-value, GO identification code and their ontology as well as functional terms. Biological processes (BP), cellular components (CC) or molecular function (MF).

Module name	p-value	GO ID	Ontology	Term name
Lightpink4	0.00016	1902498	BP	regulation of protein autoubiquitination
	0.00053	0000956	BP	nuclear-transcribed mRNA catabolic process
	0.00056	0000932	CC	P-body
	0.00057	0051823	BP	regulation of synapse structural plasticity
	0.00057	0070973	BP	protein localization to endoplasmic reticulum exit site
	0.0011	0061001	BP	regulation of dendritic spine morphogenesis
	0.0013	0010629	BP	negative regulation of gene expression
	0.0015	1900452	BP	regulation of long-term synaptic depression
	0.0016	0050807	BP	regulation of synapse organization
	0.0021	0006000	BP	fructose metabolic process

Mediumorchid3	0.0021	0015821	BP	methionine transport
	0.0021	0034514	BP	mitochondrial unfolded protein response
	0.0021	0004742	MF	dihydrolipoyllysine-residue acetyltransferase activity
	0.0021	0015191	MF	L-methionine transmembrane transporter activity
	0.0021	0047325	MF	inositol tetrakisphosphate 1-kinase activity
	0.0021	0052725	MF	inositol-1,3,4-trisphosphate 6-kinase activity
	0.0021	0052726	MF	inositol-1,3,4-trisphosphate 5-kinase activity
	0.0042	0052746	BP	inositol phosphorylation
	0.0042	0009923	CC	fatty acid elongase complex
	0.0042	0000825	MF	inositol tetrakisphosphate 6-kinase activity
Brown3	0.00037	0005005	MF	transmembrane-ephrin receptor activity
	0.00043	0005003	MF	ephrin receptor activity
	0.00098	0000188	BP	inactivation of MAPK activity
	0.0022	0046588	BP	negative regulation of calcium-dependent cell-cell adhesion
	0.0022	0060035	BP	notochord cell development
	0.0022	0005900	CC	oncostatin-M receptor complex
	0.0022	0070611	MF	histone methyltransferase activity (H3-R2 specific)
	0.0022	0070612	MF	histone methyltransferase activity (H2A-R3 specific)
	0.0022	1990275	MF	preribosome binding
	0.0044	0034970	BP	histone H3-R2 methylation
Magenta	0.00058	0016831	MF	carboxy-lyase activity
	0.00062	0071393	BP	cellular response to progesterone stimulus
	0.002	0022027	BP	interkinetic nuclear migration
	0.0022	0006334	BP	nucleosome assembly
	0.0031	0016830	MF	carbon-carbon lyase activity
	0.0038	0031497	BP	chromatin assembly
	0.0042	0046600	BP	negative regulation of centriole replication
	0.0042	0032356	MF	oxidized DNA binding
	0.0043	0097110	MF	scaffold protein binding
	0.0048	0031225	CC	anchored component of membrane

4.2.2 Tumor condition

For the tumor condition data aggregation were made by hierarchical clusters into dendrograms. This was useful to check for outliers and the possible linkages among clustering. Figure 16 shows that all samples are tumoral. But on the contrary of the healthy condition, tumor samples don't cluster under similar expression of EphB2 levels. Instead, they follow a rather particular distribution where most clusters contain samples from the same patient with both low and high expression of EphB2.

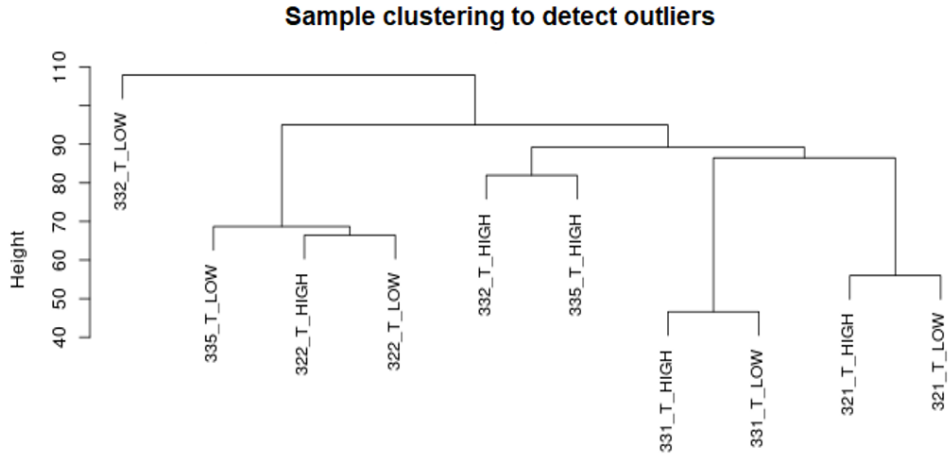

Figure 15: Hierarchical cluster represented by dendrograms based on the "average" method. This dendrogram created by hclust function, indicates the similarity between samples merging them into clusters. Autor's figure.

Modules obtained by the Dynamic Tree Cut algorithm. The correlated modules were merged, and each colored row represents a color-coded module containing a group of highly connected genes. Compared to the healthy condition, there are less modules in tumor condition as we can see in figure 12 and 17.

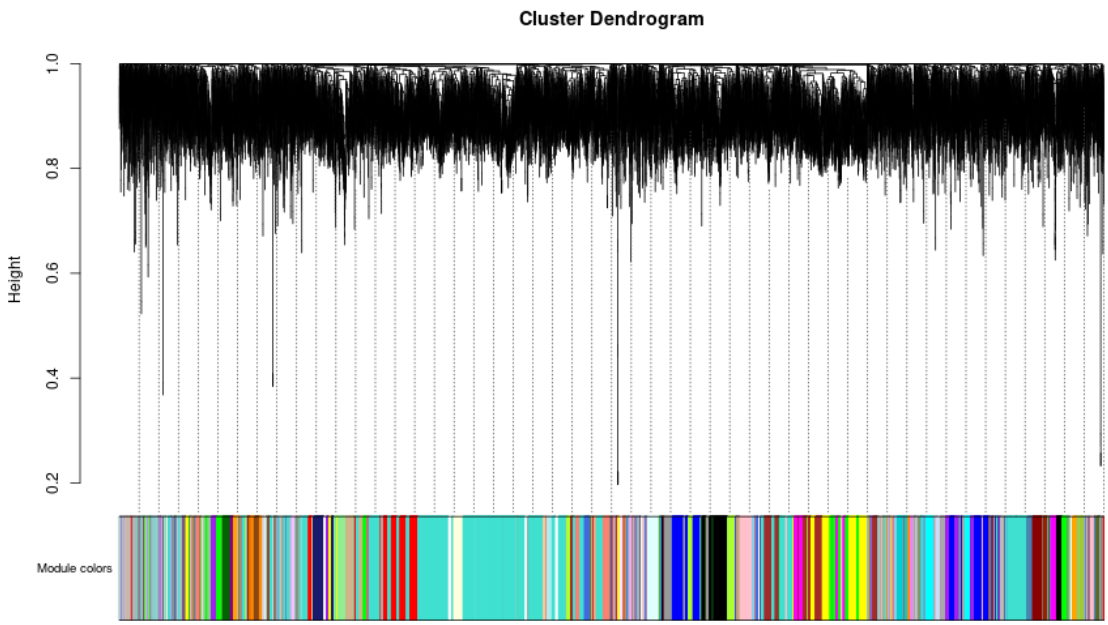

Figure 16: Identification of gene co-expression modules via hierarchical average linkage clustering by Dynamic tree cut algorithm. At the top is the hierarchical clustering plot whilst at the bottom are represented the modules with different colors. Autor's figure.

Next, a heatmap was made to see the interactions between modules and lipid traits which are the relative expression of lipid species normalized at the lipid class (figure 18). Each module gene co-expression is summarized by eigengene, then calculated the correlations of each eigengene with the lipid traits. Helping to identify modules highly correlated to the phenotype. In particular, tumor condition shares similarities with healthy condition regarding modules correlating with high expression of EphB2 and at the same time positively correlated with SM d34:1.

As we can see there are fewer modules in this heatmap compared to the healthy condition. And correlations are slightly less intensive regarding color. More modules are found with correlation values around zero.

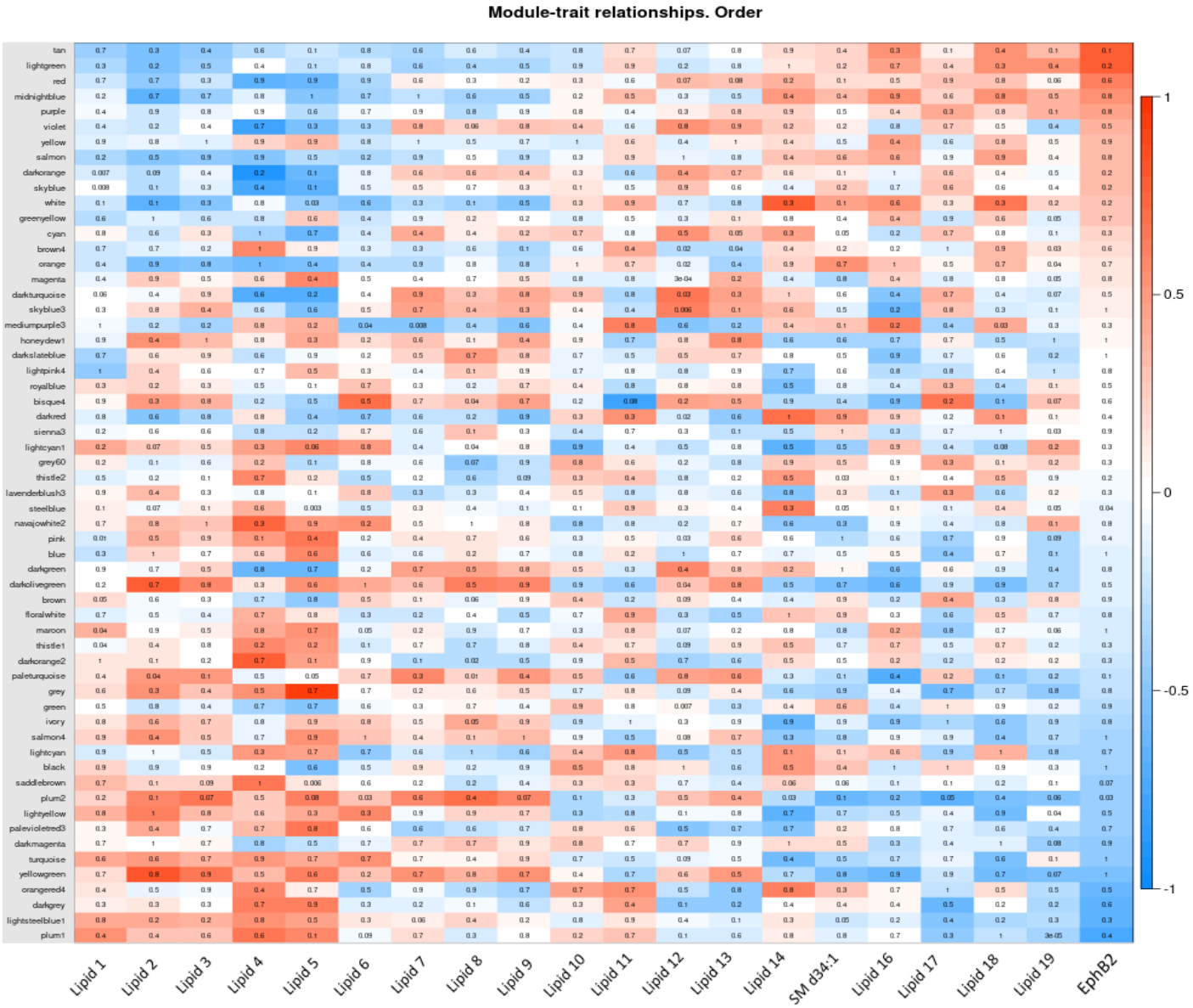

Figure 17: Heatmap of the correlation between eigengene modules and lipid traits (normalized relative expression of lipid species to lipid class). Modules are present in the y axis whereas x axis contains the lipid traits. Each row corresponds to a module eigengene, and each column to a lipid species. Every cell contains the corresponding correlation value. Values of correlation range from -1 to 1 using a color scale blue for negatively correlated and red for positively correlated. Last column on the right named position stands for the levels of EphB2 expression red for high expression and blue for low expression of EphB2. Autor's figure.

Top four best modules correlated to SM d34:1 according to significant correlation and p-value, assessed by module membership vs. gene significance plots (figure 19) were paleturquoise with correlation value of 0.6, orangered4 with correlation value of 0.49, darkorange2 with correlation value of 0.39 and yellowgreen with correlation value of 0.39.

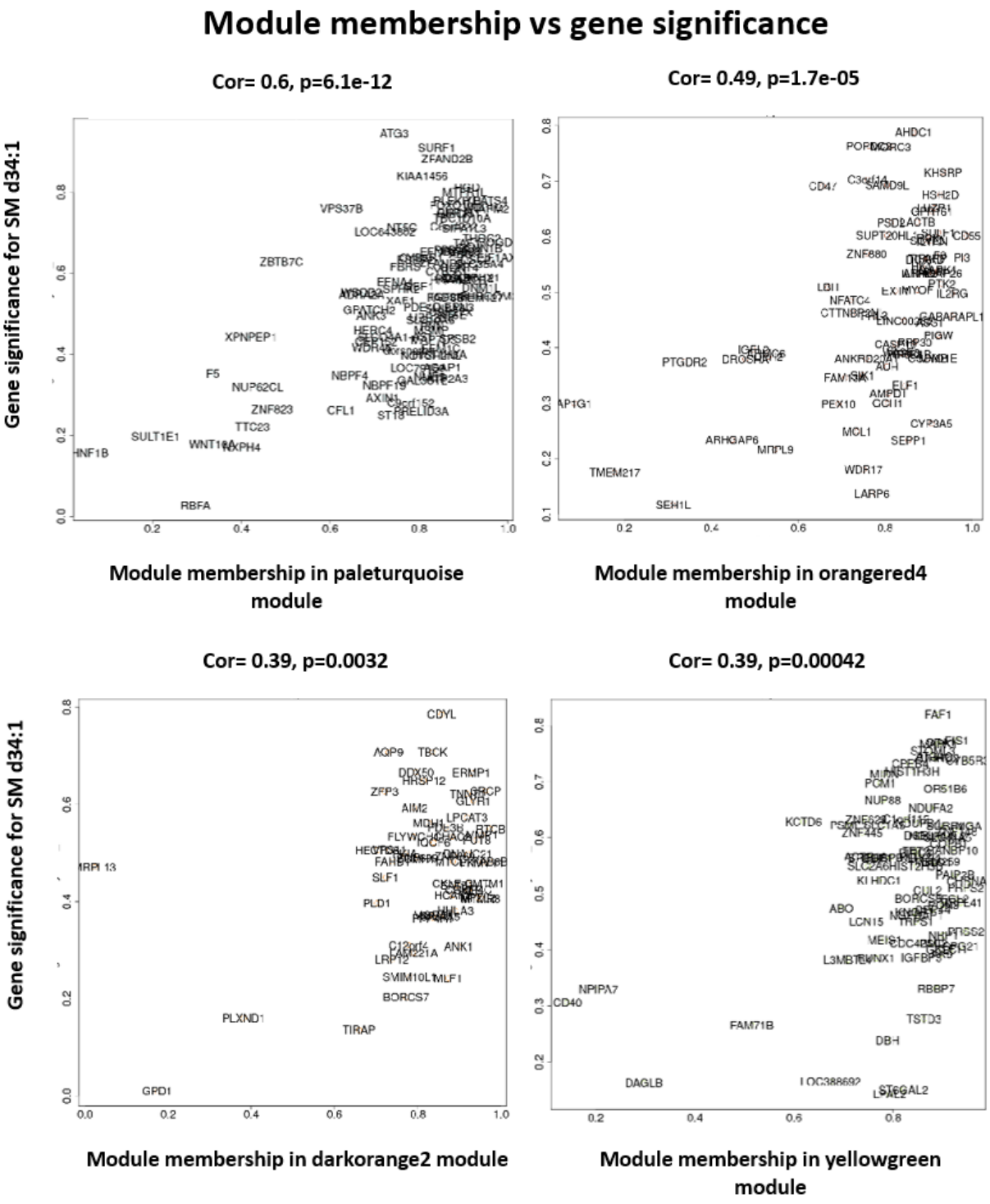

Figure 18: Plot of module membership (MM) vs gene significance (GS) for each of the top four modules. Y axis represents gene significance for SM d34:1 whereas X axis represents module membership of the genes in the specific module. Values of correlation and p-value can be found below the title of each plot. Autor's figure.

For tumor condition, enrichment analysis by Gene Ontology (GO) was also made as seen in table 4.

In this case, module paleturquoise contains genes involved mostly in biological processes participating in cell division like mitotic centrosome separation, mitotic spindle, RNA polymerase II

regulatory region sequence-specific DNA binding and centrosome separation.

However, orangered4 contains genes involved in a mix of molecular functions and biological processes. For example, regulation of B cell apoptotic process, ionotropic glutamate receptor complex and glutamate receptor activity.

Module yellowgreen genes are involved in biological processes such as transcription processes. Containing genes with functions like cellular metabolic process, tRNA 3'-end processing, RNA 5'-end processing, transferase activity and chromatin DNA binding.

Genes from darkorange2 module are also a mix of BP and MF like negative regulation of transcription by competitive promoter binding, AMP metabolic process and intracellular signal transduction.

Table 4: Significantly enriched functional terms of genes in most significant modules selected by correlation and p-values. This table contains information regarding four modules which are paleturquoise, orangered4, yellowgreen and darkorange2. For each there is information about p-value, GO identification code and their ontology as well as functional terms. Biological processes (BP), cellular components (CC) or molecular function (MF).

Module name	p-value	GO ID	Ontology	Term name
Paleturquoise	0.0009	0060322	BP	Head development
	0.0017	0015631	MF	Tubulin binding
	0.0019	0000235	CC	Astral microtubule
	0.0022	2000179	BP	Positive regulation of neural precursor cell proliferation
	0.0023	0007100	BP	Mitotic centrosome separation
	0.0027	0051299	BP	Centrosome separation
	0.0029	0030900	BP	Forebrain development
	0.0029	0072686	CC	Mitotic spindle
	0.0040	0000977	MF	RNA polymerase II regulatory region sequence-specific DNA binding
	0.0042	0001012	MF	RNA polymerase II regulatory region DNA binding
Orangered4	0.00053	0008328	CC	Ionotropic glutamate receptor complex
	0.00098	0030207	BP	Chondroitin sulfate catabolic process
	0.0014	0004970	MF	Ionotropic glutamate receptor activity
	0.0016	0050655	BP	Dermatan sulfate proteoglycan metabolic process
	0.0016	0004549	MF	tRNA-specific ribonuclease activity

	0.0018	0002902	BP	Regulation of B cell apoptotic process
	0.0027	0035235	BP	Ionotropic glutamate receptor signaling pathway
	0.0030	0008066	MF	Glutamate receptor activity
	0.0033	0001783	BP	B cell apoptotic process
	0.0033	0032281	CC	AMPA glutamate receptor complex
Yellowgreen	0.00041	0042780	BP	tRNA 3’-end processing
	0.00066	0006807	BP	nitrogen compound metabolic process
	0.0016	0044237	BP	Cellular metabolic process
	0.0018	0099116	BP	tRNA 5’-end processing
	0.0020	0051580	BP	Regulation of neurotransmitter uptake
	0.0021	0031123	BP	RNA 3’-end processing
	0.0023	0098810	BP	Neurotransmitter reuptake
	0.0027	0031490	MF	Chromatin DNA binding
	0.0030	0016740	MF	Transferase activity
	0.0033	0000966	BP	RNA 5’-end processing
Darkorange2	0.0003	0007264	BP	Small GTPase mediated signal transduction
	0.00043	0010944	BP	Negative regulation of transcription by competitive promoter binding
	0.00066	0046033	BP	AMP metabolic process
	0.0013	0019003	MF	GDP binding
	0.0016	0017137	MF	Rab GTPase binding
	0.0017	0035556	BP	Intracellular signal transduction
	0.0018	0046332	MF	SMAD binding
		0032925	BP	Regulation of activin receptor signaling pathway
		0045921	BP	Positive regulation of exocytosis
		0032400	BP	Melanosome localization

4.3 Organoid culture

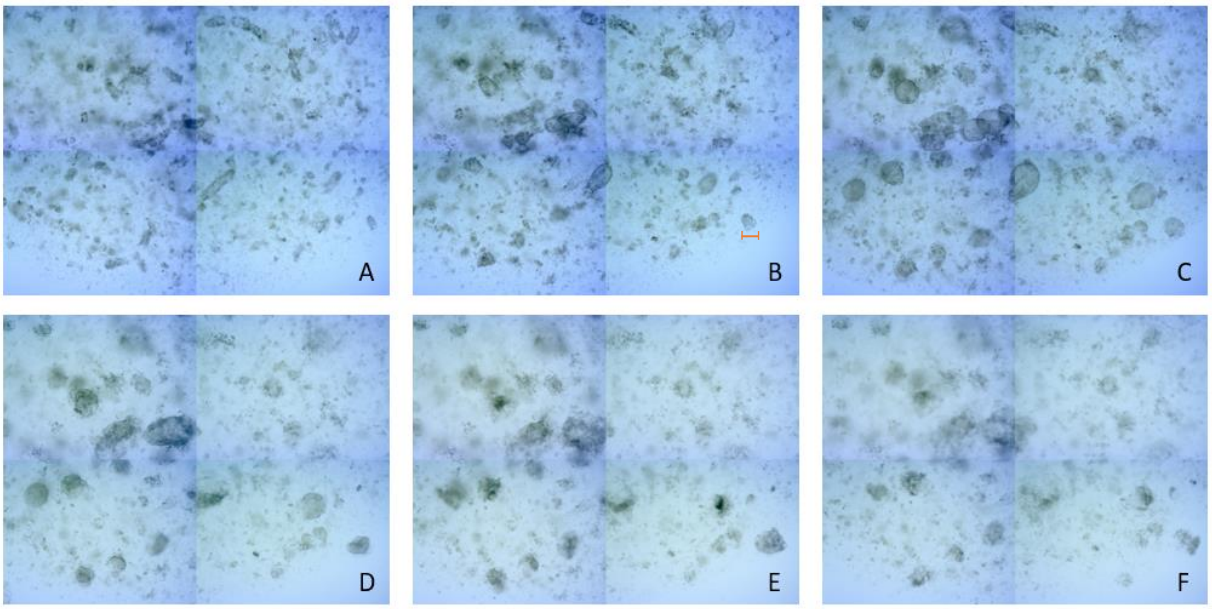

Figure 19: Organoid culture images taken from Carl Zeiss Cell Observer at 5X. A) Organoids of healthy condition 30 min after being seed. B) Organoids of healthy condition 24 hours after, additionally the scale is of 200 µm. C) Organoids of healthy condition 72 hours after. D) Organoids of healthy condition right after differentiation (0 hours). E) Organoids of healthy condition 72 hours after differentiation. F) Organoids of healthy condition 7 days after differentiation.

Given the fact that organoids require a considerable number of weeks and that especially healthy condition colon organoids are susceptible to contamination, we couldn't obtain lipidomic results. Moreover, there were less operations, thus less samples during the period I did my internship because of COVID restrictions in the hospital. Luckily, we successfully cultured healthy condition colon organoids and even had the chance to apply the differentiation process on them as seen in figure 19. All images in this figure are from the same region allowing to see how organoids change though the culture process.

In this case, 24 hours after the seed, some crypts had already started to slightly grow into forming organoids. It is not until 72 hours later that more significant growth was made. When the differentiation process started, organoids changed their aspect by becoming darker in the interior and exterior surface. To remember, in differentiation process crypt culture medium wasn't the same as the culture used for organoid growth, factors like PGE2, Nicotinamide and SB202190 were omitted. Wnt3a-conditioned medium was reduced to 5% and then omitted 48 hours prior organoid harvesting. Changes in the organoids suggested that the differentiation process was effective. Additionally, as organoids become more differentiated they tend to look like they are starting to die, because their surface looks irregular and even degraded. Most of the times it's an optic illusion because of the 3D nature of the organoids combined with the fact that microscopes show the image in 2D at a particular level, thus other levels create shadows contributing to this darkening.

5. Conclusion

Data obtained from MALDI-IMS demonstrate the existence of a decreasing sphingomyelin gradient through colonocyte differentiation. Levels of sphingomyelin are similar between healthy and tumor condition but different depending on if it is the upper or lower region of the crypt.

Through WGCNA analysis is demonstrated that expression profile between healthy and tumor is different, because of the different modules and genes expressed in both cases.

In contrast, lipidomic results to evaluate the incorporation of deuterated fatty acids in colon organoids weren't obtained because of time and sample limitations. Nonetheless, I performed a successfully organoid culture observing changes after causing the differentiation process.

6. References

Ahmed, R. L., Schmitz, K. H., Anderson, K. E., Rosamond, W. D., & Folsom, A. R. (2006). The metabolic syndrome and risk of incident colorectal cancer. *Cancer*, *107*(1), 28–36. https://doi.org/10.1002/cncr.21950

Araghi, M., Soerjomataram, I., Bardot, A., Ferlay, J., Cabasag, C. J., Morrison, D. S., De, P., Tervonen, H., Walsh, P. M., Bucher, O., Engholm, G., Jackson, C., McClure, C., Woods, R. R., Saint-Jacques, N., Morgan, E., Ransom, D., Thursfield, V., Møller, B., … Arnold, M. (2019). Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *The Lancet. Gastroenterology & Hepatology*, *4*(7), 511–518. https://doi.org/10.1016/S2468-1253(19)30147-5

Asano, T. K., & McLeod, R. S. (2004). Nonsteroidal anti-inflammatory drugs and aspirin for the prevention of colorectal adenomas and cancer: a systematic review. *Diseases of the Colon and Rectum*, *47*(5), 665–673. https://doi.org/10.1007/s10350-003-0111-9

Azzouz, L. L., & Sharma, S. (2021). *Physiology, Large Intestine.*

Barceló-Coblijn, G., & Fernández, J. A. (2015). Mass spectrometry coupled to imaging techniques: The better the view the greater the challenge. *Frontiers in Physiology*, *6*(JAN), 1–7. https://doi.org/10.3389/fphys.2015.00003

Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P. J., & Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature*, *449*(7165), 1003–1007. https://doi.org/10.1038/nature06196

Bestard-Escalas, J., Maimó-Barceló, A., Pérez-Romero, K., Lopez, D. H., & Barceló-Coblijn, G. (2019). Ins and Outs of Interpreting Lipidomic Results. *Journal of Molecular Biology*, *431*(24), 5039–5062. https://doi.org/10.1016/j.jmb.2019.08.006

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Churko, J. M., Mantalas, G. L., Snyder, M. P., & Wu, J. C. (2013). Overview of high throughput sequencing technologies to elucidate molecular pathways in cardiovascular diseases. *Circulation Research*, *112*(12), 1613–1623. https://doi.org/10.1161/CIRCRESAHA.113.300939

Colucci, P. M., Yale, S. H., & Rall, C. J. (2003). Colorectal polyps. *Clinical Medicine & Research*, *1*(3), 261–262. https://doi.org/10.3121/cmr.1.3.261

Eaden, J. A., Abrams, K. R., & Mayberry, J. F. (2001). The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut*, *48*(4), 526–535. https://doi.org/10.1136/gut.48.4.526

Futerman, A. H., & Riezman, H. (2005). The ins and outs of sphingolipid synthesis. *Trends in Cell Biology*, *15*(6), 312–318. https://doi.org/10.1016/j.tcb.2005.04.006

Garate, J., Fernández, R., Lage, S., Bestard-Escalas, J., Lopez, D. H., Reigada, R., Khorrami, S., Ginard, D., Reyes, J., Amengual, I., Barceló-Coblijn, G., & Fernández, J. A. (2015). Imaging mass spectrometry increased resolution using 2-mercaptobenzothiazole and 2,5-diaminonaphtalene matrices: application to lipid distribution in human colon. *Analytical and Bioanalytical Chemistry*, *407*(16), 4697–4708. https://doi.org/10.1007/s00216-015-8673-7

Giovannucci, E. (2001). An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *10*(7), 725–731.

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., De Sousa E Melo, F., Missiaglia, E., Ramay, H., Barras, D., … Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, *21*(11), 1350–1356. https://doi.org/10.1038/nm.3967

Harada, S., & Morlote, D. (2020). Molecular Pathology of Colorectal Cancer. *Advances in Anatomic Pathology*, *27*(1), 20–26. https://doi.org/10.1097/PAP.0000000000000247

Jaladanki, R., & Wang, J.-Y. (2011). Regulation of Gastrointestinal Mucosal Growth. *Colloquium Series on Integrated Systems Physiology: From Molecule to Function*, *3*, 1–114. https://doi.org/10.4199/C00028ED1V01Y201103ISP015

Jung, P., Sato, T., Merlos-Suárez, A., Barriga, F. M., Iglesias, M., Rossell, D., Auer, H., Gallardo, M., Blasco, M. A., Sancho, E., Clevers, H., & Batlle, E. (2011). Isolation and in vitro expansion of human colonic stem cells. *Nature Medicine*, *17*(10), 1225–1227. https://doi.org/10.1038/nm.2470

Kopczynski, D., Coman, C., Zahedi, R. P., Lorenz, K., Sickmann, A., & Ahrends, R. (2017). Multi-OMICS: a critical technical perspective on integrative lipidomics approaches. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*. https://doi.org/10.1016/j.bbalip.2017.02.003

Kwak, E. L., & Chung, D. C. (2007). Hereditary colorectal cancer syndromes: an overview. *Clinical Colorectal Cancer*, *6*(5), 340–344. https://doi.org/10.3816/CCC.2007.n.002

Labianca, R., Beretta, G. D., Kildani, B., Milesi, L., Merlin, F., Mosconi, S., Pessi, M. A., Prochilo, T., Quadri, A., Gatta, G., de Braud, F., & Wils, J. (2010). Colon cancer. *Critical Reviews in Oncology/Hematology*, *74*(2), 106–133. https://doi.org/10.1016/j.critrevonc.2010.01.010

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted

correlation network analysis. *BMC Bioinformatics*, *9*, 559. https://doi.org/10.1186/1471-2105-9-559

Li, C., Jiang, W., & Xu, Y. (2017). Omics and Bioinformatics: Time for New Data Analysis Approaches? In *Omics : a journal of integrative biology* (Vol. 21, Issue 12, p. 749). https://doi.org/10.1089/omi.2017.0120

Loewe, R. P., & Nelson, P. J. (2011). Microarray bioinformatics. *Methods in Molecular Biology (Clifton, N.J.)*, *671*, 295–320. https://doi.org/10.1007/978-1-59745-551-0_18

Lynch, H. T., & Lynch, J. F. (2000). Hereditary nonpolyposis colorectal cancer. *Seminars in Surgical Oncology*, *18*(4), 305–313. https://doi.org/10.1002/(sici)1098-2388(200006)18:4<305::aid-ssu5>3.0.co;2-a

Mahdipour-Shirayeh, A., & Shahriyari, L. (2018). Modeling Cell Dynamics in Colon and Intestinal Crypts: The Significance of Central Stem Cells in Tumorigenesis. *Bulletin of Mathematical Biology*, *80*(9), 2273–2305. https://doi.org/10.1007/s11538-018-0457-8

Maida, M., Macaluso, F. S., Ianiro, G., Mangiola, F., Sinagra, E., Hold, G., Maida, C., Cammarota, G., Gasbarrini, A., & Scarpulla, G. (2017). Screening of colorectal cancer: present and future. *Expert Review of Anticancer Therapy*, *17*(12), 1131–1146. https://doi.org/10.1080/14737140.2017.1392243

Mansoor, S., Dolkar, T., & El-Fanek, H. (2013). Polyps and polypoid lesions of the colon. *International Journal of Surgical Pathology*, *21*(3), 215–223. https://doi.org/10.1177/1066896913481060

Merlos-Suárez, A., Barriga, F. M., Jung, P., Iglesias, M., Céspedes, M. V., Rossell, D., Sevillano, M., Hernando-Momblona, X., Da Silva-Diz, V., Muñoz, P., Clevers, H., Sancho, E., Mangues, R., & Batlle, E. (2011). The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell Stem Cell*, *8*(5), 511–524. https://doi.org/10.1016/j.stem.2011.02.020

Merlos-Suárez, A., & Batlle, E. (2008). Eph-ephrin signalling in adult tissues and cancer. *Current Opinion in Cell Biology*, *20*(2), 194–200. https://doi.org/10.1016/j.ceb.2008.01.011

Müller, M. F., Ibrahim, A. E. K., & Arends, M. J. (2016). Molecular pathological classification of colorectal cancer. *Virchows Archiv : An International Journal of Pathology*, *469*(2), 125–134. https://doi.org/10.1007/s00428-016-1956-3

O'Brien, M. J., Zhao, Q., & Yang, S. (2015). Colorectal serrated pathway cancers and precursors. *Histopathology*, *66*(1), 49–65. https://doi.org/10.1111/his.12564

Olivier, M., Asmis, R., Hawkins, G. A., Howard, T. D., & Cox, L. A. (2019). The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *International Journal of Molecular Sciences*, *20*(19). https://doi.org/10.3390/ijms20194781

Pasquale, E. B. (2008). Eph-Ephrin Bidirectional Signaling in Physiology and Disease. *Cell*, *133*(1), 38–52. https://doi.org/10.1016/j.cell.2008.03.011

Quinn, P. J. (2002). Plasma membrane phospholipid asymmetry. *Sub-Cellular Biochemistry*, *36*, 39–60. https://doi.org/10.1007/0-306-47931-1_3

Telloni, S. M. (2017). Tumor Staging and Grading: A Primer. *Methods in Molecular Biology (Clifton, N.J.)*, *1606*, 1–17. https://doi.org/10.1007/978-1-4939-6990-6_1

van Meer, G., Voelker, D. R., & Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nature Reviews. Molecular Cell Biology*, *9*(2), 112–124. https://doi.org/10.1038/nrm2330

von Roon, A. C., Reese, G., Teare, J., Constantinides, V., Darzi, A. W., & Tekkis, P. P. (2007). The risk of cancer in patients with Crohn's disease. *Diseases of the Colon and Rectum*, *50*(6), 839–855. https://doi.org/10.1007/s10350-006-0848-z

Waller, A., Findeis, S., & Lee, M. J. (2016). Familial Adenomatous Polyposis. *Journal of Pediatric Genetics*, *5*(2), 78–83. https://doi.org/10.1055/s-0036-1579760

Wiseman, M. (2008). The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. *The Proceedings of the Nutrition Society*, *67*(3), 253–256. https://doi.org/10.1017/S002966510800712X

Xiong, X., Xu, W., Eberlin, L. S., Wiseman, J. M., Fang, X., Jiang, Y., Huang, Z., Zhang, Y., Cooks, R. G., & Ouyang, Z. (2012). Data processing for 3D mass spectrometry imaging. *Journal of the American Society for Mass Spectrometry*, *23*(6), 1147–1156. https://doi.org/10.1007/s13361-012-0361-7

Yang, K., & Han, X. (2016). Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends in Biochemical Sciences*, *41*(11), 954–969. https://doi.org/10.1016/j.tibs.2016.08.010

7. Annex

Table I. CRC patient's clinical information.

Patient	Age	Sex	Localization	Histological type	TNM	ANALYSIS
305	65	M	Sigmoid colon	ADC	T4b, N0	Healthy FACS samples MALDI-IMS
309	83	F	Sigmoid colon	ADC	T3, N0	Healthy FACS samples MALDI-IMS
311	75	F	Sigmoid colon	ADC	T3, N1c	Healthy FACS samples MALDI-IMS
312	72	F	Sigmoid colon	Residual Neoplasia	X, N0	Healthy FACS samples MALDI-IMS
313	71	M	Cecum	ADC	T3, N0	Tumor FACS samples MALDI-IMS
314	86	M	Cecum	ADC	T3, N0	Tumor FACS samples MALDI-IMS
315	71	F	Ascending colon	ADC	T3, N1c	Tumor FACS samples MALDI-IMS
316	82	M	Sigmoid colon	ADC	T3, N2a	Tumor FACS samples MALDI-IMS
317	48	M	Sigmoid colon	Mucinous ADC	T4b, N0	Tumor FACS samples MALDI-IMS
320	65	M	Sigmoid colon	ADC	T3, N0	Tumor FACS samples MALDI-IMS
321	85	F	Sigmoid colon	ADC	T3, N0	Healthy and Tumor FACS Human clariom S pico Affymetrix
322	72	F	Ascending colon	ADC	T4a,N0	Healthy FACS samples MALDI-IMS + Healthy and Tumor FACS Human clariom S pico Affymetrix
330	71	M	Sigmoid colon	ADC	T3, N2a	Healthy FACS samples MALDI-IMS
331	68	F	Ascending colon	Mucinous ADC	T3,N0	Healthy and Tumor FACS Human clariom S pico Affymetrix
332	78	M	Ascending colon	ADC	T3,Nb	Healthy and Tumor FACS Human clariom S pico Affymetrix
335	83	M	Ascending colon	ADC	T3, N0	Tumor FACS samples MALDI-IMS + Healthy and Tumor FACS Human clariom S pico Affymetrix
373	73	F	Sigmoid colon	ADC	T4a, N1c	Healthy & Tumor FACS samples MALDI-IMS

Abbreviations: Adenocarcinoma (ADC), not described (ND).