Inês Rodrigues Borges Ferreira
Degree in Biology

Identification of biomarkers of colorectal cancer risk and metastasis: Targeting early onset cancer

Dissertation to obtain the Master's Degree in Molecular Genetics and Biomedicine

Supervisor: Doutora Maria Cristina Mantas Albuquerque Valeroso, Senior Investigator, UIPM, IPOLFG

Jury

President: Doutora Maria Alexandra Núncio de Carvalho Ramos Fernandes

Arguer: Doutora Maria Luísa Santos de Sousa Cyrne

Supervisor: Doutora Maria Cristina Mantas Albuquerque Valeroso

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September 2019

Inês Rodrigues Borges Ferreira
Degree in Biology

Identification of biomarkers of colorectal cancer risk and metastasis: Targeting early onset cancer

Dissertation to obtain the Master's Degree in Molecular Genetics and Biomedicine

Supervisor: Doutora Maria Cristina Mantas Albuquerque Valeroso, Senior Investigator, UIPM, IPOLFG

Jury

President: Doutora Maria Alexandra Núncio de Carvalho Ramos Fernandes

Arguer: Doutora Maria Luísa Santos de Sousa Cyrne

Supervisor: Doutora Maria Cristina Mantas Albuquerque Valeroso

September 2019

Acknowledgments

Gostaria de expressar o maior agradecimento a todas as pessoas que contribuíram para o desenvolvimento deste projecto e que me apoiaram durante o decorrer deste ano.

Primeiramente, gostaria de agradecer à orientadora desta dissertação e projecto, Dra. Cristina Albuquerque, por ter confiado e acreditado em mim para desenvolver parte deste projecto que tem uma história imensa. Agradecer por toda a orientação e apoio prestado ao meu trabalho, por todos os conhecimentos transmitidos e a paciência para esclarecer todas as minhas dúvidas ou para me ajudar e compreender nos momentos mais emotivos.

De seguida gostava de fazer um agradecimento especial à Inês Francisco e à Patrícia Silva por todos os conhecimentos transmitidos sobre as técnicas, trabalho de bancada, conhecimento teórico e alguns "fun facts". Agradeço pelas respostas às minhas infindáveis perguntas e dúvidas existenciais, pelas vezes todas que me iniciaram a conta no computador e me abriram as pastas que eu nunca sabia onde estavam. Principalmente mostraram-me como ter ritmo de trabalho, a ser organizada e deram-me todas as ferramentas para trabalhar de forma autónoma e ter sentido crítico. Tudo ensinamentos essenciais que levo comigo para os meus futuros projectos.

Um enorme obrigada à Lucília Pereira, inicialmente uma colega que agora considero amiga, por toda a ajuda, apoio, paciência e ensinamentos que me passou ao longo deste ano. Obrigada por toda a simpatia, conversas na sala da cultura e por ouvires os meus desabafos sobre tudo e mais alguma coisa. Um obrigada ao Bruno Felipe, por ajudar sempre que fosse preciso de forma muito prestável, apesar de eu ocupar muitas vezes o seu computador no gabinete. Um agradecimento especial à Carina que nos primeiros meses me passou imensos conhecimentos, e tudo o que era preciso fazer no laboratório. Um obrigada à Carolina Pires, colega de mestrado e amiga de IPO por todas as conversas, desabafos e companhia.

Mãe, pai e mano, obrigada por todo o apoio incondicional, por se interessarem pelas minhas investigações e por me ajudarem a construir um futuro promissor. Obrigada por acreditarem em mim e me ajudarem a ser positiva.

Rodrigo um obrigado gigante, sem o teu apoio este ano teria sido diferente, todas as horas passadas no metro e comboio teriam sido muito mais aborrecidas. Obrigada por me ouvires, me apoiares e me dares motivação.

Abstract

Colorectal cancer (CRC) is one of the leading cancer related causes of death. Over the years, the number of individuals diagnosed with early onset CRC has been increasing, characterized by poor prognosis and often presenting distant metastasis. This increases the need of finding biomarkers for early diagnosis and the development of new precision therapies.

The transcription factor encoded by the *TCF7L2* gene is a key component in the Wnt signaling pathway and alterations in this gene have been identified in previous studies as probably related with CRC and malignant transformation. The first aim of this study was to investigate if alterations in the *TCF7L2* gene could contribute for increased CRC risk, especially of early onset CRC and to early malignant transformation. To address this the analysis of specific *TCF7L2* polymorphisms, copy number variations and gene expression was performed in DNA from tumor and normal mucosa samples from a group of individuals enriched for features associated to increased CRC risk, namely for early onset CRC. These results were related with personal and familial history of CRC and metastatic disease. The results showed an association of specific genotypes for the polymorphisms and also gene expression with increased CRC risk (namely to early onset). Specific genotypes also associated to metastases development

The second aim was to work towards a precision therapy to metastatic disease using the synergic effect of targeted therapies and nutraceutics. Therefore, expression of gene markers of EMT, stemness and cell cycle markers were analyzed in the metastatic LoVo cell line treated with GANT61, SFN and their combination, based on previous studies. It was observed that the combination of these treatments may attenuate mesenchymal phenotype by downregulating EMT markers and induce cell cycle arrest of metastatic cells.

Keywords: Colorectal cancer; Bethesda Criteria; *TCF7L2* gene; early onset; metastasis

Resumo

Cancro colorectal (CCR) é uma das maiores causas de morte, relacionadas com o cancro. Ao longo dos anos o número de indivíduos diagnosticados com CCR em idade jovem tem aumentado, é caracterizado por ser um tipo de cancro avançado com mau prognóstico e com presença de metástases distantes. Esta é uma chamada de atenção para a necessidade de encontrar novos biomarcadores que facilitem o diagnóstico precoce e de novas terapias de precisão.

O factor de transcrição gene *TCF7L2* é um componente vital na via de sinalização Wnt e alterações neste foram previamente identificadas como provavelmente relacionadas com o CCR. O primeiro objectivo deste estudo foi investigar se alterações no gene *TCF7L2* podem contribuir para o risco aumentado de desenvolvimento de CCR, especialmente em idade jovem. Foi realizada uma análise específica aos polimorfismos do gene *TCF7L2*, foram avaliadas as alterações de *copy-number*, a expressão do *TCF7L2* em amostras de tumor e mucosa normal de indivíduos com características associadas ao risco aumentado para CCR, nomeadamente CCR em idade precoce. Estes resultados foram relacionados com a história pessoal e familiar de CCR e doença metastática. Os resultados mostraram uma associação de genótipos específica para o polimorfismo e uma expressão com elevado risco para o desenvolvimento de CCR (principalmente precoce). Genótipos específicos estão também associados ao desenvolvimento de metástases.

O segundo objectivo foi obter uma nova terapia de precisão, usando o efeito sinérgico de fármacos e nutraceuticos. Foi analisada a expressão de vários marcadores do ciclo celular, *stemeness,* EMT, na linha celular LoVo tratada com GANT61, sulfrafano e a sua combinação, baseado em trabalho anterior do grupo. A combinação dos tratamentos previne o fenótipo mesenquimal por baixar a expressão dos marcadores de EMT e induzir a paragem do ciclo celular das células metastáticas.

Palavras-chave: Cancro colorectal; Critérios de Bethesda; TCF7L2; idade jovem; metástases

List of Contents

List of Figures

List of Tables

List of Abbreviations

5-FU	5-fluorouracil
A	Adenine
AC	Amsterdam criteria
AFAP	Attenuated FAP
APC	Adenomatous polyposis coli
aUPD	Acquired uniparental disomy
BC	Bethesda criteria
BER	DNA base excision repair gene
bp	Base pair
BR	Broad range
BRAF	Murine sarcoma viral oncogene homolog B
C	Cytosine
cDNA	Complementary DNA
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CMS	Consensus molecular subtypes
CRC	Colorectal cancer
CSC	Cancer stem cell
C_t	Threshold cycle
ddH_2O	Double-distilled water
ddNTPs	Dideoxynucleotides phosphate
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylenediamine tetra acetic acid
EGFR	Epidermal growth factor receptor
EMT	Epithelial to mesenchymal transition
ERK	Extracellular signal-regulated kinases

Exo I	*Exonuclease I*
F	Forward
FAP	Familial adenomatous polyposis
FastAP	*Thermosensitive Alkaline Phosphatase*
FCCTX	Familial colorectal cancer type X
FFPE	Formalin-fixed paraffin-embedded
G	Guanine
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GLI	Human glioma-associated oncogene homolog
GTP	Guanosine triphosphate
Hh	Hedgehog
HMG	High mobility group
HNPCC	Hereditary nonpolyposis colorectal cancer
HS	High-sensitivity
ICG	International Collaborative Group
IPOLFG	Instituto Português de Oncologia de Lisboa, Francisco Gentil
ITC	Isothiocyanate
KRAS	Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LEF	Lymphoid enhancer factor
LGR5	Leucine-rich repeat-containing G-protein coupled receptor 5
LOH	Loss of heterozygosity
LS	Lynch syndrome
MAP	MUTYH associated polyposis
MAPK	Mitogen Activated Protein Kinase
mCRC	CRC with distant metastases
MET	Mesenchymal to epithelial transition
MGB	Minor groove binder
MGMT	O6-alkylguanine DNA alkyltransferase
MLH1	*mutL* homologue 1
MMR	Mismatch repair
mRNA	Messenger ribonucleic acid
MSH2	*mutS* homologue 2
MSH6	mutS homologue 6

MSI	Microsatellite instability
MSI-H	High frequency MSI
MSI-L	Low frequency MSI
MUTYH	mutY DNA glycosylase
NFQ	Nonfluorescent quencher
OS	Overall survival
PCR	Polymerase chain reaction
PMS2	Postmeiotic segregation 2
PSF	Progression-free survival
qPCR	Quantitative PCR
R	Reverse
RNA	Ribonucleic acid
RT-PCR	Reverse PCR
SCNA	Somatic copy number alterations
SFN	Sulforaphane
SNP	Single nucleotide polymorphism
T	Timine
TBE	Tris-burato-EDTA buffer
TCF	T-cell factor
TCF7L2	Transcription Factor 7 like 2
TF	Transcription factor
TGF-β	Growth factor-beta
TME	Tumor microenvironment
TNM	Tumor-nodes-metastasis
TP53	Tumor protein 53
TSG	Tumor suppressor gene
UIPM	Unidade de Investigação em Patologia Molecular
UPD	Uniparental disomy
UTR	Untranslated region
UV	Ultraviolet radiation
WNT	Wingless-related integration site
ZEB 1	Zinc finger E-box binding homeobox 1

1.1 Colorectal cancer

1.1.1 Epidemiology and etiology

Colorectal cancer (CRC) is one of the leading causes of, worldwide cancer related, death thus representing the third most commonly diagnosed cancer in males and second in females, with an estimated 1.4 million cases occurring in 2012 (Favoriti et al., 2016). The majority of the cases (55%) occur in developed countries, in particular Australia, New Zealand, Europe and Northern America in comparison to Africa, South-Central Asia and Central America that represent the countries with low CRC incidence (Favoriti et al., 2016). CRC incidence rates have been rinsing in Eastern Europe and this fact could be associated to western lifestyle habits such as sedentarism, smoking prevalence, alcohol intake and consumption of processed food and red meat. Other risk factors that have a positive association with CRC are inflammatory bowel disease, either Crohn's or ulcerative colitis, family history of colorectal cancer and age. On the other hand, there are some factors associated with a lower risk of developing CRC such as physical activity, vegetable and fruit intake (Johnson et al., 2013). Colon cancer is a result of genetic predisposition and environmental factors, and the risk of incidence begins to rise between the ages of 40 and 50 years, increasing as time passes by (Favoriti et al., 2016). Overall, the 5 year relative survival rate for patients with CRC corresponds to 64.9%, however when tumors are detected at an early stage this percentage increases to 90% (Favoriti et al., 2016). The downside is that only 39% of the tumors are diagnosed at this stage (Favoriti et al., 2016), underlining the importance of an early diagnosis when tumors are not yet metastasized. Once the cancer has metastasized, the 5 year survival rate decreases for 12% (Favoriti et al., 2016).

1.1.2 Sporadic and familial colorectal cancer

The majority of CRC, about 70% develops sporadically due to somatic alterations in colon epithelial cells (Sehgal et al., 2014). Age is the most important risk factor for sporadic CRC, once over 90% of these cases occur in individuals over 50.5 years old (Al-Sohaily et al., 2012). The remaining 30% occur in a familial history predisposition context, however, only about 5% is well characterized at both clinical and molecular levels occurring due to germline alterations in specific genes.

1.1.3 Colon localization and characterization

Colon is a tubular organ with 1.5 to 1.8m long, which is divided in two parts, the proximal and distal colon. The proximal colon is composed by the ascending and transverse colon while the distal colon is composed by the descending colon and sigmoid. The functions of the colon are mainly absorption of water, nutrients and vitamins; compaction and moving the feces towards the rectum; and potassium and chloride secretion (Kahai, Mandiga and Lobo, 2019). Histologically, colon is composed by: mucosa (close to the lumen, mainly composed by the epithelium); submucosa (with blood vessels, nerves and connective tissue); *muscularis propria* (inner circular layer, responsible for the intestinal movements); and serosa (perimuscular tissue, is a thin outer layer) (Gulwani, 2013). The epithelium has crypts, small invaginations composed by differentiated cells and surrounded by pericryptal fibroblasts. At the bottom of the crypt there is the tissue stem cell niche, filled with undifferentiated

cells with self-renewal capacity and any mutation in this region can progress to tumor development, turning into cancer stem cells (CSCs).

1.1.3.1 CRC staging

At the time of diagnosis, the extent of the cancer is described by it stage. Nowadays, CRC staging is based on the tumor-nodes-metastasis (TNM) classification system of malignant tumors proposed by American Joint Committee on Cancer (Nguyen and Duong, 2018) (Table 1.1). TNM describes cancers based on their anatomy and categorizes each tumor by the status of the primary tumor (T), nodal metastasis (N) and metastatic disease (M) (Gaillard and J Bell, n.d.).

Table 1.1- CRC staging based on TNM classification system of malignant tumors and respective legend. (Table adapted from: Colon and Rectum Cancer Staging, 2009)

ANATOMIC STAGE/PROGNOSTIC GROUPS

Stage	T	N	M
0	Tis	N0	M0
I	T1	N0	M0
	T2	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
IIIA	T1–T2	N1/N1c	M0
	T1	N2a	M0
IIIB	T3–T4a	N1/N1c	M0
	T2–T3	N2a	M0
	T1–T2	N2b	M0
IIIC	T4a	N2a	M0
	T3–T4a	N2b	M0
	T4b	N1–N2	M0
IVA	Any T	Any N	M1a
IVB	Any T	Any N	M1b

Primary Tumor (T)

- T0 No evidence of primary tumor
- Tis Carcinoma in situ: intraepithelial or invasion of lamina propria[1]
- T1 Tumor invades submucosa
- T2 Tumor invades muscularis propria
- T3 Tumor invades through the muscularis propria into pericolorectal tissues
- T4a Tumor penetrates to the surface of the visceral peritoneum
- T4b Tumor directly invades or is adherent to other organs or structures[2]

Regional Lymph Nodes (N)

- N0 No regional lymph node metastasis
- N1 Metastasis in 1–3 regional lymph nodes
- N2 Metastasis in 4 or more regional lymph nodes

Distant Metastasis (M)

- M0 No distant metastasis
- M1 Distant metastasis

1.1.4 Molecular pathways behind colorectal cancer

There are three molecular pathways characterized by genetic instability, specific pathological features and mechanisms of carcinogenesis. The molecular pathways behind CRC are: chromosomal instability (CIN) pathway; microsatellite instability (MSI) pathway; and CpG island methylator phenotype (CIMP) pathway.

1.1.4.1 Chromosomal instability (CIN) pathway

One of the most used concepts to describe the transition of normal colon epithelial, to uncontrolled proliferation and then to carcinoma is the adenoma-carcinoma sequence model (Figure

1), also known as the CIN pathway. This model for colorectal tumorigenesis was first proposed by Fearon and Vogelstein where they postulated that mutational activation of oncogenes, coupled with inactivation of tumor suppressor genes and mutations in the DNA repair genes, leads to the development of colorectal neoplasia (Leslie et al., 2002).

Firstly, CIN pathway is the most prevalent molecular cause of genomic instability, counting for 65 to 70% of all sporadic CRC tumors (Zeinalian et al., 2018). CIN is characterized by the loss of heterozygosity (LOH) and by chromosomal abnormalities, corresponding to gains or losses of whole chromosomes or regions that harbor important genes in the process of colorectal carcinogenesis, such as mutations in *APC*, *KRAS*,*SMAD4* and *TP53* genes (Figure 1.1).

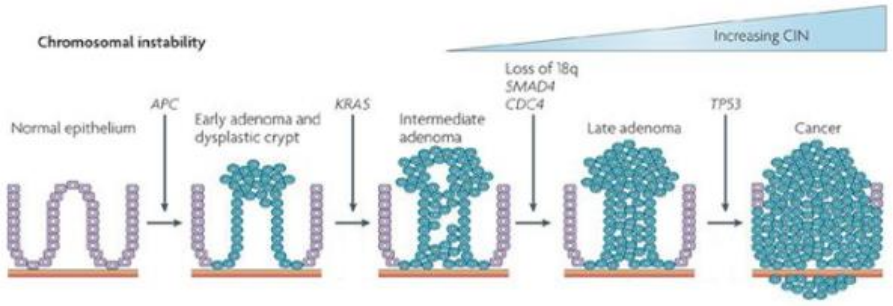


Figure 1.1- Adenoma-carcinoma sequence model in colorectal cancer. (Figure adapted from Walther et al., 2009)

The *APC* (Adenomatous polyposis coli) gene is a tumor suppressor gene, located on chromosome 5q21, that functions as a WNT signaling pathway regulator due to its capacity to down-regulate intracellular β-catenin levels. *APC* gene also plays a role in controlling cell cycle progression and stabilizing microtubules, thus promoting chromosomal stability (Al-Sohaily et al., 2012). The loss of function of this gene results in the transition of normal polarized intestinal epithelium to dysplastic adenomatous lesion, being the adenoma the precursor lesion of this pathway (Albuquerque et al., 2010). Wnt/β-catenin signaling pathway normally is one of the principal regulatory mechanisms to preserve tissue homeostasis in the organism by adjusting the regulation of self-renewal, differentiation and apoptosis in stem cell niches (Albuquerque et al., 2011). Wnt/β-catenin pathway is a conserved signal transduction cascade and in the absence of WNT signal the cytoplasmic β-catenin is phosphorylated, by the destruction complex, with subsequent ubiquitination (Figure 2). On the other hand, when this signal is present, received through the Frizzled receptors, β-catenin accumulates in the cytoplasm before translocation into the nucleus where is formed a transcriptional complex with T-cell factor (TCF)/ lymphoid enhancer factor (LEF) proteins, activating Wnt target genes (Figure 1.2) (Tomimaru et al., 2013). *APC* gene mutation affects the function of the destruction complex, composed by two kinases (GSK3 and CKI-a), APC, β-catenin and two scaffold proteins (AXIN1 and AXIN2), leading to an accumulation of β-catenin in the cytoplasm and enhanced nuclear signaling, triggering a genetic programme that initiates tumor formation (Albuquerque et al., 2011). Tissues with high levels of nuclear β-catenin are usually associated with tubular branching, invasion and EMT (Bush et al., 2013). CRC cells upregulate the expression of Wnt signaling pathway feedback inhibitors

like AXIN2, a promoter of tumor progression by upregulating the expression of Snail (a transcriptional repressor that induces EMT, which will be described later) inducing the mesenchymal phenotype and driving to metastatic activity (Wu et al., 2012).

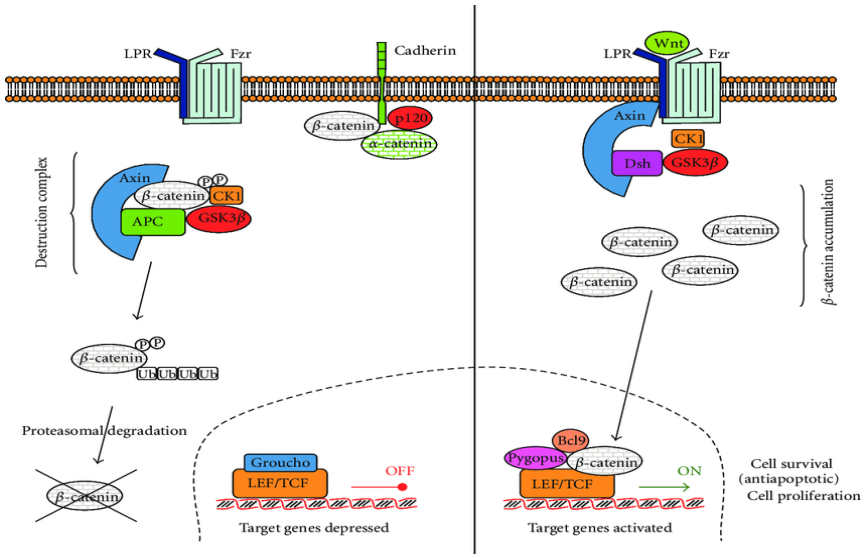


Figure 1.2- Canonical Wnt/B-catenin pathway in CRC (Figure adapted from Centelles, 2012)

KRAS (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog) gene is an oncogene that belongs to the MAPK pathway, this gene codifies to the KRAS protein, which is a GTPase that functions as an intermediate between the cell's nucleus and the MAPK pathway. When *KRAS* is mutated the cell becomes an intermediate adenoma. LOH in the long arm of chromosome 18 (18q) contributes to the evolution of the late adenoma. Also *SMAD4* gene is a tumor suppressor which mediates the transformation of growth factor-β (TGF-β) superfamily signaling, located in chromosome 18q21, usually a LOH region, that has the ability to regulate cell proliferation, differentiation and apoptosis (Alazzouzi et al., 2005). Finally, *TP53* (Tumor protein 53) gene is a tumor suppressor gene, localized in the 17p chromosome that encodes tumor protein p53 responsible for the regulation of cell growth and proliferation, inducing cell apoptosis. These mutations are one of the key steps in CRC carcinogenesis due to the loss of control of the cell cycle (Colussi et al., 2013).

1.1.4.2 Microsatellite instability (MSI) pathway

MSI pathway occurs due to mutations in the mismatch repair (MMR) genes and counts for 8% to 20% of all CRC tumors (Zeinalian et al., 2018). The MMR system is one of the essential keys in maintaining genome integrity due to its function of removing biosynthetic errors from synthesized DNA, improving the fidelity of DNA replication (Jiricny, 2006). In the genome there are short sequences of repeated DNA called microsatellites, which vary in number between individuals, being regions with high mutation rate and high contribution for the genetic diversity. The accumulation of genetic mutations in these MMR genes leads to the presence of abnormal sizes of microsatellites, inducing microsatellite instability (MSI). MSI leads to tumor cell proliferation, invasion and metastasis through activation of oncogenes or inactivation of tumor suppressors (Chen, Xu and Liu, 2018). CRC

that develop from the MSI pathway will progress from a conventional adenoma similar to the CIN pathway.

1.1.4.3 CpG island methylator phenotype (CIMP) pathway

CIMP pathway is related to epigenetic molecular changes causing alterations in gene expression or gene function without any change in the DNA sequence. This pathway is characterized by hypermethylation of promoter CpG island sites, regions rich in cytosines and guanines, leading to the silencing of some essential tumor suppressor genes or other tumor-related genes, and consequently the development of CRC tumors. Epigenetic regulations of gene expression, in normal tissue, have an important role in embryogenesis, tissue differentiation and genomic stability. This pathway represents about 35% of all CRC tumors (Zeinalian et al., 2018), which develops from a sessile or traditional serrated adenoma. CIMP tumors have a high rate of mutations in *BRAF* (Murine sarcoma viral oncogene homolog B) or *KRAS* genes and are pathologically characterized by a proximal colon location, mucinous histological type, poor differentiation and higher age of diagnosis (Nazemalhosseini Mojarad et al., 2013). *BRAF* is a proto-oncogene that encodes a protein which belongs to the RAF family of serine/threonine protein kinases, playing an important role in the regulation of MAP kinase/ERK signaling pathway, affecting cell division, secretion and differentiation. In this gene, the V600E mutation is the most common cancer-causing mutation in CRC, and other cancers.

1.1.5 Hereditary CRC syndromes

The hereditary CRC syndromes are associated with a higher risk of developing CRC. These syndromes are divided into: nonpolyposis and polyposis syndromes. The most common ones are the hereditary nonpolyposis colorectal cancer (HPNCC) and familial adenomatous polyposis (FAP), which represents 2%-5% and 1% of all CRC cases, respectively (Lynch, Lynch and Attard, 2009).

1.1.5.1 Hereditary nonpolyposis colorectal cancer (HPNCC)

HNPCC is an inherited autosomal dominant disease that can be divided into two groups: Lynch syndrome (LS) and familial colorectal cancer type X (FCCTX). In 1989, the International Collaborative Group on HNPCC (ICG-HNPCC) developed a set of criteria called “Amsterdam Criteria I” for a better diagnosis of HNPCC patients (Table 1.2). Later, in 1999, the criteria were revised to incorporate the extracolonic tumors (Sehgal et al., 2014).

Table 1.2- Amsterdam Criteria I and II. (Table taken from the article Sehgal et al., 2014)

Amsterdam Criteria I	**Amsterdam Criteria II**
At least three relatives with histologically verified colorectal cancer:	At least three relatives with HNPCC-associated cancer [CRC, endometrial, stomach, ovary, ureter/renal pelvis, brain, small bowel, hepatobiliary tract, and skin (sebaceous tumors)]:
1. One is a first-degree relative of the other two	1. One is a first-degree relative of the other two
2. At least two successive generations affected	2. At least two successive generations affected;
3. At least one of the relatives with colorectal cancer diagnosed at <50 years of age	3. At least one of the syndrome-associated cancers should be diagnosed at <50 years of age;
4. FAP has been excluded	4. FAP should be excluded in any colorectal cancer cases;

1.1.5.1.1 Lynch syndrome

Lynch syndrome is the most relevant syndrome considering that counts for approximately 3% of all CRC cases (Chen, Xu and Liu, 2018). LS is a well characterized autosomal dominant syndrome, with an increased risk of developing extracolonic cancers such as in the endometrium, ovaries, stomach, small intestine, among others. This syndrome is characterized by presenting an accelerated carcinogenesis in which small adenomas can turn into carcinomas in 2 to 3 years, nevertheless these cancers have an increased survival rate (Lynch, Lynch and Attard, 2009). Pathologically, the CRC associated with LS is usually poorly differentiated, with an excess of mucoid and signet-cell features, as well as Crohn-like reaction and a high number of infiltrating lymphocytes within the tumor and follow the MSI cancer carcinogenesis pathway (Lynch, Lynch and Attard, 2009). The onset of CRC in Lynch syndrome patients is around 45 years old, develop earlier than the general population with CRC, around 63 years old. Molecularly, the development of extracolonic cancers can derive from a mutation in the MMR genes, in particular germline mutations in the *mutS* homologue 2 (*MSH2*), *mutL* homologue 1 (*MLH1*) genes, which are responsible for 70-90% of the LS tumors, and rarely in the *mutS* homologue 6 (MSH6) and postmeiotic segregation 2 (*PMS2*) genes (Chen, Xu and Liu, 2018). These tumors, because of their MMR deficiency, are localized predominantly on the right side of the intestine and around 70% of the tumors are proximal to the splenic flexure (Lynch et al., 2003).

Due to the identification of several MMR mutations the National Cancer Institute, in 1997, organized an International Workshop on Lynch syndrome in Bethesda where they standardized a diagnostic panel of microsatellite markers (BAT25, BAT26, D2S123, D5S346, and D17S250) (Lynch et al., 2003) and created the Bethesda Criteria (BC), exposed in Table 1.3, in order to identify new

cases of LS (Sehgal et al., 2014). It can also be used immunohistochemistry to identify germline mutations in the MMR genes.

Table 1.3- Revised Bethesda criteria (BC). (Table taken from the article Sehgal et al., 2014)

Revised Bethesda Criteria
Colorectal tumors from individuals should be tested for MSI in the following situations:
1. Colorectal cancer diagnosed in a patient who is <50 years of age.
2. Presence of synchronous or metachronous colorectal, or other HNPCC-associated tumors regardless of age.
3. Colorectal cancer with microsatellite instability-high (MSI-H) histology diagnosed in a patient who is <60 years of age.
4. Colorectal cancer diagnosed in one or more first-degree relatives with an HNPCC-related tumor, with one of the cancers being diagnosed under age 50 years.
5. Colorectal cancer diagnosed in two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age.

Colorectal tumors that have more than 40% microsatellite variations are described as high frequency MSI (MSI-H), and are also referred to as MMR-deficient. This phenotype is also described in 15% of sporadic CRCs due to somatic methylation of the *MLH1* promoter region. On the other hand, CRCs with low frequency MSI (MSI-L) are the ones that have less than 40% microsatellite variations (Sehgal et al., 2014). MSS, MSI-L and tumors with intact MMR proteins are referred as MMR-proficient (Sinicrope, 2010).

1.1.5.1.2 Familial colorectal cancer type X (FCCTX)

FCCTX represents 2 a 4% of the cases of HNPCC, and refers to patients who fulfill AC I, exhibiting no evidence of a germline mutation in MMR genes and are MSS. As a result the “X” is used to describe the unknown nature of this disease (Chen, Xu and Liu, 2018). Recent studies showed that 65% of colorectal cancers associated with FCCTX follows CIN pathway by gaining the chromosomal region 20q and losing the 18 region. FCCTX families, when compared with LS families, present lower incidence of CRC, at a later onset and with greater frequency in the distal colon. Normally characterized by a poor differentiation, a more mucinous characteristic and fewer multiple tumors, but in the other hand, presents distinctive morphological features, including tumor-infiltrating lymphocytes (Chen, Xu and Liu, 2018). Studies developed by UIPM gastroenterology group show two different molecular entities for the FCCTX families, one with frequent loss of tumor suppressor gene (TSG) *loci*, that follows CIN pathway with frequent *APC* and *KRAS* somatic mutations, as well as, frequent MMR

and *MGMT* gene promoter methylation; and other with no evidence of loss of TSG *loci* and less frequent *APC* and *KRAS* somatic mutations and MMR and MGMT gene promoter methylation (Francisco et al., 2011).

1.1.5.2 Familial adenomatous polyposis (FAP)

FAP is an autosomal dominant disease characterized by the development of hundreds to thousands of adenomatous polyps in the colon and rectum, generally in early teens. FAP counts for approximately 1% of all CRC with equal distribution in men and women (Kanth et al., 2017). FAP is a genetic disorder with a germline mutation in the *APC* gene (Half, Bercovich and Rozen, 2009). Almost every case will have CRC if, at an early stage, those patients were not identified or receive adequate treatment with a prophylactic colectomy. There is also the attenuated version of FAP (AFAP) that has fewer adenomas with later development. Patients with AFAP usually have a mutation in the 5' or 3' region of the *APC* gene and patients with FAP carry mutations elsewhere in the gene. The location of the mutation seems to correlate with disease severity and the presence of extracolonic manifestations in FAP patients, with the majority of the mutations leading to premature truncation of protein synthesis (Al-Sohaily et al., 2012). Typically FAP individuals present family history of the disease, however, approximately 25% to 30% of FAP and AFAP cases are due to a *de novo* APC mutation (Kanth et al., 2017).

1.1.5.2.1 *MUTYH* associated polyposis (MAP)

MAP is an autosomal recessive disease, associated with an increased risk (80%) of developing CRC. *MUTYH* gene, is located on chromosome 1p35, is a DNA base excision repair gene (BER) that repairs DNA injuries from oxidative stress, so has an important role in protecting genomic integrity during transcription. This gene identifies and changes the sequences where adenosine wrongly binds to the residue from oxidative damage. Biallelic *MUTYH* mutations induce the development of multiple colorectal adenomas, ranging from 10 to more than 100, with a higher prevalence of serrated polyps than FAP and AFAP (Kanth et al., 2017). This condition involves a high frequency of somatic *APC* mutations, a low frequency of LOH and the tumors are usually microsatellite stable (Al-Sohaily et al., 2012).

1.1.6 Invasion and metastasis

Tumors are characterized by their ability to invade and metastasize. This process involves detachment of tumor cells from its primary site, migration, invasion of blood or lymphatic vessels, dissemination, and finally settlement in the distant site. In order to cells initiate these processes, first they need to attain a mesenchymal like phenotype, by a process called epithelial to mesenchymal transition (EMT) (Al-Sohaily et al., 2012) (Figure 1.3). EMT is a biological process that allows a polarized epithelial cell, which commonly interacts with basement membrane via its basal surface, to undergo multiple changes by a set of transcription factors (TFs) that enable it to assume a mesenchymal cell phenotype. Epithelium in transition lose polarity, adherens junctions, tight junctions, desmosomes and cytokeratin intermediate filaments in order to enhance migratory capacity,

invasiveness, resistance to apoptosis and increased production of extracellular matrix (ECM) components (Kalluri and Weinberg, 2009). These are all hallmarks of increased malignancy. The action of these EMT-TFs enables the early steps of metastasis, such as local invasion and dissemination of carcinoma cells to distant sites (Scheel and Weinberg, 2012). Therefore, EMT provides a mechanism for carcinoma cells to acquire this more aggressive phenotype. Metastatic cells usually have morphologic features of the primary tumor, and not the invasive mesenchymal phenotype, which indicates the re-differentiation of migrating cells after settling in a distant site (Al-Sohaily et al., 2012). Based on this fact, EMT is described as a transition phenomenon due to its reversibility, the capacity to occur the reverse process, mesenchymal to epithelial transition (MET). During EMT, the epithelial markers like E-cadherin, encoded by *CDH1* gene, and cytokeratins are down-regulated while the mesenchymal markers, such as N-cadherin, *SNAIL* and *Vimentin*, are up-regulated. E-cadherin is essential for the maintenance of adherens junctions that confer physical integrity and polarization to epithelial cells (Bates and Mercurio, 2005). The loss of E-cadherin promotes the activation of Wnt signaling pathway and is associated with high levels of *SNAIL* in the nucleus. On the other hand, growth factor-beta (TGF-β), in colon carcinoma, acts differently depending on the differentiation stage of the tumor, in general by switching from an early inhibitor of proliferation to a stimulator of growth and invasion during tumor progression (Bates and Mercurio, 2005). An important inducer of EMT is the ZEB1 (Zinc finger E-box binding homeobox 1) transcription factor, it is an inhibitor of the epithelial phenotype, not only represses epithelial markers like E-cadherin but also induces the expression of mesenchymal genes. Its expression promotes tumorigenesis, metastasis development and determines chemotherapy resistance. ZEB1 is expressed in epithelial colon tumor cells with germline mutations in the APC gene that result in accumulation of β-catenin in the nucleus. (Sanchez-Tillo et al., 2011).

There are three types of EMT: Type 1 EMT are all the EMTs that are associated with implantation, embryo formation and organ development; Type 2 EMT associated with wound healing, tissue regeneration and organ fibrosis; Type 3 EMT occur in neoplastic cells that have undergone genetic and epigenetic changes in genes that lead to a development of localized tumors (Kalluri and Weinberg, 2009).

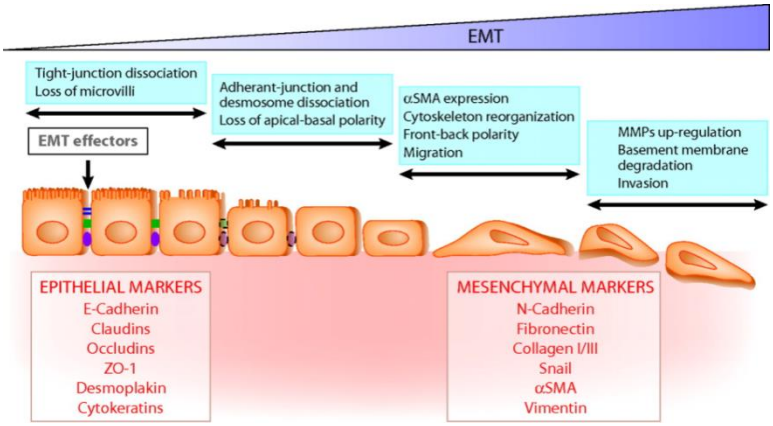


Figure 1.3- Representation of the epithelial to mesenchymal transition (EMT) and the respective EMT markers. (Figure taken from the article Aroeira et al., 2007)

EMT-TFs have been proved to maintain stemness properties and to increase tumorigenesis, suggesting that the induction of tumor initiation and formation of metastasis by the TFs is linked to the cancer stem cells (CSCs) (Brabletz et al., 2018). CSCs are a small subpopulation of tumor cells with self-renewal, differentiation and tumorigenesis abilities and are characterized by presenting surface antigens called stemness markers, such as CD133 (coded by *PROM1* gene) and CD44. CD44 is a non-kinase transmembrane glycoprotein overexpressed in tumor tissue, which activates cell signaling pathways, inducing cell proliferation and increases cell motility and survival (Chen et al., 2018). However the most selective CRC stem cell marker is the LGR5 (Leucine-rich repeat-containing G-protein coupled receptor 5), which is overexpressed in CRC relatively to the normal mucosa, this expression is detected in the early stages of cancer (Walker et al., 2011).

1.1.7 Early onset CRC

Early onset CRC, it is a heterogeneous disease that appears in young individuals between the ages of 20 to 49 years. Data shows that 15% of patients are diagnosed with CRC under the age of 50 years, with a mean age of 42.5 years (Chen et al., 2017). This percentage is increasing every year with or without family history, once approximately 20 to 25% of the cases have at least one 1st degree or 2nd degree family with CRC history. Colon and rectal cancer incidence rates are projected to increase by 90% and 124.2%, respectively, for patients with 20 to 34 years, and by 27.7% and 46% for individuals between 35 to 49 years by 2030 (Chen et al., 2017). Early onset CRC is clinically characterized by, in most cases, symptomatic patients presenting common symptoms like bleeding (59%), abdominal/rectal pain (35%), and obstruction (9%). The tumor usually (70-80%) is localized in distal to the splenic flexure of the intestine in which at least 40% involves the rectum. These CRC are frequently poorly differentiated, with mucin production and can develop synchronous or metachronous tumors, characteristics associated with inherited CRC (Perea et al., 2010). Also, these cases are associated with a more aggressive tumor phenotype, presenting increased resistance to conventional therapies. Interestingly, young patients with CRC tumor in III and IV stage of progression had a shorter duration of symptoms than patients with tumors in stage I and II. Therefore, the lack of early diagnose reflects on a bigger percentage of young patients that have been diagnosed with advanced stage CRC tumors, more than other age groups. As a consequence of delayed diagnosis, young CRC patients tend to develop distant metastases (mCRC), being, for this reason, less receptive to the treatments, leading to a greater tumor disease progression and consequently a higher risk of death, when compared to the other age groups (Perea et al., 2010). They present a poor overall survival (OS) and a low progression-free survival (PFS), among treated patients with mCRC. Sporadic CRC mutations in young individuals tend to cause more aggressive tumor phenotypes due to a higher prevalence of less favorable histological characteristics such as signet ring cell differentiation, venous invasion and perineural invasion (Stigliano et al., 2014). This somatic subtype develops based on the presence of common or rare genetic variants, with variable penetrance and still with unclear significance. Although is thought that the variants have a cumulative effect, while increasing the risk of colorectal cancer and possibly anticipating the age onset (Stigliano et al., 2014). On the other hand, germline mutations explain only a minority of the cases of early onset CRC, once only 15% to 20% of the patients have

germline mutations or and identified hereditary CRC syndrome (mutated *APC*,*TP53*, *MUTYH* genes), and half of these percentage correspond to patients with Lynch Syndrome (Murphy and Singal, 2018).

1.1.8 Identification of biomarkers

The CRC is one of the most prevalent cancers in the population with increasing percentages across the years mainly in the young individuals. Although each case of CRC can differ in the stage of the tumor, the mutations, the association or not with hereditary CRC syndromes and in the metastization, there is still a need to identify new potential biomarkers that make these cancer diseases more aggressive and less responsive to drugs. The prognostic biomarkers would be helpful in the path of target therapy for each patient.

1.1.8.1 *TCF7L2* gene

Transcription factor 7-like 2 (*TCF7L2*) gene, also known as *TCF-4*, it is a polymorphic and pleiotropic transcription factor which belongs to the TCF/LEF family and encodes a highly preserved domain named high mobility group (HMG) box DNA binding domain containing transcription factors (Angus-Hill et al., 2011). This gene is located on the long arm of chromosome 10 and its composed of 17 exons in which 5 are alternative by splicing, exons 4 and 13-16 (Figure 1.4), encoding for multiple different isoforms of the gene. *TCF7L2* gene plays a key role in the Wnt/β-catenin signaling pathway, due to its N-terminal end, that binds to nuclear β-catenin, functioning as a transcription factor complex activating gene transcription, as previously described, of genes implicated in proliferation and EMT (Bush et al., 2012). Previous studies reached a model for normal Wnt/β-catenin response, in cells with the TCF7L2 repressor, saying that there is a simultaneous activation of *TCF7L2* gene and inactivation of its proteins for an optimal transcriptional response to Wnt proteins (Tang et al., 2008). When the *TCF7L2* is mutated, usually due to a small nucleotide deletions or insertions (Duval et al., 1999), the gene loses the ability to function as a regulator resulting in an increased chance of developing colorectal cancer, due to its role in cancer cell proliferation, differentiation and metastasis development (Tang et al., 2008). Researchers have found in primary CRCs, TCF loss of function mutations that enhance cell growth in cell lines suggesting that TCF7L2 may function as a tumor suppressor (Angus-Hill et al., 2011).

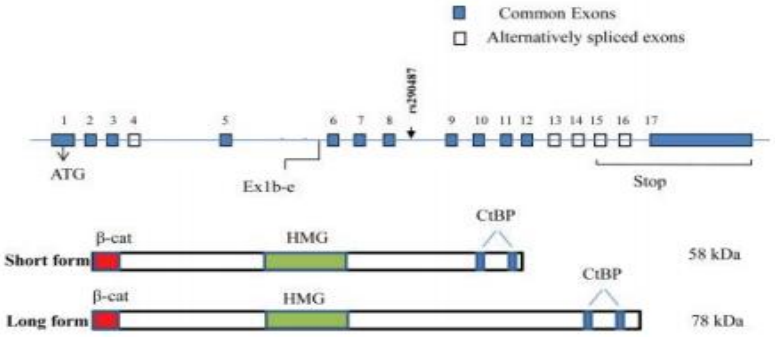


Figure 1.4- *TCF7L2* genetic structure with the exon positions and the respective transcript in the short and long form. (Figure adapted from Cingilli Vural, 2017)

1.1.8.1.1 *TCF7L2* alternative splicing

In *TCF7L2* gene are encoded multiple polymorphisms and in some cases single nucleotide polymorphisms (SNPs) which are variations of a single nucleotide in a specific DNA sequence, between species or chromosomes pairs of the same individual. In the gene are also encoded isoforms, by alternative splicing, a process where pre-mRNA molecules are segregated in different sequences and it is a fundamental step in differential gene expression. In TCF7L2 gene there are two regions of alternative splicing: exon 4, described by *Tomimaru et al.*, has an important role in the regulation of hepatocarcinome malignant phenotype; and exons 13-16 have the potential to originate thirteen different isoforms (Figure 1.5) divided into three groups (E, M and S) with different STOP sites. Hence TCF7L2 isoforms have cell-type-specifc distribution and differential gene regulation, therefore influencing tissue-specific Wnt responses (Weise et al., 2009).

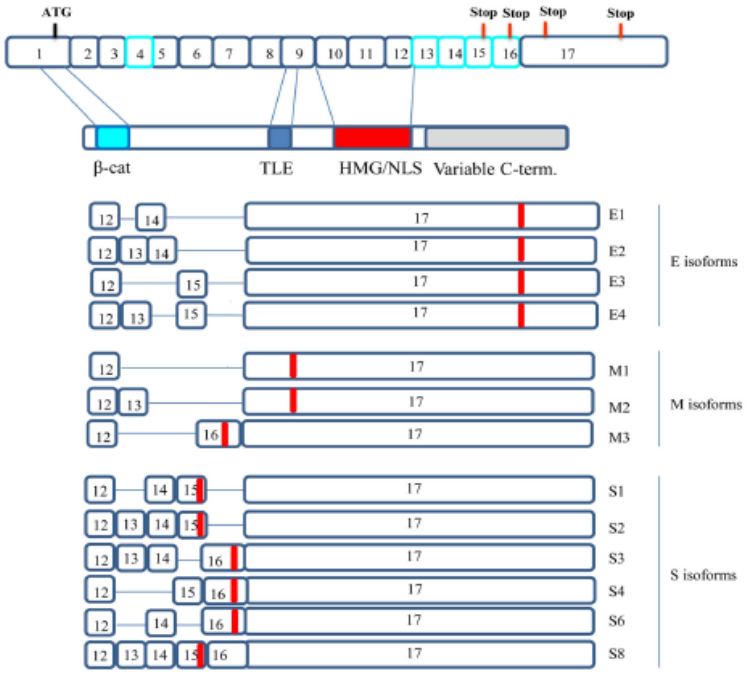


Figure 1.5- Illustration of the organization of 17 exons of *TCF7L2* gene and the respective thirteen isoforms combining the exons 13-16 after alternative splicing. ATG indicates the exon position of the translation start site and the STOP indicates the different translation stop sites. (Figure adapted from Jin, 2016)

1.1.9 Consensus molecular subtypes (CMS) of CRC

CRC has heterogeneous outcomes and for that reason it was necessary to segregate this tumors into four groups called consensus molecular subtypes (CMS1-4) which each one has different biological characteristics and gene expression patterns. CMS1 also named MSI Immune, represents 14% of all CRC, of which 2% of the patients have Lynch syndrome (Simoneaux, 2018), and is characterized by MSI tumors that have defective MMR system, a tumor microenvironment (TME) with immunogenic lymphocyte infiltration and CIMP phenotype with consequent *BRAF* mutations (Table 1.4). This tumor category has a good prognosis, until the relapse, in part due to the reduce proliferation-supporting cancer-associated fibroblast in the TME and the presence of immunity-promoting cells. This molecular subtype is present in high proportions in patients younger than 40 years old (Willauer et al., 2019). The canonical or CMS2 includes the major portion of the CRC cases

because it follows the canonical adenoma-carcinoma sequence and the gene expression consists in the loss of *APC*, activating mutation in *KRAS* and loss of *TP53*, a profile similar to the differentiated epithelial cell phenotype. This category is characterized by a high CIN, LOH and high somatic copy number alterations (SCNA), more frequently copy number losses in suppressor genes and gains in oncogenes (Table 4). This subtype has the best prognosis of the four, corresponding of 77% of 5-year survival rate (Simoneaux, 2018). CMS3 or the metabolic subtype is characterized by having CIN phenotype, low SCNA, an intermediate number of MSI cases and enrichment of multiple metabolism signatures as well as overexpression of *KRAS* that induces the cell metabolic adaptation (Table 1.4). Finally, the mesenchymal subtype or CMS4 is represented by the MSS tumors with CIN phenotype, high levels of SCNA and low hypermutation (Table 1.4). The TME is considered to be inflammatory due to a high expression of TGF-β leading to EMT and angiogenesis, these facts lead to a poor prognosis mainly because these cancers are often diagnosed at advanced stages (Simoneaux, 2018).

Table 1.4- Proposed taxonomy of CRC consensus molecular subtypes, segregated by gene expression based-molecular subtypes. (Table adapted from the article Guinney et al., 2015).

CMS1 MSI Immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
14%*	37%*	13%*	23%*
MSI, CIMP high, hypermutation	SCNA high	Mixed MSI status, SCNA low, CIMP low	SCNA high
BRAF mutations		*KRAS* mutations	
Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration, TGFβ activation, angiogenesis
Worse survival after relapse			Worse relapse-free and overall survival

Legend: CIMP (CpG Island Methylator Phenotype); MSI (microsatellite instability); SCNA (somatic copy number alterations); TGF (transforming growth factor); * (Percentage of all CRC cases)

1.1.10 Therapies for colorectal cancer

1.1.10.1 Conventional therapies

CRC treatment varies with disease progression and tumor stage, however, the common strategy for localized tumors, is surgery to remove the primary tumor and the possible existent lymphatic nodes. Nevertheless, the majorities of CRC cases are diagnosed in a stage of advanced disease or metastatic CRC, which decreases the effectiveness of treatment and diminish the survival rate. Normally, surgery is associated with the other standard treatments of radio and/or chemotherapy, that uses drugs such as 5-fluorouracil (5-FU) and leucovorin, oxaliplatin, capecitabine and irinotecan. As an example of one of the drugs, irinotecan or CPT-11, functions as an inhibitor of DNA topoisomerase I and induces single strand DNA breaks and replication arrest and is used for patients with advanced CRC that had become resistant to 5-FU treatment (Kerr and Midgley, 2001). Metastatic CRC is commonly resistant to conventional therapies, so it is usual to perform a combination of therapies in order to improve treatment results, however this combinations vary depending on the type

of tumor, stage and the metastasized local. In a case of CRC with liver metastasis is commonly used FOLFOX, which is the combination of 5-FU, leucovorin and oxaliplatin. In case of younger patients it is generally used FOLFIRI, corresponding to the name of the combination of 5-FU, leucovorin and irinotecan. In some cases in complement of neoadjuvant chemotherapy, the chemotherapy done before surgery in order to reduce tumor size, it is administered antiangiogenic drugs, like bevacizumab, and anti-EGFR therapy, such as cetuximab. The combination of treatments that act in different targets within cancer cells can give a better tumor regression and consequently better results, moreover with the advances in the research of molecular biology it has been discovered new mechanism-based target therapies for cancer cells that could be important as therapeutic targets (Figure 1.6) (Hanahan and Weinberg, 2011).

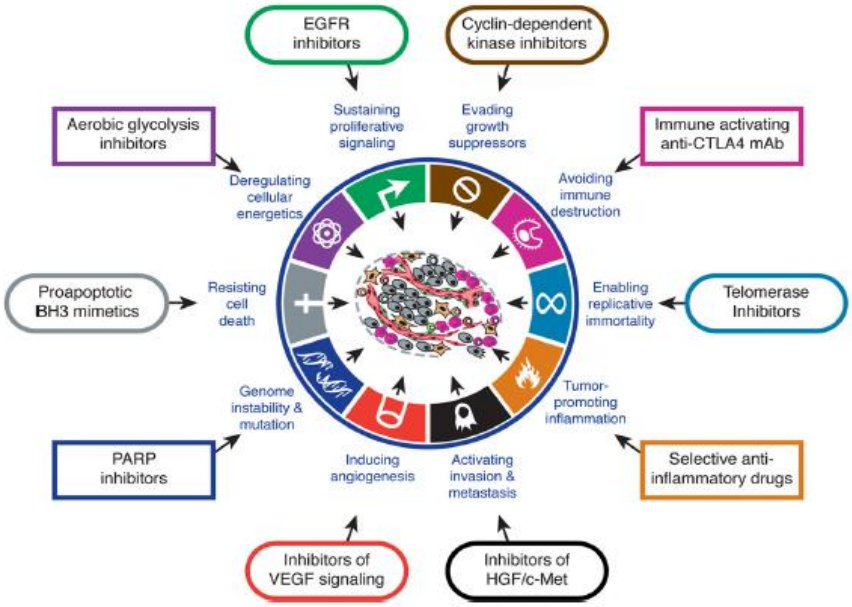


Figure 1.6- Therapeutic targeting of the Hallmarks of cancer (Figure taken from Hanahan and Weinberg, 2011)

Legend: EGFR (Epidermal growth factor receptor); anti-CTLA4 mAb (Ipilimumab, monoclonal antibody); HGF(Hepatocyte growth factor); c-Met (tyrosine-protein kinase Met); VEGF (Vascular endothelial growth factor); PARP (Poly (ADP-ribose) polymerase); BH3 (BH3-only BCL-2 family proteins are effectors of canonical mitochondrial apoptosis).

1.1.10.2 Hedgehog (Hh) signaling pathway inhibitors

The Hedgehog (Hh) signaling pathway has an important role in growth and differentiation in tissues, such as the gastrointestinal tract (Wu et al., 2017). *GLI1* (human glioma-associated oncogene homolog 1) and *GLI2* (human glioma-associated oncogene homolog 2) genes are transcription factors that regulate target genes at the distal end of the canonical Hh signaling pathway, via Smo (Smoothened) protein (Agyeman et al., 2014). Their expression is highly regulated with little expression detected in adult healthy tissues, but *GLI1* and *GLI2* genes are oncogenes that induce tumorigenesis and are constitutively activated in cancer cells. In these cases, there is a failure to terminate Hh/GLI signaling, leading to an increase of *GLI* genes activity. The role of Hh signaling

pathway in CRC remains unclear but some authors agree that the aberrant activation of this pathway induces CRC development.

GANT61 is an experimental agent in preclinical studies, functions as an inhibitor of GLI1 and has already showed extensive cytotoxicity in human models of colon cancer (Agyeman et al., 2014). The inhibition of this pathway could be very efficient in the treatment of CRC, although there is a need for a more extensive research in this area in order to understand the effect of Hh inhibitors in CRC.

Nowadays, the necessity of finding new CRC therapy combinations that could promote a more effective treatment reducing the aggressiveness of the conventional therapies is increasing and the research is turning its focus towards naturally-occurring dietary compounds with anticancer effects and less toxicity, the nutraceutics.

1.1.11 Nutraceutics

Nutraceutic is a term composed by the junction of nutrients and pharmaceutics, and stands for compounds present in foods that can provide medical or health benefits including prevention and treatment of diseases. Some natural products are seen has chemopreventive agents, acting in all stages of carcinogenesis, and these products can be nutrients, vitamins, probiotics or bioactive food components. The last one, occur naturally in low quantities in fruits, vegetables and grains.

1.1.11.1 Isothiocyanates (ITCs)

In the case of CRC, the consumption of cruciferous vegetables of the *Brassicaceae* family, such as broccoli and watercress, represent an intake in various bioactive components like folate, vitamin C, carotenoids and a high content of glucosinolates that has been associated with a reduced risk of CRC (Wu et al., 2013). The later are the precursors of isothiocyanates (ITCs) and indole-3-carbinol, after conversion into more bioactive products by plant myrosinase and/or gut microbiota activity (Pereira et al., 2017). However, the normal intake of these compounds in the diet is not sufficient to produce therapeutic effects to prevent tumor progression although being a promising chemopreventive measure. Previous studies showed the optimized conditions of supercritical fluid extraction for the selective isolation of ITCs, which is a procedure to recover the phytochemicals from vegetables while having minimal alteration of the interest product and their functional properties (Rodrigues et al., 2016).

Among all ITCs, sulforaphane (SFN) has popped out due to its antiproliferative effects and also its ability to induce apoptosis and impair metastatic CRC by acting in the reduction of cell migration, invasion and growth in anchorage-independent conditions (Pereira et al., 2017). SFN has a chemical structure 4-methyl-sulfinylbutyl isothiocyanate, and is a phytochemical that occurs in the form of a biological inactive precursor, glucoraphanin, that belongs to the group of the glucosinolates (Vanduchova, Anzenbacher and Anzenbacherova, 2018). High concentrations of SFN are present in the young broccoli sprouts. Many studies have been published about the benefits of SFN and besides

its antiproliferative effects, it is also described has an antioxidant agent (Vanduchova, Anzenbacher and Anzenbacherova, 2018).

1.2 Previous studies

As previously mentioned, *TCF7L2* gene has a very important role in CRC due to its function in the WNT signaling pathway, however its role as tumor suppressor or oncogene is still unclear. Another growing subject for the *TCF7L2* gene are the isoforms that result from the alternative splicing and are differentially expressed depending on the type of tissue having different physiological and pathophysiological effects (Le Bacquer et al., 2011). It has been reported that mutations in this gene may contribute to a higher risk to develop CRC (Tang et al., 2008). Studies of meta-analysis performed by Chen et al. indicated an association between rs7903146 *TCF7L2* polymorphism and risk of developing breast, prostate and colon cancers (Chen et al., 2013).

Previous studies performed by the group of Gastroenterology of UIPM-IPO, analyzed samples of tumor tissue, normal mucosa embedded in paraffin and peripheral blood from 131 individuals that fulfill at least one Bethesda criteria (BC), and they observed a difference between the frequencies of each genotype for the polymorphisms rs3814570 and rs10885395 of the *TCF7L2* gene in this BC group, when compared with the European and Iberian population frequencies. It was observed a higher frequency for the normal allele in homozygosity (C/C) and a reduced frequency for the alleles in heterozygosity (C/T). Moreover, in the same study it was noted that there was a clear tendency for the normal allele in homozygous to have higher polymorphism frequencies in the cohort of BC individuals than in a group of individuals with sporadic CRC, which means that they do not fulfill any BC and have no evidence to have an increased risk for CRC due to personal or familial history. Since the frequencies of the sporadic CRC individuals were similar to the controls of the Portuguese population and to the European and Iberian polymorphism frequencies, this suggests that the differences should be in the BC individuals. Taking these facts into account, there is a need to increase the cohort of individuals studied of each BC in order to observe if these frequencies maintain and also to analyze if these different frequencies are transversal to all BC, in tumor, normal mucosa and peripheral blood samples.

Previous studies, performed by the group of Gastroenterology of UIPM-IPO identified LOH in the *TCF7L2* gene, using three flanking markers, which were localized in the upstream and downstream regions of the gene as well as one intragenic. LOH was found to be associated to distant metastasis. Later, using MLPA and qPCR, the same group identified deletions in the promoter region and proximal exons of the *TCF7L2* gene associated with the development of distant synchronous metastases (Duarte, 2015). Using those techniques, the group was able to analyze the presence of allelic gains and deletions, in the tumor compared to normal mucosa, in the promoter region until exon1 and in the intron 2.

As early onset CRC is characterized by diagnosis at a late stage and being a cancer with poor prognosis due to the increased frequency of metastasis (Perea et al., 2010), considering the previous association made by the Gastroenterology group between polymorphisms in the promoter region of *TCF7L2* and the BC that includes diagnosis of CRC at young age (<50 years), there is an hypothesis

that alterations in the *TCF7L2* gene could increase the risk of CRC and the development of metastasis.

The group of Gastroenterology of UIPM-IPOLFG has been performing several studies involving cell lines that represent various types of CRC, such as LoVo cell line which is derived from supraventricular CRC metastasis. The group has been studying cell proliferation, migration, self-renewal and chemo-resistance of the cell lines in response to different drugs and also in combination with different nutraceutics, such as SFN. The results are promising in order to choose a different course of treatments depending on the cancer staging and to reduce the amount of drug used in a treatment when in combination with a nutraceutic, once the ITC-enriched extracts have a chemotherapeutic potential in CRC therapy (Pereira et al., 2017). To reinforce these previous studies and to clarify the mechanisms underlying response to treatment, there is a need to continue the evaluation of the gene expression of stemness and EMT markers, and *WNT*/β-catenin/*TCF7L2* signaling pathway genes when the LoVo cell line is treated with SFN and the combination of the nutraceutic and the conventional drug.

1.3 Aim of the Thesis

The present Master thesis comes as the final step of an extensive work performed by the group of Gastroenterology of UIPM-IPOLFG with the aim of identifying of biomarkers of colorectal cancer risk and metastasis, in particular in this study with a target in the early onset CRC. The group had already identified *TCF7L2* gene alterations as possible candidate biomarkers of CRC risk and metastasis in BC individuals. However there is the need to enlarge the patients' cohort and reinforce the association data as well as to clarify the etiology of this association. Moreover more adequate and personalized therapies are needed to improve outcomes in metastatic disease. Therefore this project had to main aims:

1) To investigate if specific polymorphisms and variations in copy-number and gene expression in the *TCF7L2* gene may contribute to a higher CRC risk, namely at a young age, and to metastatic disease.
2) To work towards a precision therapy to target metastatic disease, by continuing previous work performed by the group of Gastroenterology of UIPM-IPOLFG and evaluate the potential of GANT61, SFN and their combination in targeting the metastatic LoVo cell line.

2 Experimental Procedures

2.1 Biological material

For the molecular analysis of the polymorphisms rs3814570 and rs10885395 of the *TCF7L2* gene, localized in the promoter region of the gene (Figure 2.1), were used in this study, 16 DNA samples from formalin-fixed paraffin-embedded (FFPE) tumor tissue, 11 DNA samples from FFPE normal mucosa and 9 DNA samples from peripheral blood. For the polymorphism rs3845570 of the *TCF7L2* gene 93 DNA samples from FFPE tumor tissue, 41 DNA samples from FFPE normal mucosa and 171 DNA samples from peripheral blood were additionally analyzed. All these samples were obtained from patients diagnosed with CRC that fulfilled one of the Bethesda criteria. These samples were added to a group of cases previously studied making a total of samples analyzed of: 252 tumor samples; 201 normal mucosa samples; and 205 peripheral blood samples. Some tumor and normal mucosa samples for the rs3814570 polymorphism were analyzed by the Gastroenterology group from UIPM.

Also, for the molecular analysis of these polymorphisms were also studied 8 FFPE DNA samples from tumor tissue, 8 FFPE DNA samples from normal mucosa and 10 DNA samples from frozen biopsies, all from older individuals diagnosed with sporadic CRC.

For the copy number assays 28 DNA samples from peripheral blood from patients diagnosed with CRC that fulfilled BC were studied. Five DNA samples from peripheral blood from healthy individuals were included as reference samples. For the gene expression assays 43 RNA samples from peripheral blood from patients diagnosed with CRC that meet the BC were analyzed and were used as reference 6 RNA samples from peripheral blood from healthy individuals, representatives of the three genotypes for the polymorphism rs3845570 of the *TCF7L2* gene.

All the samples used in this study were previously extracted by the group of Gastroenterology from UIPM. DNA from FFPE tissues (tumor and normal mucosa) was extracted using the method of phenol chloroform and DNA from peripheral blood was extracted by using a salting-out method. More recently were extracted using a Maxwell® instrument according to the manufacturer's instructions and using the specific kits for each type of sample (Maxwell® RSC Whole Blood DNA Kit and Maxwell® RCS Buffy DNA FFPE Kit). The quantification of the samples was previously performed by 0.8%(w/v) agarose gel (Appendix I) and more recently using the Qubit™ 2.0 fluorometer, while for the other samples it was used the NanoDrop™ 2000 (Thermo Scientific)

Finally, it was also used the CRC LoVo cell line, which was subject to different treatments, to investigate the gene expression of several genes that contributed for the EMT, stemness, Wnt/β-catenin signaling pathway and cell cycle. The CRC LoVo cell line came from CRC supraventricular metastasis and is characterized as a MSI tumor.

Each individual included in this study was followed, or is currently followed, in the CRC risk consult of the Instituto de Oncologia Francisco Gentil de Lisboa, EPE, with informed consent of every patient. Every DNA or RNA sample was stored by the Gastroenterology from UIPM.

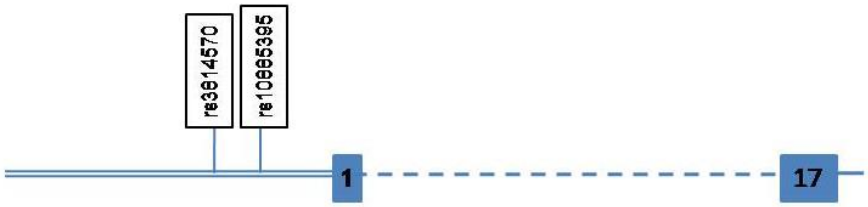


Figure 2.1- Schematic representation of the localization of the polymorphisms rs3814570 and rs10885395 in the *TFC7L2* gene. Each rectangle represents an exon. The dotted line represents the rest of the gene. The double line represents the promoter region.

2.2 Polymerase chain reaction

A polymerase chain reaction (PCR) is a technique used to amplify a specific DNA segment by a DNA polymerase, in order to create thousands of copies of the same sequence. There are a wide variety of polymerases that can be used in this reaction although the more commonly used, came from a thermophilic bacteria, *Thermus aquaticus*, named Taq. In addition to the polymerase, this reaction also requires four components equally essential: DNA template; deoxyribonucleotide triphosphate (dNTPs); primers; and ion magnesium (Mg^{2+}). PCR methods rely on thermo cycling, where the reactants are repeatedly exposed to cycles of heating and cooling that allows different temperature-dependent reactions. PCR follows three common steps: denaturation, causing the DNA melting; annealing, that leads to the hybridization of the primer to the template and this step requires different temperatures regarding the primer fragments in use; and elongation, that enables polymerase to synthesize new DNA strands. These steps are included in a series of 30 to 40 cycles that can vary with the size of the DNA fragment. Each PCR has its conditions optimized in order to obtain a good yield without unspecific bands and, for that reason, the concentration of $MgCl_2$, the quantity of primer and the annealing temperature can vary.

2.2.1 Amplification of genomic DNA by PCR

The samples of DNA used in this project were amplified in a reactional volume of 12.5µL with *Amplitaq Gold*™ kit (*Applied Biosystems*™*)*. The reaction consisted in 1µl of DNA, with a concentration of 80ng/µl, and 11.5µL of a mixture composed by: 1.25µL of *10x PCR Gold Buffer*; 1µl of dNTPs (Solis BioDyne, 100nM); *$MgCl_2$ Solution* (25mM); primer forward and reverse in a concentration of 10pmol/µL; 0,08µl of Taq (5U/µL); and ddH_2O in order to achieve the final volume.

2.2.1.1 Expand Long PCR reaction

Expand™ Long Template PCR System (Roche) is a PCR reaction used to amplify large DNA fragments, products between 5 to 20 kb. This reaction has higher yields and threefold fidelity because the kit is composed by an enzyme mix that contains a thermostable Taq Polymerase and DNA Polymerase with proofreading activity. Long PCR products require a specific buffer out of the three different buffers that differ in composition, underlying the importance of reaction optimization. Each

buffer is used to amplify different fragment sizes: Expand Long Template Buffer 1 (17.5mM $MgCl_2$, 10x) for fragments to 9kb; Expand Long Template Buffer 2 (27.5 mM $MgCl_2$, 10x) for fragments with 9 to 12kb; and Expand Long Template Buffer 3 (27.5 mM MgCl2 and detergents, 10x) for templates higher than 12kb.

For each reaction it was necessary to make two different mixes, for a final volume up 5.75µL each, and add those mixes to a 1µL of DNA (80ng/µL). The first mix was composed by: 1.75µL of dNTPs; 0.4µL of forward and reverse primer; and ddH_2O. The second mix was composed by: 1.25µL of Expand Long Template Buffer 3; 0.19 to 0.21µL of Expand™ Long Template Enzyme Mix (3.75 U); and ddH_2O. The program setup used is described in Appendix II (Table B).

All of these reactions (2.2.1 and 2.2.1.1) were made in the *Veriti™ Dx Thermo Cycler Applied Biosystems™*, following the established conditions of the program (Appendix II, Tables II.1 and II.2).

The PCR product should be stored at 4°C until the evaluation of the PCR reaction is made and then preserved at -20°C.

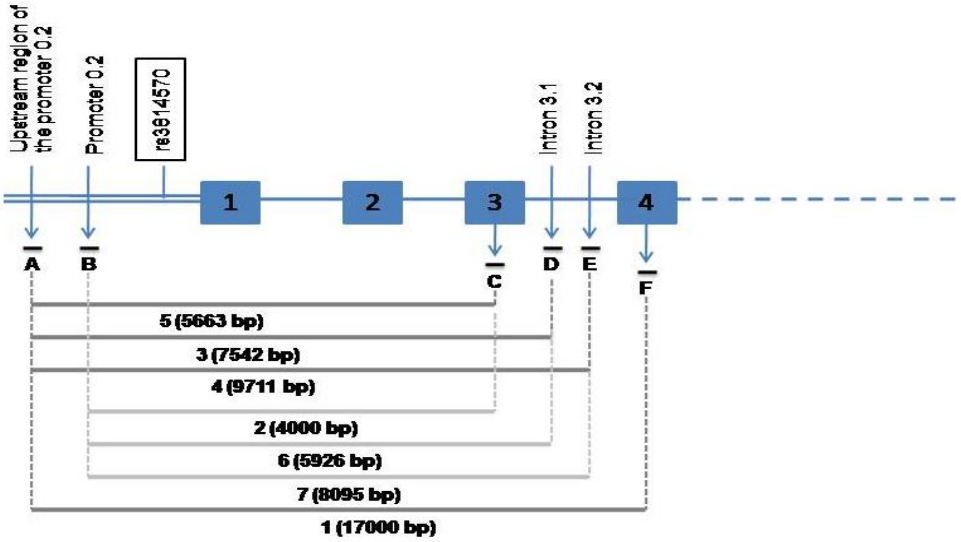


Figure 2.2- Schematic representation of the localization of the primers in the *TFC7L2* gene used in the Expand long PCR and the 7 different regions amplified. Each rectangle represents an exon and each filled line an intron. The dotted line represents the rest of the gene.

Legend: A-TCF7L2 Ups.P0.2; B- TCF7L2 P0.2; C- TCF7L2 Ex3 CN; D-TCF7L2 Int3.1; E- TCF7L2 Int3.2; F-TCF7L2 Ex4 CN.

2.2.2 PCR reaction optimization

For each PCR reaction it is necessary to do an optimization of the conditions in order to obtain an efficient amplification of the DNA fragment. The optimization involves the definition of the optimal magnesium concentration, annealing temperature, number of cycles and DNA polymerase kit. The annealing temperature for each primer can be calculated by using equation 1, where the letters A, T, C and G represent the nitrogen bases (adenine, timine, cytosine and guanine, respectively). One of the annealing temperatures normally tested is the average between the annealing temperatures of the forward and reverse primers.

$$Annealing\ Temperature\ (ºC) = 2(A + T) + 4(C + G)$$ (Equation 1)

It was necessary to do the optimization of large fragments (>4000pb) with the Expand™ Long Template PCR System (*Roche*) using different temperatures, different buffers and with or without extra magnesium. The aim was to obtain a clear band that represents the amplified DNA fragment, without the presence of unspecific bands.

2.2.3 Evaluation of the PCR efficiency by electrophoresis agarose gel

The evaluation of the PCR efficiency was made by electrophoresis in an agarose gel. This technique is based on the migration and segregation of the amplified DNA segments under the influence of an electric field, determined by its molecular weight. In order to visualize the fragments was added ethidium bromide to the gel, which is a nucleic acid intercalating agent that emits fluorescence once in the presence of UV radiation.

For the PCR products with <1000bp it was performed an electrophoresis using a 2%(w/v) agarose gel in order to evaluate yield and specificity of the amplified DNA fragment. For each PCR product it was prepared a mixture of 7µl of *Orange 1x* (loading buffer) and 3µl of the PCR product, which were loaded in each well of the gel. To ensure the molecular weight of the fragments it was also loaded, in a different well, 6.5µl of the molecular weight marker *GeneRuler™ 50bp DNA Ladder (Thermofisher Scientific™, 0.5*µg/µL*).* The electrophoresis was performed in a horizontal tin filled with *TBE* (Tris-Borate-EDTA) *buffer 1x,* at a voltage of 140V for at least 25 minutes. After the electrophoresis, the gel was visualized in the *BioDocAnalyze (Biometra)* transilluminator. All the solutions used in this step are reported in Appendix I.

For PCR products >4000bp it was performed electrophoresis using a 1%(w/v) agarose gel and for each PCR product it was prepared a mixture of 11.7µl of *Orange 1x* (loading buffer) and 5µl of the PCR product. It were used two weight markers (GeneRuler 50pb DNA Ladder and 1kb DNA Ladder), in order to reveal the bands that represent the amplified fragment. This electrophoresis is performed at 140mV for at least 1:30h. All the solutions used in this step are reported in Appendix I.

For PCR products >8000bp it was performed electrophoresis using a 0.7%(w/v) agarose gel (*SeaKem® GTG®, Genetic Technology Grade™*) in order to increase band segregation. It was used the same conditions to prepare the PCR product for loading into the gel (11.7µl of *Orange 1x* and 5µl of the PCR product), the same weight markers (GeneRuler 50pb DNA Ladder and 1kb DNA Ladder) and the same conditions of run (140mV for 1:30h minimum). All the solutions used in this step are reported in Appendix I.

2.3 Sanger sequencing

Sanger sequencing is a DNA sequencing method used to determine the sequence of nucleotide bases in a specific fragment of DNA. This method is a process of random selective integration of dNTPs and chain-terminating dideoxynucleotides phosphate (ddNTPs) by the DNA polymerase, during DNA *in vitro* replication, leading to different DNA fragment sizes due to the lack of

3'-OH group that is necessary to continue the chain. Each ddNTP is associated with a fluorophore, detected by a laser that translates the fluorescent intensity into a peak, and using specific software all peaks are assembled in an electropherogram, therefore getting the nucleotide sequence of each sample.

2.3.1 Purification of the PCR product

2.3.1.1 Enzymatic purification

Before the sequencing reaction it was necessary to do a step of enzymatic purification of the amplified material in order to eliminate any dNTPs or primers that were not consumed during the PCR reaction. Hence in each tube with PCR product it was added 1µL of *Thermosensitive Alkaline Phosphatase* (FastAP) (*1 U/µL Thermo Scientific*™) and 0.5µL de *Exonuclease I* (Exo I) *(20 U/µL Thermo Scientific*™*).* This reaction takes place in a thermocycler at 37°C for 15 minutes followed by a step of inactivation of the enzyme during 15 minutes at 85°C. Then the tubes should be preserved at -20°C until further use.

2.3.1.2 Purification from cutted bands of agarose gel

When the aim is to identify specific isoforms of a gene it is necessary to cut off from the agarose gel, after fragment separation by electrophoresis, the bands that represent that fragment to proceed to the sequencing reaction. In order to extract and purificate the DNA from the agarose gel it was used the *QIAquick Gel Extraction Kit* (QUIAGEN), which uses spin columns with a silica membrane assembly for binding of DNA in a high-salt buffer and elution with low-salt buffer or water(Qiagen.com, 2013). The purification procedure removes primers, nucleotides, enzymes, mineral oils, salts, agarose, ethidium bromide and other impurities from PCR.

The bands of interest were excised from the agarose gel with a scalpel under UV light and placed in a different eppendorf-like tube of 1.5mL. The purification was performed as described by the manufacturer (Appendix III).

2.3.2 Sequencing reaction

For each amplified sample it was made two sequencing reactions, one for the forward and the other one for the reverse strand. These reactions use a constant volume of 2µL, respectively, of sequencing buffer 5x *BigDye™ Terminator v1.1 (Applied Biosystems™)* and primer, which needs to be in a concentration of 1.6pmol/µL. The reaction requires a balance between the quantity of PCR product and *BigDye™ Terminator v1.1* mix *(Applied Biosystems™)* and that depends on the size of the amplified fragment and on the intensity of the band obtained in the previous agarose gel electrophoresis. For instance, in the presence of a fragment with higher molecular weight it was necessary to add more volume of PCR product, in case of two fragments with the same molecular weight it was necessary to add less volume of PCR product to the one with the more intense band. The remaining volume was filled with ddH_2O in order to make a total volume of 20µl. The sequencing reaction took place in a thermocycler using the program described in Appendix II, Table II.3. At the

end, the samples were preserved at 4°C no more than 24h until the DNA precipitation and purification was made.

2.3.3 DNA precipitation and purification

After the sequencing reaction, it was made a step of DNA precipitation and purification with the aim of eliminating any component that could interfere with the capillary electrophoresis, based in a salting-out method. It was used a mix solution composed by: absolute ethanol; sodium acetate and EDTA (ethylenediamine tetra acetic acid). The ethanol and sodium acetate were necessary to neutralize the acid nucleic charges, reduce the solubility and promote the DNA precipitation. On the other hand, EDTA is a chelating agent and was used to stop the enzyme activity. It was necessary to do a spin down of this mix solution in order to lodge in the bottom of the tube the excess of salt that could influence the capillary electrophoresis. Then, 54µL of the mix were distributed to new eppendorf-like tubes, the total content of the sequencing product was added and the samples were mixed by vortexing followed by incubation at room temperature for 15 minutes. This incubation time is very important in order to let the small fragments precipitate. Then the samples were centrifuged, the supernatant was carefully removed and ethanol 70%(v/v) was added to each sample in order to wash the DNA pellet by new centrifugation. Later the supernatant was again removed and the DNA pellet was dried at 37°C and stored at 4°C. The complete protocol of this step is described in Appendix IV.

2.3.4 Automatic sequencing

To the dried DNA pellet it was added 17µL of *HI-DI*™ formamide *(Applied Biosystems™)*, the samples were mixed in the vortex and the total volume was transferred to a distinct well of a microplate of 96 wells (*Platemax, Axygen*). Next, the plate was covered and placed in a thermocycler at 95°C for 5 minutes, in order to occur sample denaturation, and then was placed on ice for a few minutes followed by a spin down at room temperature at 1200rpm to make sure the samples were in the bottom of the well and without air bubbles. Finally, the plastic cover was removed, the plate was assembled and placed in the tray of the *ABI Prism™ 3130 Genetic Analyzer (Applied Biosystems™)*.

2.3.5 Results Analysis

It was used the *Sequencing Analysis v.6 (Applied Biosystems™)* software to obtain an electropherogram from each sample. The sequences were analyzed in comparison to the reference DNA sequence for the studied gene, which was obtained in the *Ensembl* database (Human genome built, GRCh38.p12), accessed through the site: https://www.ensembl.org/index.htmL.

2.4 Molecular analysis of *TCF7L2* gene polymorphisms

In this study were molecularly analyzed four polymorphisms of the *TCF7L2* gene, two of them located in the 5'UTR region of the gene (rs3814570 and rs10885395) and the other two in the exon 1 (rs373426839 and rs76088094). In order to perform these analysis it was used specific sets of primers, designed by the group of Gastroenterology of UIPM, two different pairs of primers for the 5'UTR polymorphisms and a pair of primers for the exon 1 that comprises both polymorphisms. The

PCRs were previously optimized for each primer and performed as described in 2.2, followed by Sanger sequencing as described in 2.3. All of the information about the primers and the conditions for the PCR are reported in the Appendix V (Tables V.1 and V.2). It was analyzed DNA samples from tumor and normal mucosa tissue and peripheral blood. For some cases of pared samples tumor/normal mucosa/blood it was necessary to repeat the analysis and confirm using the reverse primer in order to cross out any suspicious result.

2.5 Amplification of an extended region of *TCF7L2* gene

This assay was performed in order to amplify large fragments of the *TCF7L2* gene, mainly the promoter region until exon 4. The aim was to amplify the different isoforms of the *TCF7L2* gene in order to detect deletions or gains in the gene by seeing changes between bands after Expand Long PCR amplification and see those alterations in the sequence of the respective sample. It was used several combinations of primers forward and reverse: one with 7542bp composed by the *TCF7L2* upstream region of the promoter to intron 3; a second fragment with 9711bp which comprises the same upstream region of *TCF7L2* promoter to a different part of intron 3; a third fragment with around 4000bp which comprises the promoter region of the gene to exon 3;a fourth fragment with approximately 17000pb which included the exon3 until the exon4; and a fifth fragment with 5663bp composed by the upstream region of the TCF7L2 promoter to exon 3. The localization of the primers is represented in the figure 2.2. The primers used are described in Appendix V, Table V.3.

2.6 Quantitative PCR

Quantitative PCR (qPCR) or real-time PCR is a technique similar to the classic PCR that uses the method of DNA amplification to determine relative or absolute quantities of a sequence in a sample, by using a fluorescent reporter that binds to dsDNA. This technique enables measurements during the exponential phase of the reaction due to the presence of a fluorescent signal that proportionally increases with the quantity of amplified DNA. For each reaction of qPCR is determined the threshold cycle (C_t), which is a relative measure of the target concentration in the PCR establishing the minimal number of cycles necessary to measure fluorescence above background levels. The C_t value allows determining sample quantification present in the beginning of the reaction.

This quantification can be: absolute or relative. The absolute quantification is used to quantify samples whose concentration is unknown, by comparison with a standard curve. The relative quantification is obtained due to a normalization between the target gene and the reference gene (Equation 2), usually it is used a housekeeping gene because it is equally expressed in every cell. Then is determined the normalization between the target sample and the reference sample (Equation 3) in order to reach the expression ratio (Equation 4). In this study, was performed a relative quantification with different reference genes for both copy number and gene expression assays.

$$\Delta Ct = Ct\ target\ gene - Ct\ reference\ gene \quad \text{(Equation 2)}$$

$$\Delta\Delta Ct = \Delta Ct\ target\ sample - \Delta Ct\ reference\ sample \quad \text{(Equation 3)}$$

$$Expression\ ratio = 2^{-\Delta\Delta Ct}$$ (Equation 4)

2.6.1 Fluorometric quantification of DNA and RNA

In order to achieve a more precise and credible quantification of the DNA and RNA samples used in this study, it was performed a quantification using the Qubit™ 2.0 (Invitrogen). This is a fluorometric method that uses target-selective dyes that emit fluorescence when associated to DNA or RNA. This method is more trustworthy than the UV absorbance because it does not quantifies possible contaminants that could be present in the sample including DNA/RNA contamination. It was used the Qubit™ dsDNA HS (double-stranded DNA high-sensitivity) assay which detects DNA quantity lower than 100ng. To quantify RNA it was used the Qubit™ Broad range (BR) assay which has a quantification range of 20 to 1000ng.

The detailed protocol is specified in Appendix VI.

The information obtained with these quantifications was used to obtain a concentration of 5ng/µL and then these dilutions were again quantified in order to ensure that the dilution was correct. The quantification of the RNA samples was used for cDNA synthesis by RT-PCR.

2.6.2 Copy number variation

Copy number variation was detected by using the *Applied Biosystems*™ Taqman Copy Number Assays that are composed by a standard minor groove binder (MGB) probe necessary to evaluate the copy number of genomic DNA targets. It is also necessary to have a reference assay that has the function to detect a sequence that is known to be present in two copies in the diploid genome.

These assays have two options of running: in singleplex, when the copy number assay and the reference assay are in different wells; and in multiplex, when the copy number assay and the reference assay are in the same well being detected at the same time in a duplex qPCR reaction.

2.6.2.1 Copy number assay

To determine the copy number variation in multiplex it was made a mix, for each Taqman assay, for a total volume of 18µL composed by: 10µL of the Universal PCR Master Mix (*Applied Biosystems*™); 1µL of the Taqman Copy Number Assay (*Applied Biosystems*™); 1µL of Taqman Copy Number Reference Assay, RNase P (*Applied Biosystems*™); and ddH_2O to make the total volume. This mix was pipetted into the bottom of each well of a PCR-96 well microplate (*Axygen*). Then, into every well, is pipetted 4µL of DNA in a concentration of 5ng/µL, in triplicates for each sample, making the total volume of 20µL per well. The DNA was previously quantified by a fluorometric method described in 2.6.1. It is necessary to count with one well per assay for a negative control, where is putted 16µL of mix and 4µL of ddH_2O in order to discard the presence of contaminants. Next, the plate was sealed with Platemax UltraClear Sealing Film (*Axygen*), centrifuged at 1200rpm during 1 minute at room temperature, and was placed into the QuantStudio 5 (*Applied*

Biosystems™) where was previously programmed the software (Appendix II, Table II.5) and was made the design of the plate.

To determine the copy number variation in singleplex, in each well, was pipetted a total volume of 20µL composed by: 10µL of Universal PCR Master Mix; 1µL of the Taqman Copy Number Assay (*Applied Biosystems*™); 7µL of ddH_2O; and 2µL of DNA.

2.6.2.2 Results analysis

After the reaction was finished the results were exported from the QuantStudio 5 (*Applied Biosystems*™) to an Excel document (*Microsoft*) and the results were analyzed with resource of the Equations 1, 2 and 3.

2.6.3 Gene expression

Gene expression was detected by using the *Applied Biosystems* Taqman Gene Expression Assays that are composed by a pair of unlabeled PCR primers, a FAM or VIC dye label on the 5'end minor groove binder (MGB) and nonfluorescent quencher (NFQ) on the 3'end. This assay was made in order to measure the expression of a gene in comparison to an endogenous control, GUSB.

2.6.3.1 cDNA synthesis by reverse PCR (RT-PCR)

RT-PCR is a reaction that converts a single stranded RNA molecule into complementary DNA (cDNA). This reaction uses random primers that bound to the mRNA and a reverse transcriptase, a DNA polymerase that depends on RNA, and amplifies RNA into cDNA.

First, it was made a mix solution, for a total volume of 7.75µL per sample, composed by: RNA in a concentration of 1000ng for a maximum volume of 7.25µL; 0.5µL of random primers (3 µg/µL, *Roche*); and diethyl pyrocarbonate (DEPC) water (*MERK*) to make the final volume. The RNA samples were previously quantified using a fluorometric method as described in 2.6.2. The mix solution was made on eppendorf-like tubes of 0.2mL and was put on the thermocycler during 10 minutes at 70ºC. During this time, it was prepared a second mix for a total volume of 12.25µL, with: 4µL of buffer 4x First Strand (Invitrogen); 4µL of dNTPs (Solis BioDyne, 100nM); 2µL of dithiothreitol (DTT) (Invitrogen) that acts as an enzyme stabilizer; 0.75µL of RNaseOut™ (Invitrogen), a ribonuclease that has the role to inactivate possible RNases that could be present in the reaction; 1µL of Superscript II (Invitrogen), which is the enzyme reverse transcriptase; and 0.5µL of DEPC water. After 10 minutes in thermocycler, the program was paused to add 12.25µL of the second mix in each tube and then the program continued. The amplification conditions used in this reaction are reported in the Appendix II, Table II.4.

2.6.3.2 Gene expression assays

To determine the gene expression was made a mix solution of 18µL composed by: 10µL Universal PCR Master Mix (Applied Biosystems™); 1µL of Taqman Gene Expression Assay (Applied Biosystems™); and water. This mix was pipetted into the PCR-96 well microplate (*Axygen*) and then

was pipetted 2µL of cDNA, in triplicate, of each sample. The way of preparing the microplate is equal as described in 2.6.2.1, the plate was next placed in the QuantStudio 5 (*Applied Biosystems*™) where was previously programmed the software (Appendix II, Table II.5) and was made the design of the plate.

2.6.3.3 Results analysis

After the reaction was finished the results were exported from the QuantStudio 5 (*Applied Biosystems*™) to an Excel document (*Microsoft*) and the results were analyzed with resource of the Equations 1, 2 and 3.

2.7 Copy number variation and relative gene expression of the *TCF7L2* gene

In this study was performed the copy number variation and relative gene expression analyses of the *TCF7L2* gene in a group of individuals diagnosed with CRC, at any age and that meet the BC. It was used DNA samples from peripheral blood from these individuals for the copy number variation assays, while for the gene expression assays it was used RNA samples from peripheral blood of the BC individuals, from which was synthesized cDNA. For the copy number variation analysis was amplified the following regions: upstream of exon 1; exon 1; intron 1 to exon 2; intron 4 to the proximal region of the rs7901695; intron 16 to intron 17 (Figure 2.3). For the relative gene expression analysis was tested the expression of the regions exons 1 to 2 and exons 7 to 8. It was also analyzed a group of healthy individuals that belong to LS families, to have a comparison between healthy samples and CRC samples.

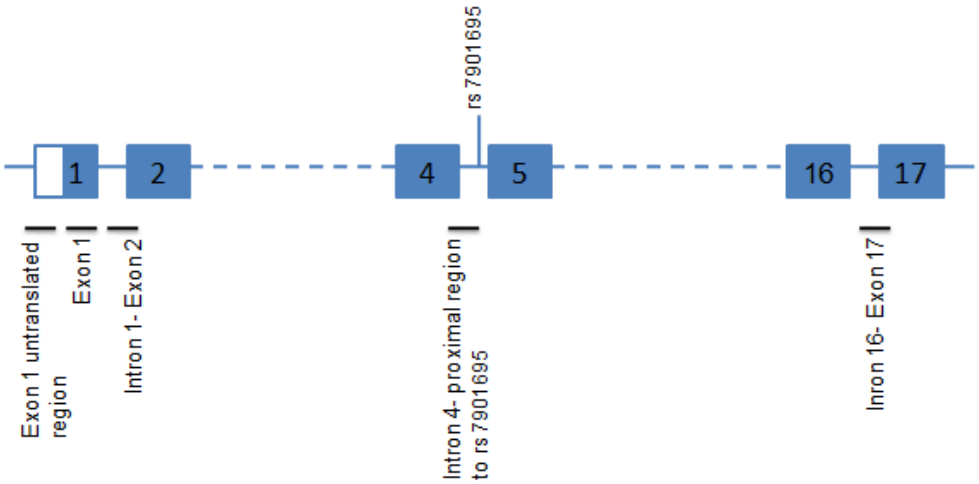


Figure 2.3- Schematic representation of the localization of the copy number assays in the *TFC7L2* gene. Each rectangle represents an exon and each filled line an intron. The dotted line represents the rest of the gene.

2.8 Cell-based Assays

2.8.1 Cellular subculture

2.8.1.1 Passaging of adherent cells

To maintain growing cells it was necessary to, when needed, make new dilutions of cells in fresh medium to any lineage. For that reason, in the laminar flow chamber, the first step was to take out the old and used medium with a pipette. This medium has an acidity indicator called phenol red, that changes the medium color of red to yellow in the presence of acidity. Then it was added a certain volume of Dulbecco's phosphate buffered saline (DPBS) *(Gibco™)* to remove the remains of medium and dead cells that could be attached to the monolayer of cells, this volume changes with the size of the t-flask (Table 2.1). This volume needs to be added always on the corner or base of the flask, on the opposite side of the monolayer wall, and then the t-flask was rocked back and forth several times. Then the DPBS was removed and discarded with a pipette and was added the respective volume of Trypsin 0.1%v/v (*Invitrogen™*) (Appendix VII), once again in the opposite corner of the monolayer wall. The t-flask was incubated at 37°C during 5 minutes until all the cells of the monolayer and clusters are dissociated, using inverted microscopy to observe the stage of the cells. Then was done a step of inactivation of the trypsin with 4 mL of fresh medium, in order to have a final concentration with a ratio of 1:5, is very important to mix the medium with the cells by doing up-and-down to dissociate all the remaining clusters. Then 4mL of the medium were taken out from the t-flask, leaving the remaining 1mL of medium with cells and was added 4mL of fresh medium, doing up-and-down again to distribute the cells in the medium.

Table 2.1- The work volume, of PBS and Trypsin volume needed to each type of flask.

Type of t-flask	Work volume	DPBS volume	Trypsin (0,1%v/v) volume
Small (25cm3)	5mL	2mL	1mL
Medium (75cm3)	10/12mL	3-4mL	3mL

2.8.1.2 Cells count

When was necessary to obtain a specific cell concentration for cellular assays it was made a cell count of the cells present in solution after trypsin inactivation. First, in an eppendorf-like tube was mixed 80µl of *Trypan Blue Stain* 0.4% (Gibco™, Invitrogen Corporation) and the 20µl of cells, for a concentration ratio of 1:5. Then to prepare the *Neubauer* Haemocytometer chamber (Figure 2.4), first, was placed a small drop of water in each side of the chamber in order to help fixing the coverslip, secondly was placed 10µl of the mixture of cells and Trypan Blue Stain in each side of the chamber, and finally, were counted all cells present in the total eight quadrants of the chamber (Figure 2.4) using inverted microscopy. At the end of the count, it is very important to carefully clean the chamber and the coverslip and redo the technique.

Cell concentration was obtained by using the equation 5:

$$Cell\ concentration\ (\frac{cells}{ml}) = \frac{\Sigma\ cell\ count}{8} \times Dilution\ factor \times 10000 \qquad \text{(Equation 5)}$$

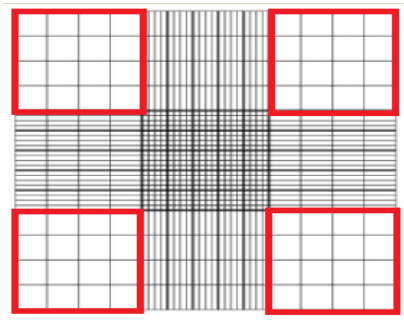


Figure 2.4- *Neubauer* Haemocytometer chamber. Each red square represents one quadrant of counting area. (Figure adapted from: http://www.microbehunter.com)

2.8.2 Quantitative PCR (qPCR)

As previously described in 2.6 it was performed a qPCR analysis in order to evaluate the amplification of certain genes in cell lines. It was used as endogenous control: GAPDH gene. The qPCR used cDNA, which was obtained by RT-PCR explained in the point 2.6.3.1 from RNA extracted from cells by using the RNeasy Mini Kit (*QIAGEN*), which the protocol is described in Appendix VIIII.

For the qPCR it was necessary to make a mix solution for each gene with a final volume of 13µL and this mix, depending on the gene that is been evaluated, needs to be done with reagents from Applied Biosystems™ or from Kapa Biosystems. In Appendix IX is described the reagent for each primer used in this study. The Applied Biosystems™ mix was composed by: 0.75µL of primer forward and 0.75µL of reverse primer of the target gene that can vary in concentration between 3 and 7.5pmol/µL according to the optimization previously made; 7.5µL of Master Mix; and water. The Kapa Biosystems mix was composed by the same solutions and the only difference was 0.3µL of ROX^{high} . These mixes were distributed into a PCR-96 well microplate (Axygen), into the bottom of each well was pipetted 13µL of mix. Next, was pipetted 2µL of cDNA with a concentration of 5ng/µL, in triplicates, into each well and 2µL of water into the negative control wells. Then, the plate was sealed with Platemax UltraClear Sealing Film (*Axygen*) and centrifuged at room temperature during 1 min at 1200rpms and after was placed in the QuantStudio 5 (ThermoFisher), where the program was previously set and the plate was designed. The program used in the qPCR is in the Appendix II (Table II.6).

2.8.2.1 Results Analysis

After the reaction was finished the results were exported from the QuantStudio 5 (*Applied Biosystems*™) to an Excel document (*Microsoft*) and the results were analyzed with resource to the equations 1, 2 and 3 from 2.6.

2.9 Relative gene expression of CRC genes of interest in LoVo cells

In this study was analyzed the relative gene expression, by qPCR, of several genes that have an important role in the CRC development in a LoVo cell line after treatment. The LoVo cells were cultivated in DMEM medium, supplemented with 10% (v/v) of FBS, 2mM of L-Glutamine and 1% (v/v) of PenStrep, in a moist atmosphere at 37ºC and 5% (v/v) CO_2. Then the cells were trypsinized and counted in order to perform the seeding in 24 well plates with a cell density of 100.000 cells per well. The cells were treated with Gant61 (15µM), with the combination of Gant61+SFN (15µM+12.5µM) and SFN (12.5µM) alone, during 24h, after seeding. As controls were used LoVo cells treated with DMSO+ETOH, in equal percentage of solvent used, and DMSO alone. It was investigated the expression of the following genes: EMT-associated markers (*VIM; CDH1; ZEB1; SNAIL; TGF-β*), stemness markers (*LGR5; PROM1; PTCH1; CD44*), Wnt/β-catenin signaling pathway genes (*TCF;* Isoform 1-6 of *TCF; AXINA2; SOX9*) and cell cycle markers (*CCNA2; CDKN1A)*. The expression of the GAPDH gene was used as a control. The aim was observe the changes in the gene expression of these genes induced by the different treatments and also to demonstrate the possibility of introducing the nutraceutics as combination of the conventional therapies.

3 Results and Discussion

3.1 Analysis of polymorphisms localized in the *TCF7L2* promoter region in individuals that meet the Bethesda criteria

In previous studies performed by the Gastroenterology group at UIPM-IPOLFG loss of heterozygosity of the *TCF7L2* gene and deletions in the promoter and proximal exons have been associated to the development of synchronous metastases (metastases that are developed in the first 6 months after the diagnosis of the primary tumor). Recently, preliminary data from this group suggested that the genotype of *TCF7L2* polymorphisms, localized in the promoter region, in tumors from patients fulfilling the BC, might be associated to increased CRC risk. Therefore, this data should be strengthened and, further research, including the analysis of additional DNA tumor samples from BC patients, was carried out in the present study to evaluate this hypothesis.

3.1.1 Association of *TCF7L2* polymorphisms in the promoter region to Bethesda criteria

Considering the haplotype formed by the 3 *TCF7L2* polymorphisms localized in the promoter region, that were analyzed in previous studies performed by the Gastroenterology group at UIPM-IPOLFG, in the present study the analysis was focused in only one of these polymorphisms, the rs3814570. Before starting the analysis of this polymorphism in additional samples, the analysis of the recent results obtained by the Gastroenterology group within this subject was updated. One hundred and forty four DNA samples from tumor tissue had been analyzed for the 3 polymorphisms, belonging to two groups of patients:1) individuals diagnosed with CRC that meet one of the Bethesda criteria; 2) a second group of individuals diagnosed with sporadic CRC.

The frequencies of each genotype for the rs3814570 polymorphism, obtained for the tumors from BC individuals, were compared with the world frequencies reported in the population data base *Ensembl*. Some differences were observed between the tumors from BC individuals and the information reported in the data base (Table 3.1). First, when comparing the frequencies obtained for the BC tumors with the European population frequencies, we observed a higher frequency of the homozygosity for the ancestral/normal allele (C/C) which appears to be compensated with a minor frequency for the heterozygous genotype (C/T): 93/144 (64.6%) versus 243/503 (48.3%) for the C/C genotype (p=0.001) and 43/144 (29.9%) versus 211/503 (41.9%) for the C/T genotype (p=0.009), respectively. The same differences were observed when comparing the frequencies observed in the BC tumors with the frequencies in the Iberian population: 93/144 (64.6%) versus 47/107 (43.9%) for C/C genotype (p=0.001) and 43/144 (29.9%) versus 45/107(42.1%) for the C/T genotype (p=0.063), respectively. As the homozygosity for the polymorphic allele (T/T) has a low frequency, the variations observed among different populations (between 1 and 15%, the majority under 10%) were not significant and no considerations could be drawn. These differences prompt the Gastroenterology group to analyze the frequency of the different genotypes in healthy individuals from the Portuguese population. A higher frequency for the C/C genotype and a lower frequency for the C/T genotype was

also observed for the BC tumors when compared with a group of healthy individuals from the Portuguese population: 93/144 (64.6%)versus 47/94 (50%) for the C/C genotype (p=0.025) and 43/144 (29.9%) versus 40/94 (42.6%) for the C/T genotype (p=0.045), respectively (Table 3.1). Considering the higher CRC risk associated to the Bethesda Criteria, one can hypothesize that the increased frequency of the C/C genotype may be associated with an increased risk of developing CRC.

Table 3.1- Genotype frequency for the rs3814570 and rs10885395 polymorphisms of the *TCF7L2* gene in the European, Iberian and Portuguese populations respectively. Also, the representation of the frequencies of each genotype for the polymorphism rs3814570 of the *TCF7L2* gene in individuals diagnosed with CRC and meets one of the BC and individuals with sporadic CRC.

Frequency of each genotype for the polymorphism	**rs3814570**			**rs10885395**		
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	**Normal (C/C)**	**Het. (C/T)**	**Hom. (T/T)**
In the European population	243/503 **(48.3%)**	211/503 **(41.9%)**	49/503 **(9.7%)**	252/503 **(50.1%)**	208/503 **(41.4%)**	43/503 **(8.5%)**
In the Iberian population	47/107 **(43.9%)**	45/107 **(42.1%)**	15/107 **(14.0%)**	50/107 **(46.7%)**	44/107 **(41.1%)**	13/107 **(12.1%)**
In the Portuguese population	47/94 **(50%)**	40/94 **(42.6%)**	7/94 **(7.5%)**	47/94 **(50%)**	40/94 **(42.6%)**	7/94 **(7.5%)**
In BC individuals	93/144 **(64.6%)**	43/144 **(29.9%)**	8/144 **(5.6%)**	-	-	-
In individuals with sporadic CRC	11/20 **(55%)**	9/20 **(45%)**	0/20 **(0%)**	-	-	-

In order to reinforce this hypothesis the genotype frequencies for rs were also analyzed in a series of tumors from sporadic CRC patients (age at diagnosis >60 years and without evidence of meeting the BC). However, the genotype frequencies in the sporadic CRC patients were very similar to the ones observed in the Portuguese, European and Iberian populations: 11/20 (55%) vs. 93/144 (64.6%), 9/20 (45%) vs. 43/144 (29.9%) for the C/C genotype (p=0.410) and heterozygous genotypes (p=0.173) of the rs3814570 polymorphism, respectively (Table 3.1).

The fact that there were frequency differences of each genotype between the BC individuals and the sporadic CRC, and the frequencies of the latter were similar to the population controls suggested that the differences rely on the fact that these people meet the Bethesda Criteria, as previously described (Santos et al., 2018). Therefore, the observed differences in the BC group do not seem to be related with the development of CRC in general.

3.1.2 Association of *TCF7L2* polymorphisms in the promoter region to Bethesda criteria BC1, BC2 and BC3

Instead, these alterations appear to be associated with a higher risk of developing CRC: at a young age, meeting the BC1; with synchronous, metachronous or other Lynch syndrome-related tumors meeting the BC2; with features of MSI-H tumors like mucous production of Crohn-like inflammatory infiltrate and age at diagnosis less than 60 years, meeting the BC3; with the existence of

one or more first-degree relatives diagnosed with CRC under age 50 meeting the BC4; or finally with a family history in first or second-degree relatives, regardless of age, meeting BC5.

Taking into account the characteristics of each BC, would these alterations be associated with all BC in general or would be particular of specific BC's? In order to find an answer to this question, previously analyzed BC tumors were stratified by BC (Table 3.2).

Table 3.2- Representation of the frequencies of each genotype for the rs3814570 polymorphism in tumors from patients fulfilling the BC, stratified by BC cases, from previous work performed by the Gastroenterology group from UIPM-IPOLFG.

BC	rs3814570			p value
	Normal (C/C)	Het. (C/T)	Hom. (T/T)	
BC1	44/64 **(68.8%)**	15/64 **(23.4%)**	5/64 **(7.8%)**	0.648
BC2	15/21 **(71.4%)**	5/21 **(23.8%)**	1/21 **(4.8%)**	
BC3	9/12 **(75%)**	2/12 **(16.7%)**	1/12 **(8.3%)**	
BC4	4/8 **(50%)**	4/8 **(50%)**	0/8 **(0%)**	
BC5	10/19 **(52.6%)**	8/19 **(42.1%)**	1/19 **(5.3%)**	

Tumors from patients fulfilling the BC1, BC2 and BC3 presented a higher frequency of the C/C genotype [(67.8%), 15/21 (71.4%) and 9/12 (75%)] and lower frequency for the C/T genotype [44/64 *vs*. 15/64 (23.4%), 5/21 (23.8%) and 2/12 (16.7%)], respectively (Table 3.2). In contrast, BC4 and BC5 appeared to present similar frequencies to the ones observed in the Portuguese population 4/8 (50%) and 10/19 (52.6%) for the C/C genotype and 4/8 (50%) and 8/19 (42.1%) for the C/T genotype (Table 3.2). These results suggest that the possible increased risk for developing CRC associated to the rs3814570 polymorphism is preferentially related with BC1, BC2 and BC3 criteria and not with BC as a whole.

Looking to the number of cases analyzed for each BC, the BC2, BC3, BC4 and BC5 as well as the group of sporadic CRC were poorly represented therefore the analysis of additional samples performed in the present study, was focused in these groups, especially in BC4. In particular, 59 samples have been analyzed: 8 DNA tumor samples from sporadic CRC and 7, 15, 20 and 9 DNA tumor samples from BC2, BC3, BC4 and BC5, respectively. The frequencies of each genotype obtained for rs3814570 are represented in table 3.3.

First, the increase in the number of sporadic CRC cases reinforced the similarity of the genotype frequencies with those obtained for the healthy individuals from the Portuguese population [15/28 (53.6%), 12/28 (42.9%) and 1/28 (3.6%) *vs*. 47/94 (50%), 40/94 (42.6%) and 7/94 (7.5%) for the C/C, C/T and T/T genotypes, respectively] (Table 3.1 and 3.3).

Regarding the genotype frequencies for each BC, after the enlargement of the cohort, when comparing the results depicted in tables 3.2 and 3.3 it is possible to observe that the frequencies obtained for BC4 and BC5 were confirmed to be similar to sporadic CRC and to the healthy individuals. Notably, the differences between BC1, BC2 and BC3 comparing with BC4 and BC5, were strengthened. By grouping BC1 with BC2 and BC3, considering the similarity among genotype frequencies, and compared with BC4 and BC5 combined, we could observe that the former showed a

significantly higher frequency of the C/C genotype and lower frequency of the C/T genotype when compared with the latter [83/120 (69.2%), 29/120 (24.2%) and 8/120 (6.7%)*vs*. 27/57 (47.4%), 27/57 (47.4%) and 3/57 (5.3%) for the C/C, C/T and T/T genotypes, respectively (P=0.005)] (Table 3.4). These results suggest that the higher frequency of the C/C genotype and lower frequency of the C/T genotype for the rs3814570, localized in the *TCF7L2* promoter region, is associated to a personal rather than to a familial history of CRC. This raises the hypothesis that this loss of heterozygosity may occur more frequently in the cells of the intestinal epithelium than at a germline level. If it occurred frequently at a germline level we would maybe observe a reduction of the heterozygous frequency for the BC4 and BC5 as these criteria involve family history of CRC and specific germline alterations should segregate with the disease. In contrast BC1, BC2 and BC3 are associated with personal history of CRC (young age at diagnosis, presence of synchronous or metachronous tumors, or histological features typical of Lynch syndrome tumors under the age of 60 years).

Table 3.3- Representation of the frequencies of each genotype for the rs3814570 polymorphism in tumors from patients with sporadic CRC and fulfilling the BC, in the latter stratified by BC, after the analysis of 59 additional samples included in the present study.

Sporadic CRC/ BC	**rs3814570**			**p value**
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	
Sporadic CRC	15/28 **(53.6%)**	12/28 **(42.9%)**	1/28 **(3.6%)**	0.136
BC1	44/64 **(68.8%)**	15/64 **(23.4%)**	5/64 **(7.8%)**	
BC2	20/28 **(71.4%)**	7/28 **(25%)**	1/28 **(3.6%)**	
BC3	20/27 **(74.1%)**	5/27 **(18.5%)**	2/27 **(7.4%)**	
BC4	12/28 **(42.9%)**	15/28 **(53.6%)**	1/28 **(3.6%)**	
BC5	14/28 **(50%)**	12/28 **(42.9%)**	2/28 **(7.1%)**	

Table 3.4- Representation of the frequencies of each genotype for the rs3814570 polymorphism in tumors from patients fulfilling the BC, grouped in BC1-3 and BC4-5.

BC	**rs3814570**			**p value**
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	
BC1-3	83/120 **(69.2%)**	29/120 **(24.2%)**	8/120 **(6.7%)**	0.005
BC4-5	27/57 **(47.4%)**	27/57 **(47.4%)**	3/57 **(5.3%)**	

3.1.3 Association of *TCF7L2* polymorphisms in the promoter region to increased risk for developing CRC at young age and synchronous/metachronous CRC

To evaluate if this loss of heterozygosity also occurs at the germline level, the next step in this work was to analyze the rs3814570 genotype frequencies in DNA from peripheral blood samples of BC patients. Table 3.5 shows the frequencies of each genotype, stratified by BC, for the peripheral blood DNA.

When comparing the frequencies in tumor samples (Table 3.3) with the frequencies in blood samples (Table 3.5) there are some differences between the two. In the case of BC2, the genotype frequencies obtained in blood DNA were similar to those obtained in tumor samples [15/23 (62.2%),

6/23 (26.1%) and 2/23 (8.7%) *vs*. 20/28 (71.4%), 7/28 (25%) and 1/28 (3.6%) for the C/C, C/T and T/T genotypes, respectively]. However, in the case of BC1, there was a reduction of the frequency of the C/C genotype and an increase of the heterozygous frequency [52/96 (54.2%), 35/96 (36.5%) and 9/96 (9.4%) *vs.* 44/64 (68.8%), 15/64 (23.4%) and 5/64 (7.8%) for the C/C, C/T and T/T genotypes and for the blood and the tumor DNA samples, respectively. For BC3, this effect was even more pronounced and the genotype frequencies were similar to those observed in sporadic CRC and in healthy individuals. Indeed, for BC3, the frequencies in blood differed from the tumor tissue frequencies: 6/12 (50%) *vs.*20/27 (74.1%) and 6/12 (50%) *vs.*5/27 (18.5%), for the C/C (p=0.141) and heterozygous genotypes (p=0.043), respectively. The small cohort of DNA peripheral blood samples analyzed may contribute for these differences, although it appears that this loss of heterozygosity is less likely to occur at the germline level for BC3. The genotype frequencies observed in blood DNA from BC4 and BC5 were very similar to those observed in tumor DNA (Table 3.5).

Table 3.5- Representation of the frequencies of each genotype for the rs3814570polymorphism in DNA samples from peripheral blood from patients fulfilling the BC, stratified by criteria.

BC	**rs3814570**			**p value**
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	
BC1	52/96 **(54.2%)**	35/96 **(36.5%)**	9/96 **(9.4%)**	0.540
BC2	15/23 **(62.2%)**	6/23 **(26.1%)**	2/23**(8.7%)**	
BC3	6/12 **(50%)**	6/12 **(50%)**	0/12 **(0%)**	
BC4	8/19 **(42.1%)**	10/19 **(52.6%)**	1/19 **(5.3%)**	
BC5	17/42 **(40.5%)**	19/42 **(45.2%)**	6/42 **(14.3%)**	

Table 3.6- Representation of the frequencies of each genotype for the rs3814570polymorphism in DNA samples from peripheral blood from patients fulfilling the BC, stratified by criteria and grouped by similarity of the previous results: BC1; BC2 and BC3-5.

BC	**rs3814570**			**p value vs. CB3-5**
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	
BC1	52/96 **(54.2%)**	35/96 **(36.5%)**	9/96 **(9.4%)**	0.132
BC2	15/24 **(62.5%)**	7/24 **(29.2%)**	2/24**(8.3%)**	0.004
BC3-5	31/73 **(42.5%)**	35/73 **(47.9%)**	7/73 **(9.6%)**	-

In table 3.6, the BC that showed similar frequencies in blood (BC3-BC5) when compared to sporadic CRC and healthy individuals were grouped. Although in blood samples the genotype frequencies differed between BC1 and BC3-5, with BC1 presenting a lower frequency of heterozygosity (p=0.132), only the genotype frequencies obtained for BC2 differed significantly from BC3-5 (p=0,004) and consequently from the observed in the blood samples from healthy individuals.

The reduction of the frequency of homozygosity for the normal allele genotype (C/C), observed in blood DNA samples from BC1-3 and the increase of the frequency of the heterozygous genotype, led us to formulate a new hypothesis: Would this reduction of the homozygous frequency be particular of the tumor or it may occur already in the normal mucosa?

At this point, the frequencies of the genotypes for the rs3814570 polymorphism were analyzed in the DNA samples from normal mucosa, mostly the paired tissue of the tumors analyzed, i.e. belonging to the same individual. Table 3.7 shows the frequencies of each genotype for the rs3814570, in the normal mucosa samples, stratified by BC. The frequencies observed in normal mucosa were very similar to the frequencies observed in tumor samples, where BC1, BC2 and BC3 presented a higher frequency of homozygosity for the normal allele genotype (C/C) and a lower frequency of the heterozygous genotype (C/T). When the criteria were grouped base on their similarity, BC1-3 differed significantly from BC4-5 (p=0.021) (Table 3.8) and consequently from sporadic CRC and from healthy individuals.

Table 3.7- Representation of the frequencies of each genotype for the rs3814570polymorphism in DNA samples from paired normal mucosa from patients fulfilling the BC, stratified by criteria.

Sporadic CRC/BC	**rs3814570**			**p value**
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	
Sporadic CRC	8/16 **(50%)**	8/16 **(50%)**	0/16 **(0%)**	0.187
BC1	41/55 **(74.5%)**	10/55 **(18.2%)**	4/55 **(7.3%)**	
BC2	14/22 **(63.6%)**	7/22 **(31.8%)**	1/22 **(4.5%)**	
BC3	20/25 **(80%)**	4/25 **(16%)**	1/25 **(4%)**	
BC4	4/8 **(50%)**	4/8 **(50%)**	0/8 **(0%)**	
BC5	8/14 **(57.1%)**	6/14 **(42.9%)**	0/14 **(0%)**	

Table 3.8- Representation of the frequencies of each genotype for the rs3814570polymorphism in DNA samples from paired normal mucosa from patients fulfilling the BC, stratified by criteria and grouped by similarity of the previous results: BC1; BC2 and BC3-5.

Sporadic CRC/BC	**rs3814570**			**p value**
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	
Sporadic CRC	8/16 **(50%)**	8/16 **(50%)**	0/16 **(0%)**	0.021
BC1-3	75/102 **(73.5%)**	21/102 **(20.6%)**	6/102 **(5.9%)**	
BC4-5	12/23 **(52.2%)**	11/23 **(47.8%)**	0/23 **(0%)**	

Taking into consideration the similarity between the genotype frequencies observed in tumor and in normal mucosa and that the genotype frequencies of homozygosity in the normal mucosa, in the groups BC1-3, were higher than the same frequencies in blood samples. The following hypothesis was formulated: Could the genotype for the rs3814570 polymorphism change, in some BC patients (in particular BC1-3), from heterozygous in the blood DNA to homozygous in the tumor?

To address this question, paired samples of tumor, normal mucosa and the respective blood samples of BC individuals that meet the BC (mainly BC1-3) were analysed. Only cases that in the blood presented the heterozygous genotype were analyzed in order to evaluate if they had lost the heterozygosity in the tissue. The frequencies with which these changes occurred are represented in table 3.9 and it were depicted the changes from normal mucosa/peripheral blood samples to tumor samples and from peripheral blood samples to normal mucosa samples, classified according the approximate reduction in the contribution of one of the two alleles. In the table 3.10 the results for

each paired samples that presented changes among the three (blood, normal mucosa and tumor) were detailed. Overall there were more changes registered from normal mucosa/peripheral blood samples to tumor samples and in this case it is observed that BC1 was the BC that showed more changes: 12.5% for slight allelic disequilibrium, 20.8% for 30-40% of allelic disequilibrium and 29.2% for less than 30% of allelic disequilibrium to full LOH, although BC1 was the group more represented when compared to the remaining groups. Nevertheless, the category less than 30% of allelic disequilibrium to full LOH was the one that had more impact in both situations, presenting the higher frequencies: BC1 29.2% versus 27.8%; BC2 10% versus 25%; BC3 37.5% versus 28.6%, for the normal mucosa/peripheral blood samples to tumor samples and peripheral blood samples to normal mucosa samples, respectively. Regarding the BC4, a very small group of individuals was studied, not sufficient to take conclusions about its frequencies. Although the analysis of more cases comparing the respective tissues and blood would help to have more robust frequencies, the hypothesis that some cases, mainly from BC1, BC2 and BC3, change their genotype form heterozygous in the peripheral blood to homozygous in the normal mucosa, besides the already expected from blood to tumor, was confirmed.

The LOH in peripheral blood or in the tissue could be associated with a phenomenon called uniparental disomy (UPD) or acquired UPD, respectively, caused by the inheritance of two copies of maternal or parental chromosomes as a result of a meiotic error (Engel, 1980), inducing LOH but no copy number alterations (Andersen et al., 2007). The mechanism underlying this event is still a subject of research however it is thought to be related with mitotic recombination, a non-disjunction phenomenon or deletion and re-duplication events (Andersen et al., 2007).

UPD has been described in several malignancies, including solid tumors, and was suggested as a mechanism that would alter the expression of genes involved in carcinogenesis (Torabi et al., 2015). First studies described that UPD was an alternative mechanism that induced breast cancer, basal cell carcinoma, bronchial cancer and leukemia in which the tumor suppressor genes were inactivated by the deletion of the wildtype allele, followed by a gain of the mutated allele (Melcher et al., 2007). This phenomenon is confirmed when chromosomal regions are heterozygous in the normal mucosa and homozygous is the corresponding carcinoma genotype while the copy number in both tissues is identical (Melcher et al., 2007). In ovarian cancer, acquired UPD (aUPD) was described during tumor initiation and progression, also known as copy neutral LOH. This is an usual event on cancer, which occur due to chromosome loss followed by duplication of the other chromosome leading to whole chromosome aUPD or a somatic recombination that leads to a segmental aUPD, but in both cases copy number does not change (Tuna et al., 2015). In aUPD regions, homozygous mutated genes that have a role in initiation and progression of cancer may be associated with tumor type or subtype, risk of disease transformation and patients' survival time (Tuna et al., 2010). For example, MAP (MUTYH associated polyposis) carcinomas display a higher frequency of somatic UPD, implying an inherited or acquired predisposition to UPD, with presence of fragile sites prone to recombination or chromosomal instability, but on the other hand, sporadic CRC presents physical loss of chromosomal material so UPD is less frequent (Makishima and Maciejewski, 2011). Also, MSI has been shown to be associated with increased frequency of UPD (Makishima and Maciejewski, 2011). However in this

study this association between the allelic changes and MSI was not observed (p=1.00) (Table 3.11). A study performed by Keyvan Torabi and collaborators showed that *TCF7L2* gene has 16.7% of UPD/UPP (Uniparental polysomy) and 6.7% of copy number losses, indicating a prevalence of UPD over genomic loss (Torabi et al., 2015). Nevertheless, it has to be confirmed in which cases there are copy-neutral changes and in which cases there is copy-number loss/gain.

Table 3.9- Representation of the frequencies of BC1-4 cases that presented changes from normal mucosa/ peripheral blood samples to tumor samples and from peripheral blood samples to tumor samples. Each group was stratified in cases with no alterations, with a slight allelic disequilibrium, with 30-40% of allelic disequilibrium and with less than 30% of allelic disequilibrium to complete LOH.

Type of sample		**No alteration**	**Slight allelic disequilibrium**	**30-40% of Allelic disequilibrium**	**Less than 30% of Allelic disequilibrium**
Normal mucosa/Peripheral blood to tumor	**BC1**	8/24 **(33.3%)**	3/24 **(12.5%)**	5/24 **(20.8%)**	7/24 **(29.2%)**
	BC2	3/10 **(30%)**	5/10 **(50%)**	0/10 **(0%)**	1/10 **(10%)**
	BC3	3/8 **(37.5%)**	1/8 **(12.5%)**	1/8 **(12.5%)**	3/8 **(37.5%)**
	BC4	3/5 **(60%)**	0/5 **(0%)**	0/5 **(0%)**	2/5 **(40%)**
Peripheral blood to Normal mucosa	**BC1**	7/18 **(38.9%)**	1/18 **(5.6%)**	5/18 **(27.8%)**	5/18 **(27.8%)**
	BC2	4/8 **(50%)**	2/8 **(25%)**	0/8 **(0%)**	2/8 **(25%)**
	BC3	4/7 **(57.1%)**	1/7 **(14.3)**	0/7 **(0%)**	2/7 **(28.6%)**
	BC4	2/4 **(50%)**	0/4 **(0%)**	2/4 **(50%)**	0/4 **(0%)**

Table 3.10- Representation of changes in the allele contribution in the cases that presented alterations in the genotype among tumor, normal mucosa and peripheral blood samples. The cases were organized by individual (Tumor/normal mucosa/peripheral blood), stratified by BC. The percentages represent the allelic disequilibrium.

Sample (Tumor/NM/Blood)	**BC**	**Tumor**	**Normal mucosa**	**Peripheral blood**
Cas4089/-/Cas3912s	1	C/C	-	C/T
Cas3889/3890/Cas3890s		T/T	T/T	C/T
Cas3835/3836/Cas3836s		C/C	C/C	C/T
Cas3647/3648/L2336		C/t (70-75%; 25-30%)	C/T	C/T
Cas3627/3628/L2170		T/T	c/T (20-25%; 75-80%)	C/T
Cas3591/3592/L2167		C/t (60-70%; 30-40%)	c/T (30-40%; 60-70%)	C/T
Cas3425/3426/Cas3426s		T/T	C/T	C/T
Cas3319/-/L1864		C/C	-	C/T
Cas3275/3276/L2133		C/T	C/C	C/T
Cas2967/2968/Cas2968s		-	C/C	C/T
Cas2903/2904/L1769		C/T	c/T (30-40%; 60-70%)	C/T
Cas2869/2870/Cas2870s		C/C	C/T	C/T
Cas4559/4560/Cas4560s	2	C/C	C/C	C/T
Cas3485/-/-		C/t (75-80%; 20-25%)	-	-
Cas3371/3372/Cas3372s		C*/T	C/C	C/T
Cas3793/-/Cas3794s	3	T/T	-	C/T
Cas3557/3558/L2297		C/C	C/C	C/C

Cas3481/-/-		c/T (15-25%; 75-85%)	-	-
Cas3271/3272/Cas3272s		c/T (30-40%; 60-70%)	T/T	C/T
Cas2361/2362/L1702		c/T (20-30%; 70-80%)	C/t (90-95%; 5-10%)	C/t (60-70%; 30-40%)
Cas3471/3472/L2331	4	C/t (70-85%; 25-30%)	C/t (60-70%; 30-40%)	C/T
Cas2585/2586/L2068		c/T (20-25%; 80-85%)	C/t (60-70%; 30-40%)	C/T

Legend: - no information/sample nonexistent; * -slight allelic disequilibrium; lower case- accentuated allelic disequilibrium; NM- normal mucosa.

Table 3.11- Representation of the frequencies of the genotypes for rs3814570 polymorphism in DNA tumor samples from BC individuals stratified by MSS, MSI and MSI-L.

Microsatellite instability status	**rs38145570**			**p value**
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	
MSS	72/109 **(66.1%)**	30/109 **(27.5%)**	7/109 **(6.4%)**	1.00
MSI	18/26 **(69.2%)**	7/26 **(26.9%)**	1/26 **(3.7%)**	
MSI-L	2/2 **(100%)**	0/2 **(0%)**	0/2 **(0%)**	

Legend: MSS- microsatellite stability; MSI- Microsatellite instability; MSI-L- Microsatellite instability low

3.1.4 Association of *TCF7L2* polymorphisms in the promoter region to synchronous CRC metastases

As previously demonstrated by the Gastroenterology group at UIPM-IPOLFG, loss of heterozygosity of the *TCF7L2* gene and deletions in the promoter and proximal exons have been associated to the development of synchronous metastases. In order to improve this data and evaluate if this association was also observed for *TCF7L2* polymorphisms in the promoter region information regarding the development of synchronous metastases was also evaluated in this study. Accordingly, it was also possible to observe an association between the homozygosity for the normal allele of the polymorphism (C/C) and synchronous metastases. In M1 tumor cases the frequency of homozygosity for the normal allele (C/C) was significantly higher than the frequency of heterozygosity: 23/27 (85.2%) vs. 3/27(11.1%) (p=0.03) (Table 3.12). In the case of normal mucosa, it was possible to observe a tendency for the same association (p=0.274) (Table 3.13). This tendency appears to occur only very slightly in blood samples (p=0.479) (Table 3.14) although the cohort should be enlarged to make these considerations.

Table 3.12- Representation of the frequency of the genotypes for thers3814570 polymorphism in a group of DNA tumor samples from BC and BC1-3 individuals, respectively, stratified by the presence (M1) or absence (M0) of synchronous metastases.

Metastasis	**rs3814570**			**rs3814570: for BC1-3**		
	Normal (C/C)	**Het. (C/T)**	**Hom. (T/T)**	**Normal (C/C)**	**Het. (C/T)**	**Hom. (T/T)**
M0	44/76 **(57.9%)**	27/76 **(35.5%)**	5/76 **(6.6%)**	27/42 **(64.3%)**	11/42 **(26.2%)**	4/42 **(9.5%)**
M1	23/27 **(85.2%)**	3/27 **(11.1%)**	1/27 **(3.7%)**	21/24 **(87.5%)**	2/24 **(8.3%)**	1/24 **(4.2%)**
p value	0.030			0.158		

Table 3.13- Representation of the frequency of the genotypes for thers3814570 polymorphism in a group of DNA normal mucosa samples from BC and BC1-3 individuals, respectively, stratified by the presence (M1) or absence (M0) of synchronous metastases.

Metastasis	rs3814570			p value
	Normal (C/C)	Het. (C/T)	Hom. (T/T)	
M0	41/64 **(64.1%)**	20/64 **(31.3%)**	3/64 **(4.7%)**	0.274
M1	17/20 **(85%)**	3/20 **(15%)**	0/20 **(0%)**	

Table 3.14- Representation of the frequency of the genotypes for thers3814570 polymorphism in a group of DNA blood samples from BC and BC1-3 individuals, respectively, stratified by the presence (M1) or absence (M0) of synchronous metastases.

Metastasis	rs3814570			rs3814570: for BC1-2		
	Normal (C/C)	Het. (C/T)	Hom. (T/T)	Normal (C/C)	Het. (C/T)	Hom. (T/T)
M0	21/41 **(52.2%)**	16/41 **(39%)**	4/41 **(9.8%)**	13/22 **(59.1%)**	8/22 **(36.4%)**	1/22 **(4.5%)**
M1	12/21 **(57.2%)**	9/21 **(42.9%)**	0/21 **(0%)**	8/11 **(72.7%)**	3/11 **(27.3%)**	0/11 **(0%)**
p value	0.479			0.802		

3.2 Comparison between the genotype for rs3814570 and rs10885395 and copy-number variation and gene expression in tumor, normal mucosa and peripheral blood samples from selected individuals diagnosed with CRC that meet the Bethesda criteria

In this part of the study, 16 individuals that meet one of the BC were selected for concomitant evaluation and characterization of the genotype for rs3814570 and rs10885395 polymorphisms, copy-number variation and gene expression in tumor, normal mucosa and peripheral blood samples from the same individual. This work is still in progress however some the results obtained are documented in Table 3.15. Copy-number analysis is being completed with Multiplex ligation-dependent probe amplification (MLPA) analysis performed by the Gastroenterology group-UIPM.

One of the cases (Cas2635/Cas2636/L1848) presented an allelic disequilibrium in the blood DNA (L1848) and in the tumor tissue (Cas2635). Also the contribution of the C and T alleles of the polymorphism was inverted between the tumor and blood, which means that the higher peak represented in the tumor DNA sequencing was the lower in the blood DNA sequencing (Figure 3.1). The cases with allelic disequilibrium showed a contribution of each allele that was not 50:50. This may be explained by the loss of heterozygosity and understood when these results were compared with the MLPA information. The cases that showed allelic disequilibrium were: Cas4168; Cas3431; the pair Cas4241/L2738 and the pair Cas3717/A1976. Those disequilibria are represented in the figure 3.4.

Although data is preliminary, in Cas3431 and Cas3317 the allelic disequilibrium observed in the tumor may be explained by the copy-number gain observed in these tumors by MLPA. The cases that were homozygous for the polymorphism, even though some presented copy-number variation, detected by MLPA, they were not informative regarding allelic disequilibrium.

Table 3.15- Analysis of DNA from tumor, normal mucosa and peripheral blood samples from 16 individuals diagnosed with CRC that meet one of the BC. The analysis consists in: the genotype result for the rs38145570 and rs10885395 polymorphisms of the *TCF7L2* gene; MLPA results from the DNA tumor and normal mucosa tissues; and copy number variation and gene expression results from the DNA blood samples.

Sample	rs38145570	rs10885395	Deletion Promoter region/exon 1	Del Intron2	Gain	Copy number	Gene expression
Cas2635	C/Ct	C/t	-	-	-	-	-
Cas2636	T/T	C*/T	-	-	-	-	-
L1848	C/T*	C/t	-	-	-	”Normal”	Reduction
Cas3411	C/C	C/C	1	0	0	-	-
Cas3412	C/C	C/C	0	0	1	-	-
L2030	C/C	C/C	0	0	0	”Normal”	Reduction
Cas4231	C/T	C/T	0	0	0	-	-
Cas4232	C/T	C/T	0	0	0	-	-
Cas3339	C/C	C/C	1	1	0	-	-
Cas3340	C/C	C/C	0	0	1	-	-
A2008	C/C	C/C	0	0	0	”Normal”	”Normal”
Cas3505	C/C	C/C	0	0	0	-	-
Cas3506	C/C	C/C	0	0	0	-	-
L2254	C/C	C/C	0	0	0	Gain	Reduction
Cas4095	C/T	C/T	0	0	0	-	-
Cas4096	C/T	C/T	0	0	0	-	-
Cas4169	C/C	C/C	0	0	0	-	-
Cas4170	C/C	C/C	-0	-0	0	-	-
Cas4171	C/C	C/C	0	0	0	-	-
Cas4172	C/C	C/C	0	0	0	-	-
Cas4167	C/T	C/T	0	0	1	-	-
CAs4168	C/T	C/t	1	1	0	-	-
L2494	C/T	C/T	0	0	0	Gain	Reduction
Cas3287	C/C	C/C	-	-	-	-	-
Cas3288	C/C	C/C	-	-	-	-	-
Cas3431	c/T*	C/t	-	-	1	-	-
Cas3553	T/T	T/T	-	-	1	-	-
Cas3554	T/T	T/T	-	-	-	-	-
Cas3554s	T/T	T/T	-	-	-	-	-
Cas4203	C/C	C/C	0	0	1	-	-
Cas4221	C/T	C/t	-	-	-	-	-
L2738	C/T	C/t	-	-	-	Loss	-
Cas3317	c/T*	C/t	-	-	1	-	-
A1976	C/T	C/t	-	-	-	Loss**	Reduction
Cas3543	C/C	C/C	-	-	1	-	-
A1649	C/C	C/C	-	-	-	-	-

Legend: 0- no; 1- yes; - without information; *- polymorphism opposite to the tumor; **- copy number loss only in the region intron1/exon 2; black- homozygous for the normal allele (C); grey- heterozygous for the polymorphism; red- homozygous for the polymorphism; lowercase- lower contribution of the allele -allelic disequilibrium;

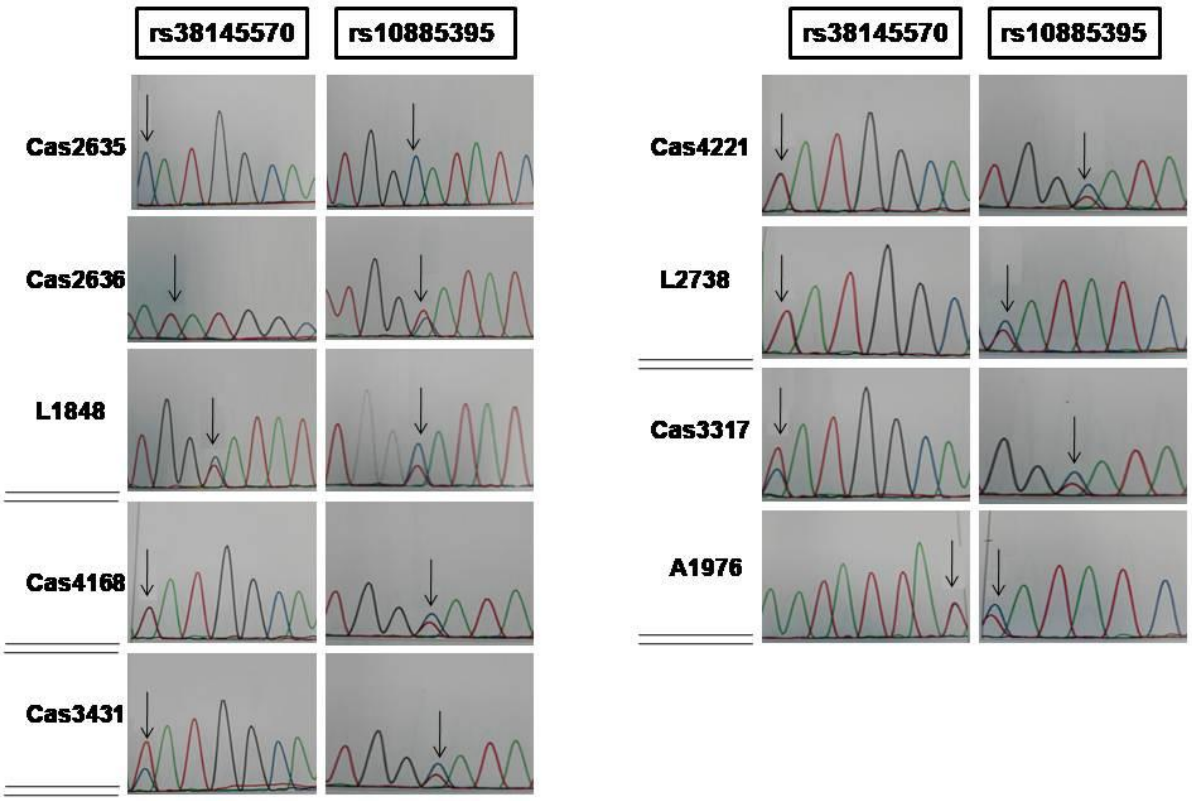


Figure 3.1- Representative figure of the analysis of rs38145570 and rs10885395 polymorphisms of the *TCF7L2* gene in DNA samples extracted from tumor (Cas2635, Cas3431, Cas4221, Cas3317), from normal mucosa (Cas2636, Cas4168) and from peripheral blood (L1848, L2738, A1976) from CB individuals. For each polymorphism it is represented the sequence obtained with the forward primer. The arrow represents the alteration corresponding to the polymorphism where the allelic disequilibrium occurs. The double line separates the cases by individual.

3.3 Copy number analysis of the *TCF7L2* gene in DNA samples from peripheral blood from BC individuals

It was hypothesized that could exist a relation between the rs3814570 polymorphism genotypes and the copy number variation in the *TCF7L2* gene, and that, eventually, it could influence the risk of CRC development.

In this study, analysis of copy number in the *TCF7L2* gene was performed for 28 DNA samples from peripheral blood in individuals diagnosed with CRC that meet one of the BC (Figure 3.1). Copy number variation was analyzed for five different copy-number assays localized in the following regions of the *TCF7L2* gene: the untranslated region of exon 1; the exon 1; the border region intron 1/exon 2; intron 4 (close to rs7901695); and finally the border region intron 16/exon 17 (Figure 2.1). The first three regions were chosen due to their proximity to the promoter region of the gene and to rs3814570 and because of the probability of some change in this region could influence the transcription of the *TCF7L2* gene. The intron4/proximity to rs7901695 and intron16/exon17 assays were chosen to cover a central and 3´primer region of the gene and because of the localization of the main alternative splicing sites of the gene (exon 4 and exons 13-17) (Cingilli Vural, 2017). Copy number evaluation was calculated relatively to reference samples from healthy individuals.

Figure 3.2 represents the results obtained for the copy number variation analysis for a series of blood DNA samples from BC individuals, according with the genotypes obtained for the rs3814570 polymorphism of the *TCF7L2* gene. A possible relation between the result from the copy number variation and the heterozygous or homozygous status for the polymorphism was evaluated. The following outcomes were analyzed: if the copy number variation was around 1 (between 0.75 and 1.25) the region of the gene did not suffer any copy number variation; if the result was higher than 1.25 (around 1.5 in the case of blood samples), that corresponded to a gain of a genomic region; and if the copy number variation was lower than 0.75 (around 0.5 in the case of blood samples) the region had a deletion. Within each sample the results were concordant among the first three assays (untranslated region of exon 1; exon 1 and intron 1/exon 2) and tend to have similar results for each sample: neutral, gain or deletion. The same occurred for the assays intron 4/proximity to rs7901695 and intron 16/exon 17, that presented a normal copy-number (near 1) for all cases. However, in a global view, the results did not seem to follow a pattern, in other words appeared to be no relation between the copy number values and the different polymorphism genotypes. For the homozygous for the normal allele and for the heterozygous genotype cases there were samples without copy-number alteration, gains and losses, with no tendencies to one result.

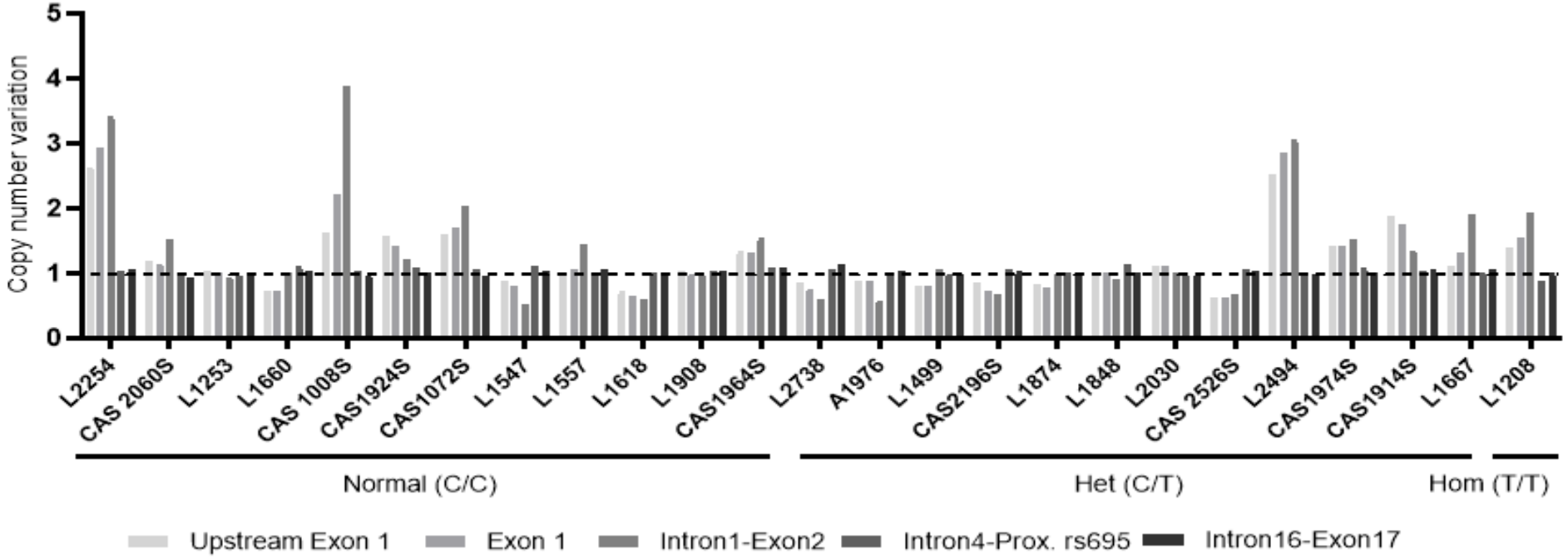


Figure 3.2- Copy number variation analysis of the *TCF7L2* gene in DNA samples from peripheral blood from individuals diagnosed with CRC and that meet one of the Bethesda criteria (BC individuals). The analysis was performed for five different copy number assays designed for five different regions of the *TCF7L2* gene: upstream of exon1; exon 1; intron 1 to exon 2; intron 4/proximal region of the rs7901695; and intron16 to exon 17.

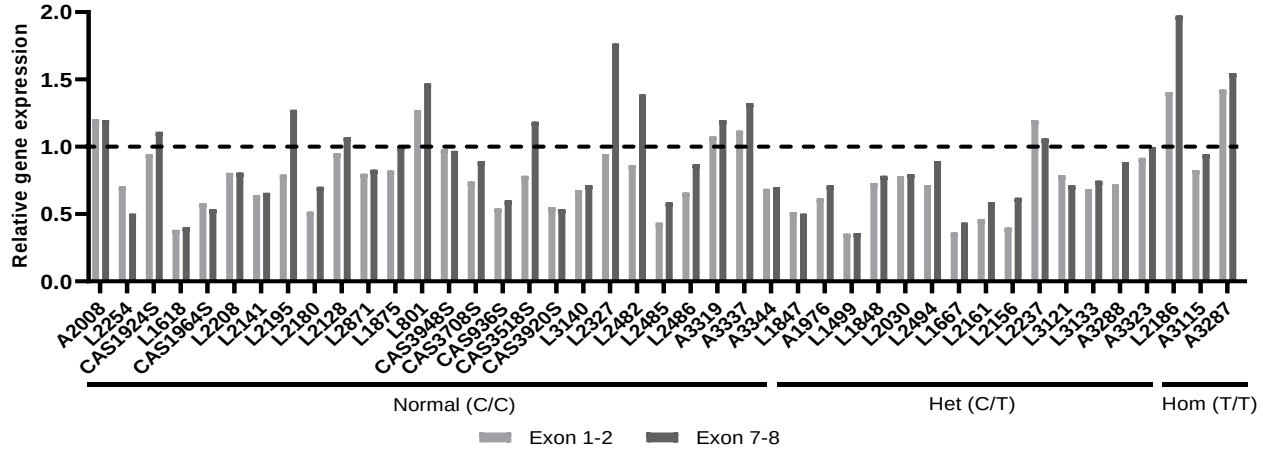
3.4 Analysis of *TCF7L2* relative gene expression in RNA samples from individuals diagnosed with CRC that meet one of the Bethesda criteria

The next step, was to observe if there was a relation between the different genotypes for the rs3814570 polymorphism and the gene expression of the *TCF7L2* gene, and ultimately if the polymorphisms could contribute for increased risk of developing CRC.

In this study, relative gene expression analysis was performed for RNA samples from 43 individuals diagnosed with CRC that meet one of the BC (BC individuals). cDNA was synthesized from the RNA samples from peripheral blood. In order to evaluate if the genotype obtained for the rs3814570 polymorphism, localized in the promoter region, has an influence in the expression of *TCF7L2* gene, gene expression assays targeting exon 1/exon 2 and exon 7/exon8 borders, respectively, have been analyzed. Exons 7/exon 8 region was analyzed because it is a region that does not suffer alternative splicing and thus it is represented in almost every transcript, representing the global expression of the *TCF7L2* gene.

The relative gene expression analysis was obtained using as reference 6 RNA samples from healthy individuals of the Portuguese population (representing the different genotype forms of the polymorphism (Appendix XI, Table XI.1). In the table XII.2, from Appendix XII, and in the figure 3.2 are represented the samples that were analyzed, the respective result for the exons 1-2 and exons 7-8, and also the information about the genotype for the rs3814570 polymorphism.

Figure 3.3- Graphic representation of the*TCF7L2* relative gene expression analysis in RNA samples from individuals diagnosed with CRC that meet one of the Bethesda criteria, for the exons 1-2 and exons 7-8 regions, sorted by genotype for the rs3814570 polymorphism.



For each genotype (C/C, C/T and T/T) there were cases of arbitrary ”normal”, increased and reduced gene expression relative to reference samples, and in almost every case the result for the exons 1-2 matched the result for the exons 7-8, except in five cases: L2195, Cas3518s, L2327, L2482 and L2186 (Figure 3.3). Then, the results were grouped (Table 3.16 and Figure 3.4) in order to evaluate the variation of gene expression within and among each genotype. Comparing the results for both exons 1-2 and 7-8 assays, the samples that were homozygous for the normal allele (C/C) presented an overall tendency to have a higher relative gene expression than the heterozygous cases: 0.7875 *vs*. 0.6599 (mean values for exons 1-2) and 0.934 *vs*. 0.721 (mean values for exons 7-8), respectively.

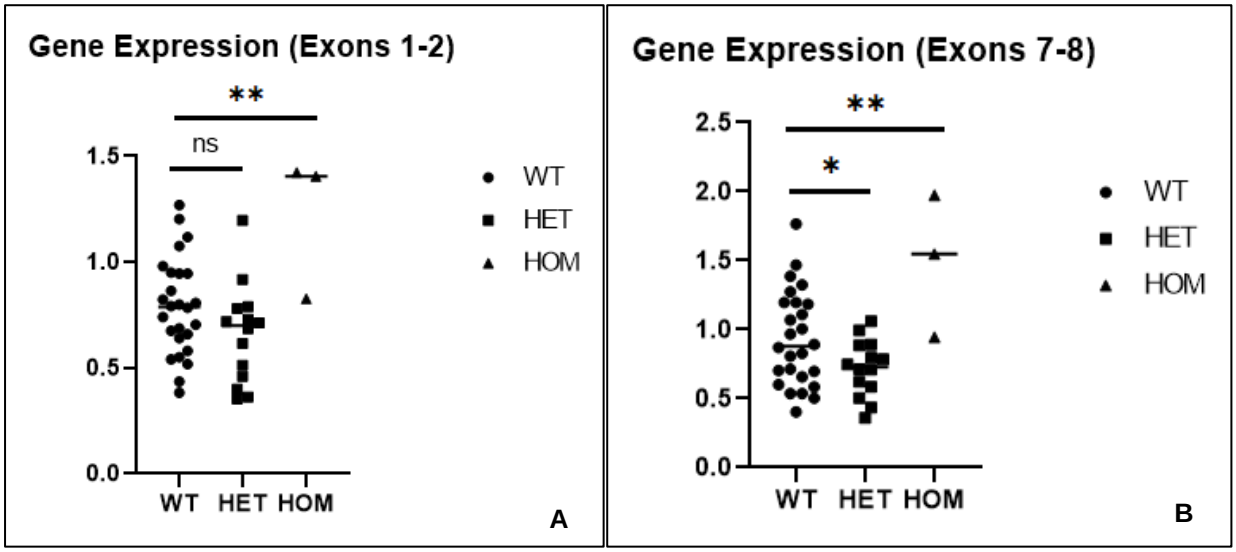


Figure 3.4- Graphic representation of the relative gene expression analysis in RNA samples from individuals diagnosed with CRC that meet one of the Bethesda criteria. A relative gene expression results for the border region exon1- exon 2; B, relative gene expression results for the border region exon7-exon 8.

Table 3.16- Frequency of cases that have a superior value than the mean of the expression values obtained for each genotype, for rs3814570 polymorphism, for exons 1-2 and exons 7-8, respectively. The mean values for exons 1-2 and exons 7-8 assays were 0.776 and 0.903, respectively.

Assays	Normal (C/C)	Het (C/T)	Hom (T/T)	P value
Exons 1-2	14/26 **(53.85%)**	4/14 **(28.57%)**	3/3 **(100%)**	0.058
Exons 7-8	12/26 **(46.15%)**	2/14 **(14.29%)**	3/3 **(100%)**	0.012

In order to better understand the variation of the expression among different genotypes, the mean value obtained for each assay and for each genotype was compared with the expression values obtained for that specific assay and genotype. Accordingly, Table 3.16 represents the frequency of cases, of each genotype, that presented a gene expression value superior than the mean. Once again, the homozygous cases for the normal allele (C/C) presented a higher frequency of cases that have a superior relative gene expression value when compared with the heterozygous cases: 14/26 (53.85%) *vs*. 4/14 (28.57%) for exons 1-2 (p=0.853) and 12/26 (46.15%) *vs*.2/14 (14.29%) for exons 7-8 (p=0.003), respectively. These differences among the genotypes are reinforced by the scatter-plot

analysis represented in figure 3.4. As the exon 7-8 region reflect the total *TCF7L2* gene expression, including virtually all the coding transcripts, the significant difference between homozygous and heterozygous genotypes obtained for exons 7-8 assume particular relevance.

Although there is no statistical significance for the exons 1-2, it is important to note that there is a tendency for the homozygous for the normal allele genotype to have superior gene expression values than the heterozygous genotype, which could be improved with the study of a larger cohort of individuals.

In the point 3.1 of this study it was demonstrated that the genotype frequencies of the *TCF7L2* rs3814570 polymorphism, obtained for the individuals that fulfill BC1, BC2 and BC3 differed from the European, Iberian and Portuguese populations, not only at the tumor level but also in the normal mucosa and blood DNA. BC1, BC2 and BC3 presented an increased frequency of homozygosity for the normal allele (C/C) and a lower frequency of heterozygosity that may be associated with the superior relative gene expression of *TCF7L2* observed for the homozygous genotype. When there is a higher frequency of homozygosity for the normal allele of *TCF7L2* and its expression is increased, the gene loses the ability to function as a regulator of the WNT pathway resulting in an increased activation and chance of developing CRC, due to its role in cancer cell proliferation and metastasis development (Tang et al., 2008).

3.5 Mapping gains/deletions in the promoter region and proximal exons of the *TCF7L2* gene

The identification of gains and deletions at the promoter and proximal exons of the *TCF7L2* gene prompt us to map the start and end positions of these structural alterations. The first approach was to sequence a region encompassing part of the promoter region (including rs3814570) until intron 4 (fragment 1, Material and methods section 2.2), (~17000bp). It was not possible to amplify the 17Kbp fragment, possibly due to the large size of the fragment, and only various unspecific fragments (confirmed by sequencing) were obtained.

In alternative, due to the difficulty to amplify this fragment, a region encompassing part of the promoter until exon 3 (fragment 2, Experimental procedures section 2.2) (~4000bp) was amplified. However, the amplification of the 4Kbp fragment did not reveal differences among samples and there was any evidence of gains or losses. This may be explained by the fact that the amplified fragment possibly does not encompass the region deleted or gained as their limits may be localized upstream of the forward primer and/or after exon 3.

Taking this into account, a new set of primers was designed in order to amplify a larger region and maximize the probability to encompass the deletions/gains: one forward primer localized in the *TCF7L2* promoter, but 1580bp upstream of the TCF7L2 P0.2 primer, and two reverse primers, one in the beginning of intron 3 (TCF7L2 Int3.1) and a second in the last part of this intron (TCF7L2 Int3.2). Accordingly, two fragments were amplified (fragments 3 and 4, Experimental procedures section 2.2), the TCF7L2 Ups P0.2 forward primer with the TCF7L2 Int3.1 reverse primer, resulting in a fragment with 7542bp, and TCF7L2 Ups P0.2 forward primer with the TCF7L2 Int3.2, resulting in a fragment with 9711bp. Despite some differences that were observed between samples, it was not possible to

map the region of gain or deletion by sequencing. The fragments that were sequenced contained a small part of the promoter sequence but the rest of the sequence was not specific. The *TCF7L2* promoter has high homology with other sequences within the gene and this may be one of the causes of the unspecific fragments.

Other different combinations using the primers previously mentioned were also tested: TCF7L2 Ups P0.2 combined with TCF7L2 exon 3 (5663bp), combined with TCF7L2 Int3.1 (5926bp), or with TCF7L2 Int3.2 (8095bp) as reverse primers, respectively (Figure 3.5). Although this analysis was inconclusive the last combination with the TCF7L2 Int3.2 primer appeared to be the most promising but the complete sequence of this region was not possible to obtain. Indeed, both Cas2526s and Cas2196s (Figure 3.5) presented copy number deletions and showed a similar pattern, differing slightly from L1847 and L2254, the first not presenting copy number alteration and the latter presenting gain. However, this amplification needs further optimization to eliminate the unspecific fragments.

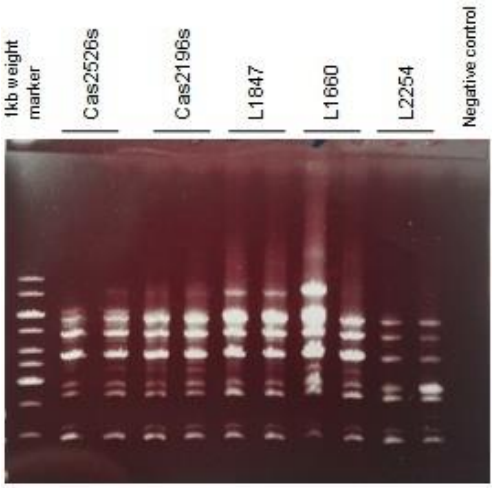


Figure 3.5- Representative image of the amplification of *TCF7L2* gene fragment encompassing part of the promoter region until intron 3 (9711bp forward primer and TCF7L2 Int3.2 reverse primer) after agarose gel electrophoresis. Molecular weight marker - 1Kb ladder. Each sample is represented in duplicate.

3.6 Potential treatments for metastatic CRC: analysis of the expression of target genes in response to treatment of the metastatic LoVo cell line with Gant61 in combination with sulforaphane

The last part of this study attempted to explore potential treatments for metastatic CRC that was the focus of this part of the project. These findings follow a previous work of the Gastroenterology group from UIPM-IPOLFG developed with the objective of identifying targeted treatments for different types of CRC while reducing the toxicity of the conventional treatments with a synergy with nutraceutics. In this context, potential nutraceutics that could have a positive impact in the CRC treatment, probably in combination with the conventional therapies were studied.

The present study was focused in the LoVo cell line that originated from left supraventricular CRC metastases, representative of the molecular subtype CMS1, where the MSI tumors are included, with high metastatic potential and poor prognostic after recurrence (Berg et al., 2017). It was chosen

this particular cell line in order to develop a target therapy for the metastatic CRC, an aggressive form of CRC due to its advanced stage. Previous results of the Gastroenterology group from UIPM-IPOLFG showed that different tumoral subtypes with distinct molecular and cellular characteristics present different therapeutic responses, for example the SW480 and HCT116 cell lines have 5-FU chemoresistance (Santos et al., 2018). Hence the need to investigate new personalized therapeutic strategies more efficient to each subtype.

LoVo cells were treated, during 24h, with three different conditions: treated with Gant61 (15µM); treated with a synergy of Gant61 (15µM) and sulforaphane (SFN) (12.5µM); and treated only with SFN (12.5µM). Gant61 is a GLI1 inhibitor and has already showed cytotoxicity activity in human models of colon cancer (Agyeman et al., 2014) while SFN is a nutraceutic, that belong to the ITC group, that has an antiproliferative effect with ability to impair metastatic CRC by influencing the cell migration, invasion and growth in anchorage-independent conditions (Pereira et al., 2017). These concentrations of each treatment were chosen based on the results of a previous study performed by the group of Gastroenterology from UIPM-IPOLFG (Santos et al., 2018) where they demonstrate that a dosage of 15µM of GANT61 inhibits cell migration in the wound-healing assays and the best combination of GANT61 and SFN was the dosage of 15µM and 12.5µM with an IC_{50} of 23.1µM (the previous results are shown in Appendix XIII). The cells were cultured with the different treatments, then the cells were collected and was extracted the RNA in order to synthesize cDNA. In the present study RT-qPCR analysis was performed evaluating the gene expression of the following genes: *CCNA2; VIM; CDH1; GLI1; LGR5; PROM1; CDKN1A; TCF;* Isoform 1-6 of *TCF; AXIN2; TGF-β; ZEB1; SNAIL; SOX9; CTNNB1 (β-catenin); CTCF; CD44*. These genes were chosen in order to study the EMT-associated markers (*VIM; CDH1; ZEB1; SNAIL;TGF-β*), the stemness markers (*LGR5; PROM1; PTCH1; CD44*), the Wnt/β-catenin signaling pathway (*TCF;* Isoform 1-6 of *TCF; AXINA2; SOX9*) and cell cycle markers (*CCNA2;CDKN1A*).

To investigate the anti-metastatic effect of the different treatments in LoVo cell line, the gene expression of the EMT-associated markers, *VIM* (coding for *Vimentin* and usually methylated in carcinomas),*CDH1* (coding for E-cadherin), *ZEB1* (an activator of EMT, induces tumorigenesis and metastization)*; SNAIL* (repressor of E-cadherin); and it was evaluated as well *TGF-β*. The results showed gene expression values for these genes around 1 or lower for the three treatments, meaning that the gene expression of the EMT-associated markers had a normal or downregulated expression (Figure 3.6). The downregulation of the EMT-associated markers would be a positive result because EMT is the transition of epithelial to a mesenchymal phenotype, the cells start losing their tight junctions, and their migratory and invasion capacities increases until the development of metastasis (Kalluri and Weinberg, 2009). A reduced expression of these genes after treatment suggests that the treatment may have impact in the metastases development. Between treatments the results were very similar (Table 3.17): 0.78 versus 0.53 for *VIM* gene; 1.39 versus 1.25 for *CDH1* gene; 1.22 versus 1.15 for *ZEB1* and 0.90 versus 0.94 for *SNAIL* gene, for Gant61 and Gant61+SFN treatment, respectively. These results are supported by a slight upregulation of the *TGF-β:* 1.19 versus 1.54, for Gant61 and Gant61+SFN treatment, respectively (Table 3.17). This suggests that the combination of the nutraceutic with Gant61 has the same effect as the Gant61 alone but may have benefits in terms

of reducing cytotoxicity to normal cells due to the lower dosages that are used. In general, the compounds tested did not seem to modulate the cell adhesion via E-cadherin once there was not a significant effect on the expression of *CDH1*, *ZEB1* and *SNAIL*. On the other hand, *VIM* was downregulated suggesting a reduced EMT, translated into a more epithelial phenotype.

To demonstrate the stemness effect after the proceeded treatments in the LoVo cell line, the gene expression of the stemness markers*, PTCH1* (a receptor for Hh pathway, it is involved in tumorigenesis), *LGR5* (functions as a biomarker of adult stem cells and is one of the target genes of the Wnt/β-catenin pathway), *PROM1* (coding for CD133) and *CD44* (a cell surface adhesion receptor) was performed. These markers may be related with the tumor recurrence commonly associated to the CSCs, which can prevail after chemotherapy and have contribution for the tumor relapse (Pereira et al., 2017). The expression of the *PTCH1* was similar to the control: 1.03 versus 0.92, for Gant61 and Gant61+SFN treatment, respectively. On the other hand, the expression of the remaining markers was slightly increased: 1.59 versus 1.78 for the *LGR5* gene and 1.41 versus 1.43 for the *PROM1* gene, for Gant61 and Gant61+SFN treatment, respectively (Table 3.17). The expression of CD44 was downregulated: 0.46 versus 0.42 for the *CD44* gene, for Gant61 and Gant61+SFN treatment, respectively (Table 3.17) .For these markers the SFN treatment did not have the same impact on the expression as the other two treatments: 0.46 versus 0.42 versus 1.04, for Gant61 treatment, Gant61+SFN and SFN treatment, respectively. This result indicates that the SFN treatment alone does not have the same effect (reduction of gene expression) as the other treatments in this cell line (Table 3.17). These results suggest that both treatments did not induce significant modulation of the expression of the stemness markers, even thought there is a slight increase, not significant, of the *LGR5* expression.

Then to investigate the Wnt/β-catenin signaling pathway, the genes that belong to the pathway, such as *TCF,* isoform 1-6 of *TCF, AXIN2, SOX9,* which is an inhibitor of the Wnt pathway and β-catenin, were evaluated. The results showed, after treatment, a tendency for a neutral to reduced gene expression of these genes: 1.14 versus 1.07 for the *TCF*, 1.27 versus 1.28 for the isoform 1-6 of the *TCF*, 0.78 versus 0.94 for the *AXIN2* gene, 1 versus 1.06 for the *SOX9* gene and 0.32 versus 0.31 for the β-catenin, for Gant61 and Gant61+SFN treatment, respectively (Table 3.17). Once again both treatments appear to have similar results in the Wnt/β-catenin signaling pathway. Also, the SFN treatment had a similar result when was tested the expression of *SOX9* gene (1.12) suggesting that the nutraceutic alone have the ability to keep this gene with a normal expression, not leading to the inhibition of the Wnt signaling pathway. On the other hand, the SFN solo treatment appears to not influence the expression of β-catenin (1.06) in the same way as the other two treatments.

Finally, in order to investigate cell division and proliferation the expression of the cell cycle markers, *CCNA2* (coding for *CyclinA2,* responsible for division), and *CDKN1A* (coding for p21 responsible for the cell cycle regulation). Treatment with Gant61 and Gant61+SFN significantly induced *CDKN1A* (p21) upregulation in LoVo cells: 4.17 versus 3.83, for Gant61 and Gant61+SFN treatment, respectively (Table 3.17). Although p21 overexpression usually suggests induction of the cell cycle arrest at G1/S phases, p21 has also been involved in the G2/M transition in response to

DNA damage after oxidative stress, assuming a negative regulatory role (Pereira et al., 2017). It is also been reported that the ITCs can cause oxidative stress, inducing DNA damage and promoting G2/M cell cycle arrest in several tumors, such as in colon cancer cells (Pereiraet al., 2017). Therefore, the upregulation of p21 is related to cell cycle arrest and possibly to the DNA damage in the tumor cells, caused by the treatment. The results also show a downregulation of the *CCNA2* gene in LoVo cells: 0.48 versus 0.53, for Gant61 and Gant61+SFN treatment, respectively (Table 3.17). This decrease in expression may be related with the *CyclinA2* gene decrease during the G2 phase of the cell cycle, phase in which the cell-cycle was arrested by the treatments in course.

In general, the Gant61 and the combination of Gant61+SFN treatments show very promising results at a gene expression level, suggesting that both treatments could reduce mesenchymal phenotype by downregulating the expression of the EMT-associated markers, could modulate Wnt signaling pathway and could induce cell cycle arrest of metastatic cells. Although the SFN treatment alone had some promising results, it is necessary to test other markers. Furthermore, the combination Gant61+SFN by using lower dosages may reduce cytotoxicity of normal cells.

Table 3.17- Representation of results of gene expression analysis after treatment of LoVo cell line with Gant61, the combination of Gant61 and sulforaphane (SFN) and SFN, during 24h. This analysis was performed by RT-qPCR, using LoVo cells treated with DMSO and ethanol as reference sample.

Genes	Treatments		
	GANT61	GANT61+SFN	SFN
CCNA2	0,48	0,53	-
CDKN1A	4,17	3,83	-
VIM	0,78	0,53	-
CDH1	1,39	1,25	-
TGFB	1,19	1,54	-
ZEB 1	1,22	1,15	-
SNAIL	0,9	0,94	-
LGR5	1,59	1,78	-
PROM1	1,41	1,43	-
CD44	0,46	0,42	1,04
TCF	1,14	1,07	-
TCF ISO1-6	1,27	1,28	-
AXINA 2	0,78	0,94	-
SOX 9	1	1,06	1,12
B-CAT	0,32	0,31	1,06
CTFC	0,79	0,9	1,31
GLI1	0,65	1,08	-
PTCH1	1,03	0,92	-

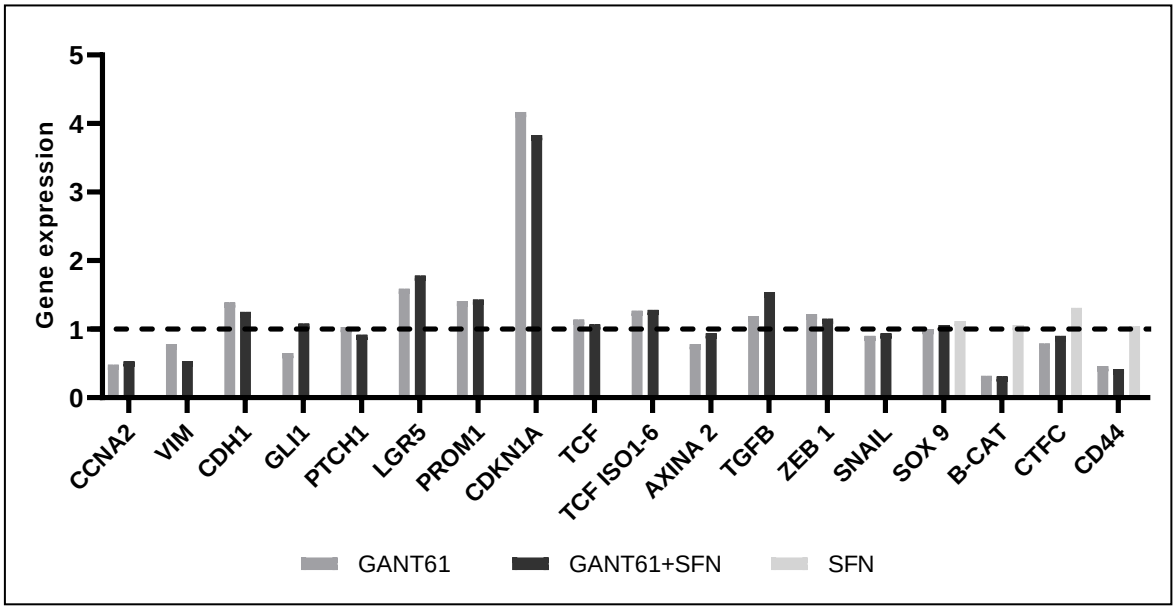


Figure 3.6- Graphic representation of the results of gene expression analysis after treatment of LoVo cell line with Gant61, the combination of Gant61 and sulforaphane (SFN) and SFN, during 24h.This analysis was performed by RT-qPCR, using LoVo cells treated with DMSO and ethanol as reference sample. The line traced in 1 indicates the reference control for gene expression.

4 Conclusion

The present Master thesis continues the extensive work performed by the group of Gastroenterology of UIPM-IPOLFG with the aim of identifying biomarkers of colorectal cancer risk and metastasis. The number of individuals diagnosed with early onset CRC has increased, being often characterized with advanced cancer with poor prognosis and distant metastasis. This reinforces the need of finding biomarkers that could facilitate early diagnosis and disease monitoring and of contributing for the development of new precision therapies.

Previous data from the Gastroenterology group showed that *TCF7L2* specific deletions are associated to synchronous metastases and suggested that the genotype of *TCF7L2* polymorphisms, localized in the promoter region, in tumours from patients fulfilling the BC, might be associated to increased CRC risk.

In the present work the analysis of the rs3814570 polymorphism of the *TCF7L2* gene, in individuals diagnosed with CRC fulfilling the BC and in individuals with sporadic CRC, reinforced the association of those polymorphisms to the fulfilment of the BC. Furthermore, the higher frequency of homozygosity for the normal/ancestral allele of this polymorphism, in tumours from individuals fulfilling the BC, was not transversal to all criteria but it was found to be specific of the BC1, BC2 or BC3. This fact suggest that this polymorphic status is associated with early age of onset of CRC and with the development of synchronous or metachronous tumors. This same association was also observed in DNA samples from normal mucosa and, in the case of DNA samples from peripheral blood, this association is maintained at least for the development of synchronous or metachronous tumours. Therefore, this finding in the normal colonic epithelial cells and at the germline level, pinpoints, for the first time, the homozygosity for the normal/ancestral allele of rs3814570 as a potential risk factor for the development of synchronous or metachronous CRC and early onset of CRC.

The finding that in some individuals, the heterozygous genotype change to homozygous at the normal mucosa or at the tumour level, either as a full loss of heterozygosity or as potential mosaicism led us to suggest the association to aUPD, a mechanism that alter the expression of the genes involved, in this case *TCF7L2*, having impact in carcinogenesis. The fact that the increase in homozygosity was also observed in blood in CB2, suggest that UPD at the *TCF7L2* promoter may also occur at the germline level and may explain the risk factor for CRC development, since the homozygosity status was associated to increased expression of *TCF7L2* in peripheral blood, in this study.

Notably, the homozygosity for the normal/ancestral allele of rs3814570 was also associated to the development of synchronous metastases which suggest that aUPD in this region of *TCF7L2* may also be a risk factor for metastasis and poor prognosis.

The work performed towards a precision therapy to metastatic CRC using the cell line derived from metastasis LoVo, led us to test the synergic effect of targeted drugs and nutraceutics. The results led us to suggest that the combination between GANT61 and SFN could attenuate the mesenchymal phenotype by downregulating the EMT markers and induce cell cycle arrest of metastatic cells, however more assays need to be performed in order to have more robust conclusions.

5 References

Agyeman, A., Jha, B., Mazumdar, T. and Houghton, J. (2014). Mode and specificity of binding of the small molecule GANT61 to GLI determines inhibition of GLI-DNA binding. Oncotarget, 5(12), pp4492-503.

Alazzouzi, H., Alhopuro, P., Salovaara, R., Sammalkorpi, H., Javinen, H., Mecklin, J.,et al. (2005). SMAD4 as a Prognostic Marker in Colorectal Cancer. *Clinical Cancer Research*, 11(7), pp.2606-2611.

Albuquerque, C., Baltazar, C., Filipe, B., Penha, F., Pereira, T., Smits, R., et al. (2010). Colorectal cancers show distinct mutation spectra in members of the canonical WNT signaling pathway according to their anatomical location and type of genetic instability. *Genes, Chromosomes and Cancer*, 49(8), pp.746-759.

Albuquerque, C., R.M. Bakker, E., van Veelen, W. and Smits, R. (2011). Colorectal cancers choosing sides. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1816(2), pp.219-231.

Al-Sohaily, S., Biankin, A., Leong, R., Kohonen-Corish, M. and Warusavitarne, J. (2012). Molecular pathways in colorectal cancer. *Journal of Gastroenterology and Hepatology*, 27(9), pp.1423-1431.

Andersen, C., Wiuf, C., Kruhøffer, M., Korsgaard, M., Laurberg, S. and Ørntoft, T. (2007). Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis*, 28(1), pp.38-48.

Angus-Hill, M., Elbert, K., Hidalgo, J. and Capecchi, M. (2011). T-cell factor 4 functions as a tumor suppressor whose disruption modulates colon cell proliferation and tumorigenesis. *Proceedings of the National Academy of Sciences*, 108(12), pp.4914-4919.

Aroeira, L., Aguilera, A., Sánchez-Tomero, J., Bajo, M., del Peso, G., Jiménez-Heffernan, et al.(2007). Epithelial to Mesenchymal Transition and Peritoneal Membrane Failure in Peritoneal Dialysis Patients: Pathologic Significance and Potential Therapeutic Interventions. *Journal of the American Society of Nephrology*, 18(7), pp.2004-2013.

Bates, R. and Mercurio, A. (2005). The epithelial-mesenchymal tansition (EMT) and colorectal cancer progression. *Cancer Biology & Therapy*, 4(4), pp.371-376.

Berg, K., Eide, P., Eilertsen, I., Johannessen, B., Bruun, J., Danielsen, et al. (2017). Multi-omics of 34 colorectal cancer cell lines - a resource for biomedical studies. *Molecular Cancer*, 16(116). doi: 10.1186/s12943-017-0691-y

Brabletz, T., Kalluri, R., Nieto, M. and Weinberg, R. (2018). EMT in cancer. *Nature Reviews Cancer*, 18(2), pp.128-134.

Bush, B., Brock, A., Deng, J., Nelson, R. and Sumter, T. (2012). The Wnt/β-catenin/T-cell factor 4 pathway up-regulates high-mobility group A1 expression in colon cancer. *Cell Biochemistry and Function*, 31(3), pp.228-236.

Centelles, J. (2012). General Aspects of Colorectal Cancer. *ISRN Oncology*, [online] 2012, pp.1-19. Available at: https://www.researchgate.net/publication/233841076_General_Aspects_of_Colorectal_Cancer [Accessed 16 Mar. 2019].

Chen, C., Zhao, S., Karnad, A. and Freeman, J. (2018). The biology and role of CD44 in cancer progression: therapeutic implications. *Journal of Hematology & Oncology*, 11(1), pp.1-23.

Chen, E., Xu, X. and Liu, T. (2018). Hereditary Nonpolyposis Colorectal Cancer and Cancer Syndromes: Recent Basic and Clinical Discoveries. *Journal of Oncology*, 2018, pp.1-11.

Chen, F., Sundaram, V., Chew, T. and Ladabaum, U. (2017). Advanced-Stage Colorectal Cancer in Persons Younger Than 50 Years Not Associated With Longer Duration of Symptoms or Time to Diagnosis. *Clinical Gastroenterology and Hepatology*, 15(5), pp.728-737.e3.

Chen, J., Yuan, T., Liu, M. and Chen, P. (2013). Association between TCF7L2 Gene Polymorphism and Cancer Risk: A Meta-Analysis. *PLoS ONE*, 8(8), p.e71730.

Cingilli Vural, H. (2017). Tcf7l2 Gene Polymorphism in T2dm with Patients in Turkish Population. *Journal of Clinical Epigenetics*, 03(03), pp.1-4.

Colon and Rectum Cancer Staging. (2009). 7th ed. [pdf] Definitions American Joint Committee on Cancer, p.1. Available at: https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/staged.html [Accessed 15 Aug. 2019].

Colussi, D., Brandi, G., Bazzoli, F. and Ricciardiello, L. (2013). Molecular Pathways Involved in Colorectal Cancer: Implications for Disease Behavior and Prevention. *International Journal of Molecular Sciences*, 14(8), pp.16365-16385.

Duarte, M., (2015). Envolvimento de variantes genéticas específicas na suscetibilidade para o risco aumentado de cancro do cólon e reto e na agressividade tumoral. Escola Superior de Saúde Egas Moniz, Caparica

Duval, A., Gayet, J., Zhou, X., Iacopetta, B., Thomas, G. and Hamelin, R. (1999). Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability. *Cancer Research*, 1(59), pp.4213-5.

Engel, E. (1980). A new genetic concept: Uniparental disomy and its potential effect, isodisomy. *American Journal of Medical Genetics*, 6(2), pp.137-143.

Favoriti, P., Carbone, G., Greco, M., Pirozzi, F., Pirozzi, R. and Corcione, F. (2016). Worldwide burden of colorectal cancer: a review. *Updates in Surgery*, 68(1), pp.7-11.

Francisco, I., Albuquerque, C., Lage, P., Belo, H., Vitoriano, I., Filipe, B., et al.(2011). Familial colorectal cancer type X syndrome: two distinct molecular entities?. *Familial Cancer*, 10(4), pp.623-631.

Gaillard, F. and J Bell, D. (n.d.). *TNM staging system | Radiology Reference Article | Radiopaedia.org*. [online] Radiopaedia.org. Available at: https://radiopaedia.org/articles/tnm-staging-system-2 [Accessed 12 Aug. 2019].

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, et al.,(2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, 21(11), pp.1350-1356.

Gulwani, H. (2013). *Normal histology*. [online] Pathologyoutlines.com. Available at: https://www.pathologyoutlines.com/topic/colonhistology.html [Accessed 4 May 2019].

Half, E., Bercovich, D. and Rozen, P. (2009). Familial adenomatous polyposis. *Orphanet Journal of Rare Diseases*, 4(22), pp.1-23.

Hanahan, D. and Weinberg, R. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), pp.646-674.

Jin, T. (2016). Current Understanding on Role of the Wnt Signaling Pathway Effector TCF7L2 in Glucose Homeostasis. *Endocrine Reviews*, 37(3), pp.254-277.

Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nature Reviews Molecular Cell Biology*, 7(5), pp.335-346.

Johnson, C., Wei, C., Ensor, J., Smolenski, D., Amos, C., Levin, B. and Berry, D. (2013). Meta-analyses of colorectal cancer risk factors. *Cancer Causes & Control*, 24(6), pp.1207-1222.

Kahai, P., Mandiga, P. and Lobo, S. (2019). *Anatomy, Abdomen and Pelvis, Large Intestine*.[online] Treasure Island (FL): StatPearls Publishing LLC. Available at: https://www.ncbi.nlm.nih.gov/books/NBK470577/[Accessed 10 Mar.2019]

Kalluri, R. and Weinberg, R. (2009). The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation*, 119(6), pp.1420-1428.

Kanth, P., Grimmett, J., Champine, M., Burt, R. and Samadder, J. (2017). Hereditary Colorectal Polyposis and Cancer Syndromes: A Primer on Diagnosis and Management. *American Journal of Gastroenterology*, 112(10), pp.1509-1525.

Kerr, D. and Midgley, R. (2001). Conventional cytotoxic and novel therapeutic concepts in colorectal cancer. *Expert Opinion on Investigational Drugs*, 10(6), pp.1011-1019.

Le Bacquer, O., Shu, L., Marchand, M., Neve, B., Paroni, F., Kerr Conte, J., Pattou, F., Froguel, P. and Maedler, K. (2011). TCF7L2 splice variants have distinct effects on β-cell turnover and function. *Human Molecular Genetics*, 20(10), pp.1906-1915.

Leslie, A., Carey, F., Pratt, N. and Steele, R. (2002). The colorectal adenoma-carcinoma sequence. *British Journal of Surgery*, 89(7), pp.845-860.

Lynch, H., Lynch, J. and Attard, T. (2009). Diagnosis and management of hereditary colorectal cancer syndromes: Lynch syndrome as a model. *Canadian Medical Association Journal*, 181(5), pp.273-280.

Lynch, H., Lynch, J., Shaw, T. and Lubiński, J. (2003). HNPCC (Lynch Syndrome): Differential Diagnosis, Molecular Genetics and Management - a Review. *Hereditary Cancer in Clinical Practice*, 1(1), pp.7-18.

Murphy, C. and Singal, A. (2018). Establishing a research agenda for early-onset colorectal cancer. *PLOS Medicine*, 15(6), p.e1002577.

Makishima, H. and Maciejewski, J. (2011). Pathogenesis and Consequences of Uniparental Disomy in Cancer. *Clinical Cancer Research*, 17(12), pp.3913-3923.

Melcher, R., Al-Taie, O., Kudlich, T., Hartmann, E., Maisch, S., Steinlein, C., et al. (2007). SNP-Array genotyping and spectral karyotyping reveal uniparental disomy as early mutational event in MSS- and MSI-colorectal cancer cell lines. *Cytogenetic and Genome Research*, 118(2-4), pp.214-221.

Nazemalhosseini Mojarad, E., JK Kuppen, P., Asadzadeh Aghdaei, H. and Reza Zali, M. (2013). The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench*, 6(3), pp.120-128.

Nguyen, H. and Duong, H. (2018). The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy (Review). *Oncology Letters*, 16(1), pp.9-18.

Perea, J., Alvaro, E., Rodríguez, Y., Gravalos, C., Sánchez-Tomé, E. and Rivera, B., et al., (2010). Approach to early-onset colorectal cancer: Clinicopathological, familial, molecular and immunohistochemical characteristics. *World Journal of Gastroenterology*, 16(29), pp.3697-3703.

Pereira, L., Silva, P., Duarte, M., Rodrigues, L., Duarte, C., Albuquerque, C., et al., (2017). Targeting Colorectal Cancer Proliferation, Stemness and Metastatic Potential Using Brassicaceae Extracts Enriched in Isothiocyanates: A 3D Cell Model-Based Study. *Nutrients*, 9(4), p.368.

Qiagen.com. (2013). *DNA Gel Extraction Kit: QIAquick Gel Extraction Kit*. [online] Available at: https://www.qiagen.com/us/products/discovery-translational-research/dna-rna-purification/dna/qiaquick-gel-extraction-kit/#productdetails [Accessed 12 Jun. 2019].

Rodrigues, L., Silva, I., Poejo, J., Serra, A., Matias, A., Simplício, A., et al., (2016). Recovery of antioxidant and antiproliferative compounds from watercress using pressurized fluid extraction. *RSC Advances*, 6(37), pp.30905-30918.

Sanchez-Tillo, E., de Barrios, O., Siles, L., Cuatrecasas, M., Castells, A. and Postigo, A. (2011). β-catenin/TCF4 complex induces the epithelial-to-mesenchymal transition (EMT)-activator ZEB1 to regulate tumor invasiveness. Proceedings of the National Academy of Sciences, 108(48), pp.19204-19209.

Santos, M., (2018).Estudo de biomarcadores e de novas terapias na tumorigénese e metastização do cancro do cólon e do reto, Mestrado, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa,Caparica.

Scheel, C. and Weinberg, R. (2012). Cancer stem cells and epithelial–mesenchymal transition: Concepts and molecular links. *Seminars in Cancer Biology*, 22(5-6), pp.396-403.

Sehgal, R., Sheahan, K., O'Connell, P., Hanly, A., Martin, S. and Winter, D. (2014). Lynch Syndrome: An Updated Review. *Genes*, 5(3), pp.497-507.

Simoneaux, R. (2018). The Four Colorectal Cancer Consensus Molecular Subtypes. *Oncology Times*, 40(6), pp.10-11.

Sinicrope, F. (2010). DNA mismatch repair and adjuvant chemotherapy in sporadic colon cancer. *Nature Reviews Clinical Oncology*, 7(3), pp.174-177.

Stigliano, V., Sanchez-Mete, L., Martayan, A. and Anti, M. (2014). Early-onset colorectal cancer: A sporadic or inherited disease?. *World Journal of Gastroenterology*, 20(35), pp.12420-12430.

Tang, W., Dodge, M., Gundapaneni, D., Michnoff, C., Roth, M. and Lum, L. (2008). A genome-wide RNAi screen for Wnt/ -catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. *Proceedings of the National Academy of Sciences*, 105(28), pp.9697-9702.

Tomimaru, Y., Xu, C., Nambotin, S., Yan, T., Wands, J. and Kim, M. (2013). Loss of exon 4 in a human T-cell factor-4 isoform promotes hepatic tumourigenicity. *Liver International*, pp.1536-1548. 1536–1548.

Torabi, K., Miró, R., Fernández-Jiménez, N., Quintanilla, I., Ramos, L., Prat, E., et al. (2015). Patterns of somatic uniparental disomy identify novel tumor suppressor genes in colorectal cancer. *Carcinogenesis*, 36(10), pp.1103-1110.

Tuna, M., Ju, Z., Smid, M., Amos, C. and Mills, G. (2015). Prognostic relevance of acquired uniparental disomy in serous ovarian cancer. *Molecular Cancer*, 14(29), pp.1-10.

Tuna, M., Smid, M., Zhu, D., Martens, J. and Amos, C. (2010). Association between Acquired Uniparental Disomy and Homozygous Mutations and HER2/ER/PR Status in Breast Cancer. *PLoS ONE*, 5(11), p.e15094.

Vanduchova, A., Anzenbacher, P. and Anzenbacherova, E. (2018). Isothiocyanate from Broccoli, Sulforaphane, and Its Properties. *Journal of Medicinal Food*, 22(2), pp.1-6.

Walker, F., Zhang, H., Odorizzi, A. and Burgess, A. (2011). LGR5 Is a Negative Regulator of Tumourigenicity, Antagonizes Wnt Signalling and Regulates Cell Adhesion in Colorectal Cancer Cell Lines. *PLoS ONE*, 6(7), p.e22733.

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I. and Kerr, D. (2009). Genetic prognostic and predictive markers in colorectal cancer. *Nature Reviews Cancer*, 9(7), pp.489-499.

Weise, A., Bruser, K., Elfert, S., Wallmen, B., Wittel, Y., Wöhrle, S., et al.,(2009). Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/β-catenin targets. *Nucleic Acids Research*, 38(6), pp.1964-1981.

Willauer, A., Liu, Y., Pereira, A., Lam, M., Morris, J., Raghav, et al. (2019). Clinical and molecular characterization of early-onset colorectal cancer. *Cancer*, 125(12), pp.2002-2010.

Wu, C., Zhu, X., Liu, W., Ruan, T. and Tao, K. (2017). Hedgehog signaling pathway in colorectal cancer: function, mechanism, and therapy. *OncoTargets and Therapy*,10, pp.3249-3259.

Wu, Q., Yang, Y., Vogtmann, E., Wang, J., Han, L., Li, H. and Xiang, Y. (2013). Cruciferous vegetables intake and the risk of colorectal cancer: a meta-analysis of observational studies. *Annals of Oncology*, 24(4), pp.1079-1087.

Wu, Z., Brabletz, T., Fearon, E., Willis, A., Hu, C., Li, X., et al.(2012). Canonical Wnt suppressor, Axin2, promotes colon carcinoma oncogenic activity. *Proceedings of the National Academy of Sciences*, 109(28), pp.11312-11317.

Zeinalian, M., Hashemzadeh-Chaleshtori, M., Salehi, R. and Emami, M. (2018). Clinical Aspects of Microsatellite Instability Testing in Colorectal Cancer. *Advanced Biomedical Research*, 7(28). Available at: http://www.advbiores.net/text.asp?2018/7/1/28/225593 [Accessed 5 Mar.2019]

7 Appendix

Appendix I: Preparation of the solutions used in agarose electrophoresis

1. **TBE Buffer 1x:**
    a. Measure 200mL of *TBE Buffer 10%* (0,89M Tris Borato ph 8,3 + 20mM Na_2 EDTA – *National Diagnostics*). Add 1800mL of ddH_2O, for a total volume of 2L and mix by inversion.

2. **2% (w/v) agarose gel:**
    a. Weight 5g of agarose (*Cleaver, Scientific Ltd*) in a 500mL capacity Erlenmeyer.
    b. Measure 250mL of *TBE buffer 1x,* and mix them together.
    c. Heat the mixture in the microwave for 5 minutes until the agarose is clearly dissolved, during this time, stir the solution to homogenize.
    d. In the *hotte*, add 12.5µl of ethidium bromide (*Panreac,AppliChem*) stirring once again.
    e. Pour gently the solution into a specific electrophoresis tray with 30 well combs in its specific places. Let the agarose gel set, in the hotte, for at least 25 minutes.

3. **1% (w/v) agarose gel:**
    a. Weight 1.5g of agarose (*Cleaver, Scientific Ltd*) in a 250mL capacity Erlenmeyer.
    b. Measure 150mL of *TBE buffer 1x*, and mix them together.
    c. Heat the mixture in the microwave for 5 minutes until the agarose is clearly dissolved, during this time, stir the solution to homogenize.
    d. In the *hotte*, add 7.5µl of ethidium bromide (*Panreac, AppliChem*) stirring once again.
    e. Pour gently the solution into a specific electrophoresis tray with 20 well combs in its specific places. Let the agarose gel set, in the hotte, for at least 25 minutes.

4. **0.8% (w/v) agarose gel:**
    a. Weight 2g of agarose (*Cleaver, Scientific Ltd*) in a 250mL capacity Erlenmeyer.
    b. Measure 150mL of *TBE buffer 1x*, and mix them together.
    c. Heat the mixture in the microwave for 5 minutes until the agarose is clearly dissolved, during this time, stir the solution to homogenize.
    d. In the *hotte*, add 7.5µl of ethidium bromide (*Panreac, AppliChem*) stirring once again.
    e. Pour gently the solution into a specific electrophoresis tray with 20 well combs in its specific places. Let the agarose gel set, in the hotte, for at least 25 minutes.

5. **0.7% (w/v) agarose gel:**
    a. Weight 1.05g of agarose (*SeaKem® GTG®, Genetic Technology Grade*™) in a 250mL capacity Erlenmeyer.
    b. Measure 150mL of *TBE buffer 1x*, and mix them together.

c. Heat the mixture in the microwave for 5 minutes until the agarose is clearly dissolved, during this time, stir the solution to homogenize.
d. In the *hotte*, add 7.5µl of ethidium bromide (*Panreac, AppliChem*) stirring once again.
e. Pour gently the solution into a specific electrophoresis tray with 20 well combs in its specific places. Let the agarose gel set, in the hotte, for at least 20 minutes.

6. **Orange 5x:** Add together 1.5g of Ficoll *(Sigma),* 0.125g of Orange G 5x *(Sigma)* to 50mL of ddH_2O under agitation. Aliquot and store at -20°C.

7. **Orange 1x:** In an *eppendorf-like* tube of 1.5ml mix 1000µL of ddH_2O with 400µl of *Orange 5x*, vortex the solution to homogenate and store at 4°C.

8. **Preparation of the weight markers:** The weight markers *GeneRuler 50bp DNA Ladder (Thermofisher Scientific, 0.5*µg/µL*), Lambda DNA/HindIII* marker (Fermentas, 0.5µg/µL) and 1Kb marker (Thermo Scientific, 0.5µg/µL) are prepared for a final volume of 100µl with the respective proportion: 70µl of ddH_2O; 25µl of *Orange 5x*; and 5µl of the selected weight marker. Mix the solution in the vortex and preserve it at 4°C.

Appendix II- Amplification programs

Table II.1- Amplification program for PCR reactions using the *Amplitaq™ Gold* kit (*Applied Biosystems™*).

Step	Temperature (°C)	Time	Number of cycles
Initial Denaturation	95	5 minutes	1
Denaturation	94	50 seconds	40
Annealing	Variable	30 seconds	
Elongation	70	50 seconds	
Final Elongation	70	7 minutes	1
Hold	15	∞	---

Table II.2- Long Expand PCR reaction program.

Step	Temperature (ºC)	Time	Number of cycles
Initialization	92	2 minutes	1
Denaturation	92	10 seconds	10
Annealing	Variable	2 minutes	
Elongation	68	variable*	
Second denaturation	92	15 seconds	25
Second annealing	Variable	30 seconds	
Second elongation	68	variable*+20 seconds**	
Final elongation	68	7 minutes	1
Hold	15	∞	---

Legend: *-Elongation time depends on fragment length: use 2 minutes for up to 3kb, 4 minutes for 6kb, 8 minutes for 10kb, 15 minutes for 20kb and 20minutes for 30kb; **- cycle elongation for each successive cycle. For example, cycle no.11 is 20 seconds longer than cycle no.10 and cycle no.12 is 40 seconds longer than cycle no.10, etc.

Table II.3- Sanger sequencing reaction program.

Step	Temperature (°C)	Time	Number of cycles
Initial Denaturation	96	5 minutes	1
Denaturation	95	10 seconds	25
Annealing	Variable	5 seconds	
Elongation	60	4 minutes	
Hold	4	∞	---

Table II.4- RT-PCR program

Step	Temperature (ºC)	Time	Number of cycles
Denaturation	70	10 minutes	1
Hold	10	∞	
cDNA synthesis	42	60 minutes	
Elongation	70	15 minutes	
Hold	4	∞	

Table II.5- Copy number and gene expression qPCR program.

Step	Temperature (ºC)	Time	Number of cycles
Initialization	50	2 minutes	1
Hold	95	2 minutes	1
Denaturation	95	15 seconds	40
Annealing/ Elongation	60	60 seconds	

Table II.6- qPCR program.

Step	Temperature (ºC)	Time	Number of cycles
Initialization	50	2 minutes	1
Hold	95	10 minutes	1
Denaturation	95	15 seconds	40
Annealing/ Elongation	60	60 seconds	
Melt curve stage	95	15 seconds	∞
	60	60 seconds	
	95	1 second	

Appendix III: Purification with QIAquick Gel Extraction Kit

a. Add 300µL of Buffer QG Solubilization Buffer to each 100mg of agarose gel to a eppendorf-like tube, where were previously stored the DNA fragments that were excised from the agarose gel.
b. Incubate at 50ºC for 10 minutes maximum or until the gel slice is completely dissolved. It is very important to verify if the color of the solution maintains yellow after the slice have dissolved to make sure that the pH is ≤7.5, which is the optimal pH for DNA binding.
c. Add 100µL of isopropanol to increase the yield of DNA fragments with sizes that could be inferior 500bp and superior 4kb and mix by inversion.
d. Place a QIAquick spin column in a provided 2mL collection tube and apply the sample to the center of the column.
e. Centrifuge for 1 minute at 130000rpm, at room temperature.
f. Discard the flow-through.
g. Place back in the collection tube and add 500µL of Buffer QG Solubilization Buffer.
h. Centrifuge for 1 minute at 130000rpm, at room temperature.
i. Wash the column with 750µL of Buffer PE, add in the center of the column and let stand for 2 minutes
j. Centrifuge for 1 minute at 130000rpm, at room temperature.
k. Discard the flow-through
l. Centrifuge again the column in order to dry the QIAquick membrane.

m. Discard the collection tube and place the column in a new 1.5mL eppendorf like tube.
n. Add 30µL of elution buffer right in the center of the column, let the tube stand for 1 minute before the last centrifugation for 1 minute at 130000rpm, at room temperature.
o. Store the purified DNA at -20ºC.

Appendix IV: Protocol for DNA precipitation and purification

a. Prepare a mix solution composed by 50µL absolute ethanol; 2µL of sodium acetate (3M, pH4.5); and 2µL of EDTA (125mM, pH8) per sample, in a eppendorf-like tube.
b. Vortex the solution followed by a spin down.
c. Distributed 54µL of the mix into new eppendorf-like tubes of 1.5mL, one tube per sample.
d. Add the total volume of the sequencing product, around 20µL, to each tube.
e. Vortex and incubate the samples at room temperature for 15 minutes.
f. Centrifuge the samples for 30 minutes at 14000rpm and 4°C, always making sure that the tubes have the same orientation in the centrifuge due to the lack of visualization of the DNA pellet.
g. Remove the supernatant carefully with a pipette.
h. Add 100µL of ethanol 70% (v/v) to each eppendorf-like tube.
i. Centrifuge for 15 minutes, following the previous conditions (14000rpm, 4°C).
j. Remove the supernatant as previously.
k. Dry the samples at 37°C for 10 minutes with the lids open for better evaporation of the ethanol.
l. Store the samples at 4°C.

Solutions:

Sodium acetate (3M, pH4.5) - Dissolve 24.8g of sodium acetate in ddH_2O for a final volume of 100ml. Adjust the pH level by using acetic acid.

EDTA (125mM, pH8) - Dissolve 4.65g of EDTA at pH8 for a final volume of 100mL of ddH_2O, adjusting the pH level by using NaOH.

Ethanol 70% (v/v) - Mix 30ml of ddH_2O to 70ml of absolute ethanol and homogenize the solutions.

Appendix V- PCR reaction conditions for each primer

Table V.1- Sequence, molecular weight and PCR conditions of specific *TCF7L2* gene polymorphisms.

***TCF7L2* gene polymorphism**	**Primer sequence**		**Fragment molecular weight (bp)**	**Annealing temperature (°C)**	**$MgCl_2$ * (µL)**	**Primer ** (µL)**
rs3814570	F	5'-TGGCCTTGTCAATCTCGG-3'	181	62	1.25	0.3
	R	5'-TTTTTCAGGGAAGGGTGGGA-3'				
rs10885395	F	5'-CCACTCTACATTTAAAAGAATC-3'	130	55	1.25	0.3
	R	5'-GTGGGACTTAACATTTCATG-3'				

F- Forward; R- Reverse; *$MgCl_2$ concentration: 25mM; **Primer concentration: 10pmol/µL

Table V.2- Sequence, molecular weight and PCR conditions of *TCF7L2* gene exon 1.

***TCF7L2* gene**	**Primer sequence**		**Fragment molecular weight (bp)**	**Annealing temperature (ºC)**	**$MgCl_2$* (µL)**	**Primer** (µL)**
TCF7L2 SEQ 1	F	5'-CTTCCAAAATTGCTGCTGGTG-3'	241	64	1.5µL	0.15
	R	5'-CCCGAGGGGCTTTTCCTA-3'				

F- Forward; R- Reverse; *$MgCl_2$ concentration: 25mM; **Primer concentration: 10pmol/µL

Table V.3- Sequence, molecular weight and Expand PCR Long conditions of *TCF7L2* gene.

***TCF7L2* gene**	**Primer sequence**		**Molecular weight (bp)**	**Annealing temperature (ºC)**
TCF7L2 UPS P0.2	F	5'-TAAGGAACCATCACCCACGC-3'	7542	55
TCF7L2 int 3.1	R	5'-TCCCGTTTAAGATGCCCCAC-3'		56
TCF7L2 int 3.2	R	5'-AGGGCATCAGTCAAAGCTCC-3'	9711	53

F- Forward; R- Reverse; int- intron

Appendix VI: DNA or RNA quantification using the Qubit™ fluorometer

a. Bring the standards solutions, from the specific kit (dsDNAHs or Br RNA assays), from 4ºC and thaw at room temperature for at least 30 minutes.
b. Set two assay tubes for the standard solutions and one tube for each sample.
c. For each tube, prepare a working solution composed by: 199µL of Qubit™ buffer and 1µL of Qubit™ reagent (200x) from the specific assays, and mix by vortex to homogenate.
d. Distribute 198µL of the working solution in the assay tubes for each sample and 190µL in the assay tubes for the standard solutions.

e. Vortex the standard solutions 1 and 2 and pipette 10µL of each to the respective standard assay tubes.
f. Pipette 2µL of each sample to the sample assay tubes.
g. Vortex all tubes and incubate the tubes for 2 minutes at room temperature.
h. Insert the standard tubes in the Qubit™ 2.0 Fluorometer to calibrate and then take reads the samples.

Appendix VII: Trypsin dilution (0.1% v/v)

In order to make a trypsin dilution of 0.1% v/v, is used 1mL of Trypsin 2.5% (*Invitrogen™*) and is diluted in 24mL of PBS in a *Falcon* tube of 50mL. This solution is mixed first by up-and-down with a pipette followed by inversion of the falcon. This *Falcon* tube needs to be properly closed with *parafilm* and stored at 4°C.

Appendix VIII: Total RNA extraction from cells with RNeasy Mini Kit (*QIAGEN*)

a. In the laminar flow chamber, prepare the columns and the 1.5mL collector tubes for each sample and make the sample identification.
b. Thaw the samples in ice.
c. Pipette 600µL of 70% ethanol to each collector tube that contain the cells and mix by up-and-down.
d. Transfer 700µL to the center of each column and centrifuge the samples for 30 seconds at 14000rpm at room temperature.
e. Discard the flow-through and add the rest of the 500µL left in the collector tube that contains the cells.
f. Centrifuge for 30 seconds at 14000rpm at room temperature.
g. Discard the flow-through and add to each column 700µL of Buffer RW1.
h. Centrifuge for 30 seconds at 14000rpm at room temperature and discard the flow-through.
i. Add 500µL of Buffer RPE to center of each column.
j. Centrifuge for 30 seconds at 14000rpm at room temperature and discard the flow-through.
k. Add again 500µL of Buffer RPE to center of each column.
l. Centrifuge for 2 minutes at 14000rpm at room temperature, prepare new collector tubes of 1.5mL.
m. Put the column in the new collector tubes and discard the used ones.
n. Centrifuge for 1 minute at 14000rpm at room temperature, discard the collector tube and place the columns in new eppendorf-like tubes of 1.5mL.
o. Add 40mL of RNase free water in the center of the column.
p. Centrifuge for 1 minute at 14000rpm at room temperature, with the lids of the eppendorf-like tubes open.
q. Pipette the 40mL that flow-through the column into the center of the column.
r. Centrifuge for 1 minute at 14000rpm at room temperature, with the lids of the eppendorf-like tubes open.

s. Discard the columns and store the RNA at -80ºC.

Appendix IX: Primers used in the qPCR analysis

Table IX.1- Sequences and conditions of the primers used in the qPCR analysis. These primers correspond to the genes of interest for the CRC gene expression study in a LoVo cell line.

Gene	Primer sequence		Primer concentration (pmol/µL)	Commercial Kit
GAPDH	F	5'-TGCACCACCAACTGCTTAGC -3'	7.5	Power SYBR® Green PCR Master Mix
	R	5'-GGCATGGACTGTGGTCATGAG -3'		
TCF7L2	F	5'-GCTGATGTCAAATCGTCTCTAGTCA -3'	5	Power SYBR® Green PCR Master Mix
	R	5'-TCTTGGCCGCTTCTTCCA-3		
TCF7L2 (Isoform 1-6)	F	5'-CCTCTCTAGATGTCTAACAAAGTGC-3'	5	KAPA SYBR® FAST qPCR Kit Master Mix Universal
	R	5'-CACACTTACCAGCCGACGTA-3'		
VIM	F	5'-GCTCAATGTTAAGATGGCCCTTG-3'	5	KAPA SYBR® FAST qPCR Kit Master Mix Universal
	R	5'-AAGAGGCAGAGAAATCCTGCTC-3'		
LGR5	F	5'-GAGGATCTGGTGAGCCTGAGAA-3'	5	KAPA SYBR® FAST qPCR Kit Master Mix Universal
	R	5'-CATAAGTGATGCTGGAGCTGGTAA-3'		
CDH1	F	5'-GATCCATTTCTTGGTCTACGCC-3'	5	Power SYBR® Green PCR Master Mix
	R	5'-CTGCTTGGATTCCAGAAACGG-3'		
CDKN1A	F	5'- GGTGGACCTGGAGACTCTC-3'	5	Power SYBR® Green PCR Master Mix
	R	5'- GTGGTAGAAATCTGTCATGCTGG-3'		
CCNA2	F	5'- AAGAGGCAGAGAAATCCTGCTC-3'	5	KAPA SYBR® FAST qPCR Kit Master Mix Universal
	R	5'- GTGCAACCCGTCTCGTCTTC-3'		
SOX9	F	5'-CAGCGAACGCAGCACATCAAGAC-3'	5	Power SYBR® Green PCR Master Mix
	R	5'-CTGTAGGCGATCTGTTGGGG-3'		
ZEB1	F	5'-CCTGTGCAGTTACACCTTTGC-3	3	Power SYBR® Green PCR Master Mix
	R	5'-CAGACTGCGTCACATGTCTTTG-3'		
TGF-β1	F	5'-CCTGGACACCAACTATTGCTTC-3	3	Power SYBR® Green PCR Master Mix
	R	5'-GGACCTTGCTGTACTGCGTG-3'		
CD44	F	5'-GCGCAGATCGATTTGAATATAACC-3'	5	KAPA SYBR® FAST qPCR Kit Master Mix Universal
	R	5'-TCCGTCCGAGAGATGCTGTAG-3'		
AXIN2	F	5'-GGTTCTGGCTATGTCTTTGCAC-3'	3.5	Power SYBR® Green PCR Master Mix
	R	5'-CCCACACGATAAGGAGGAATTC-3'		
β-CATENIN	F	5'-TGCCAAGTGGGTGGTATAGAG-3'	3.5	Power SYBR® Green PCR Master Mix
	R	5'-TGGGATGGTGGGTGTAAGAG-3'		
PTCH1	F	5´-GCATACCTCCTAGGTAAACCTC-3'	5	Power SYBR® Green PCR Master Mix
	R	5'-GCATTTCCTCCCAGCTGTCC-3'		
CTCF	F	5'-GAACCCATTCAGGGGAAAAGC-3'	5	Power SYBR® Green PCR Master Mix
	R	5'-TTCGGGCTATGACTGTGTC-3'		
PROM1	F	5'-CCTTCCTGAGGAAATGCTTTGC-3'	3	KAPA SYBR® FAST

	R	5'-GGTTCTTACCTGGTGATTTGCC-3'		qPCR Kit Master Mix Universal

Legend: F- Forward; R-Reverse; Power SYBR® Green PCR Master Mix (*Applied Biosystems*); KAPA SYBR® FAST qPCR Kit Master Mix Universal (KappaBiosystems).

Appendix X:

Table X.1- Genotypes of the rs38145570 polymorphism of the *TCF7L2* gene in DNA samples from peripheral blood from individuals diagnosed with CRC and that meet the Bethesda criteria.

Sample	rs38145570	BC	Sample	rs38145570	BC
A1961	C/C	1	**L2721**	C/T	1
A2198	C/T		**L2722**	C/C	
A2280	T/T		**L2871**	C/C	
A2328	C/C		**L2888**	C/C	
Cas2736s	C/C		**A2008**	C/C	2
Cas2740s	C/C		**A2328**	C/C	
Cas2764s	C/C		**A2564**	C/T	
Cas2870s	C/T		**Cas3372s**	C/T	
Cas2886s	C/C		**Cas3578s**	C/C	
Cas2968s	C/T		**Cas3790s**	C/C	
Cas2982s	C/C		**Cas3876s**	C/C	
Cas3026s	C/C		**L801**	C/C	
Cas3040s	C/C		**L1824**	C/C	
Cas3042s	C/C		**L1964**	C/C	
Cas3090s	T/T		**L1985**	T/T	
Cas3092s	C/C		**L2021**	C/T	
Cas3126s	C/C		**L2043**	C/C	
Cas3214s	C/C		**L2184**	C/T	
Cas3348s	C/C		**L2221**	C/C	
Cas3406s	T/T		**L2254**	C/C	
Cas3426s	C/T		**L2304**	C/C	
Cas3444s	C/T		**L2666**	C/T	
Cas3488s	C/T		**L2672**	C/C	
Cas3490s	C/T		**A1906**	C/C	3
Cas3530s	C/C		**Cas3272s**	C/T	
Cas3548s	C/C		**Cas3794s**	C/T	
Cas3630s	C/C		**L1977**	C/T	
Cas3664s	C/T		**L2378**	C/C	
Cas3694s	C/C		**L2505**	C/T	
Cas3736s	C/T		**Cas2858s**	C/T	4
Cas3738s	C/C		**Cas3630s**	C/T	
Cas3744s	C/T		**Cas3636s**	C/T	
Cas3750s	C/T		**L850**	T/T	
Cas3752s	T/T		**L1587**	C/T	
Cas3792s	C/C		**L1898**	C/C	

Cas3806s	C/T		**L2047**	C/C	
Cas3832s	C/C		**L2054**	C/C	
Cas3836s	C/T		**L2140**	C/T	
Cas3850s	C/C		**L2170**	C/T	
Cas3884s	T/T		**L2215**	C/T	
Cas3890s	C/T		**L2240**	C/C	
Cas3894s	C/C		**L2263**	C/T	
Cas3912s	C/T		**L2331**	C/T	
Cas3928s	C/C		**L2536**	C/C	
Cas3990s	C/T		**A708**	C/C	5
Cas4006s	C/C		**A1682**	C/T	
Cas4054s	C/T		**A1910**	C/C	
Cas4076s	C/C		**Cas2476s**	C/T	
Cas4088s	C/C		**Cas3098s**	C/C	
Cas4100s	C/C		**Cas3554s**	T/T	
Cas4194s	C/C		**Cas3708s**	C/C	
L833	C/C		**Cas3948s**	C/C	
L887	C/C		**Cas3976s**	C/C	
L1067	T/T		**Cas4004s**	C/C	
L1657	C/C		**Cas4038s**	T/T	
L1675	C/C		**L3040**	C/C	
L1769	C/T		**L782 (2ºEXT)**	C/C	
L1864	C/T		**L1103**	T/T	
L1866	C/C		**L1133**	C/T	
L1875	C/C		**L1474**	C/T	
L1880	C/C		**L1560**	C/T	
L1896	T/T		**L1840**	C/T	
L1912	C/T		**L1873**	T/T	
L1966	C/C		**L1892**	C/T	
L2001	T/T		**L1902**	C/T	
L2052	C/C		**L1986**	C/T	
L2118	C/T		**L2014**	C/T	
L2133	C/T		**L2139**	C/T	
L2167	C/T		**L2171**	C/T	
L2170	C/T		**L2237**	C/T	
L2186	T/T		**L2275**	C/C	
L2243	C/C		**L2315**	C/C	
L2244	C/C		**L2342**	C/C	
L2258	C/T		**L2397**	C/C	
L2336	C/T		**L2453**	C/T	
L2469	T/T		**L2464**	C/T	
L2497	C/C		**L2473**	C/C	
L2502	C/T		**L2495**	C/T	
L2528	C/C		**L2531**	C/T	
L2563	C/T		**L2550**	C/C	

L2568	C/C		**L2562**	C/C	
L2631	C/T		**L2731**	C/T	
L2692	C/C		**L2770**	T/T	

Legend: Homozygosity for the ancestral/normal allele (C/C); heterozygous genotype (C/T); homozygosity for the polymorphic allele (T/T).

Appendix XI:

Table XI.1- Representation of the genotype results for tumor, normal mucosa and peripheral blood samples of all cases analyzed. The cases analyzed were organized by individual (Tumor/normal mucosa/peripheral blood), divided by BC. The percentages represent the allelic disequilibrium.

Sample (Tumor/NM/Blood)	Tumor	Normal mucosa	Peripheral blood	BC
Cas4269/4270/L2631	c/T (30-40%; 60-70%)	C/T	C/T	
Cas4267/4268/L2563	C/T	C/T	C/T	
Cas4093/4094/L2721	C/T	C/T	C/T	
Cas4089/-/Cas3912s	C/C	-	C/T	
Cas3989/-/Cas3990s	C/T	-	C/T	
Cas3941/3942/L2502	C/T	C/T*	C/T	
Cas3889/3890/Cas3890s	T/T	T/T	C/T	
Cas3835/3836/Cas3836s	C/C	C/C	C/T	
Cas3769/3768/Cas3750	C*/T	C/T	C/T	
Cas3687/L2258	C/T*	-	C/T	
Cas3663/3664s	C/T*	-	C/T	
Cas3647/3648/L2336	C/t (25-30%; 70-75%)	C/T	C/T	
Cas3627/3628/L2170	T/T	c/T (20-25%; 75-80%)	C/T	1
Cas3591/3592/L2167	C/t (30-40%; 60-70%)	c/T (30-40%; 60-70%)	C/T	
Cas3571/3572/A2198	C/t (30-40%; 60-70%)	C/t (30-40%; 60-70%)	C/T	
Cas3501/3502/L1912	C/T	-	C/T	
Cas3489/3490/Cas3490s	c/T (30-40%; 60-70%)	c/T (30-40%; 60-70%)	C/T	
Cas3487/-/Cas3488s	C/T	-	C/T	
Cas3425/3426/Cas3426s	T/T	C/T	C/T	
Cas3319/-/L1864	C/C	-	C/T	
Cas3277/3278/L2118	c/T (30-40%; 60-70%)	c/T (30-40%; 60-70%)	C/T	
Cas3275/3276/L2133	C/T	C/C	C/T	
Cas2967/2968/Cas2968s	-	C/C	C/T	
Cas2903/2904/L1769	C/T	c/T (30-40%; 60-70%)	C/T	
Cas2869/2870/Cas2870s	C/C	C/T	C/T	
Cas3999/-/A2564	C/T*	-	C/T	
Cas4559/4560/Cas4560s	C/C	C/C	C/T	
Cas4471/4472/4472s	C*/T	C/T*	C/T	
Cas4155/4156/Cas4156s	C/C	C/C	C/C	2
Cas4107/-/L2577	C/C	-	-	
Cas3605/3606/L2184	C*/T	C/T	C/T	
Cas3485/-/-	C/t (20-25%; 75-80%)	-	-	
Cas3419/3420/L2021	C*/T	C*/T	C/T	

Cas3405/-/Cas3406s	T/T	-	T/T	
Cas3371/3372/Cas3372s	C*/T	C/C	C/T	
Cas2747/2748/L1982	C/C	C/C	C/C	
Cas2471/2472/L1694	C/C	C/C	C/C	
/CAS2350/-	-	C/C	-	
-/Cas2320/-	-	C/C	-	
Cas4577/4778	C/T*	C/T*		3
Cas4536s/-/-	-	-	C/T*	
Cas3793/-/Cas3794s	T/T	-	C/T	
Cas3717/-/-	C/C	-	-	
Cas3714/-/-	T/T	-	-	
Cas3557/3558/L2297	C/C	C/C	C/C	
Cas3537/3538/L2505	C/T	C*/T	C/T	
Cas3481/-/-	c/T (15-25%; 75-85%)	-	-	
Cas3461/-/-	C/C	-	-	
Cas3271/3272/Cas3272s	c/T (30-40%; 60-70%)	T/T	C/T	
Cas3223/3224/L1977	C/T*	C/T	C/T	
Cas3093/-/-	C/C	-	-	
Cas2905/-/-	C/C	-	-	
Cas2843/-/-	C/C	-	-	
Cas2793/2794/-	C/C	-	-	
Cas2695/2696/Cas2696s	C/C	C/C	C/C	
Cas2577/-/-	C/C	-	-	
Cas2491/2492/L1710	C/C	C/C	C/C	
Cas2361/2362/L1702	c/T (20-30%; 70-80%)	C/t (5-10%; 90-95%)	C/t (30-40%; 60-70%)	
Cas2349/-/-	C/C	-	-	
Cas2729/2730/L1749	C/C	C/C	C/C	4
Cas3471/3472/L2331	C/t (25-30%; 70-85%)	C/t (30-40%; 60-70%)	C/T	
Cas2663/-/Cas2664s	C/C	-	C/C	
Cas2585/2586/L2068	c/T (20-25%; 80-85%)	C/t (30-40%; 60-70%)	C/T	
Cas2453/2454/A1612	C/C	C/C	C/C	

Legend: - no information/sample nonexistent; * -slight allelic disequilibrium; lower case- accentuated allelic disequilibrium; NM- normal mucosa

Appendix XII:

Table XII.1- Relative gene expression analysis, for exons 1-2 and exons 7-8, of healthy individuals from LS families that were used as references in the gene expression study. Also, it is represented the rs38145570 and rs10885395 polymorphism genotype of the *TCF7L2* gene, of each individual.

Sample	Gene expression assays		Polymorphism	
	Exons 1-2	Exons 7-8	rs38145570	rs10885395
L2868	0.782754	0.765399	C/T	C/T
L2923	1.351465	1.301372	T/T	T/T
L2962	0.501602	0.390219	C/T	C/T
L3009	0.905124	0.860249	C/T	C/T
L3054	0.6355	0.55233	C/C	C/C
L3080	1.104821	1.162454	C/C	C/C

Legend: Homozygosity for the ancestral/normal allele (C/C); heterozygous genotype (C/T); homozygosity for the polymorphic allele (T/T).

Table XII.2- Relative gene expression analysis of DNA samples from individuals diagnosed with CRC that meet one of the Bethesda Criteria, for the exons 1-2 and exons 7-8 regions.

Gene Expression		Sample	Polymorphism
Exons 1-2	Exons 7-8		rs38145570
1,203308	1,19728	**A2008**	C/C
0,703819	0,502227	**L2254**	C/C
0,945098	1,110121	**CAS1924S**	C/C
0,382956	0,402227	**L1618**	C/C
0,580092	0,535556	**CAS1964S**	C/C
0,806493	0,807326	**L2208**	C/C
0,640756	0,65762	**L2141**	C/C
0,792347	1,274839	**L2195**	C/C
0,518156	0,702219	**L2180**	C/C
0,950677	1,070652	**L2128**	C/C
0,799847	0,829125	**L2871**	C/C
0,822747	1,004346	**L1875**	C/C
1,269886	1,468532	**L801**	C/C
0,980519	0,968434	**CAS3948S**	C/C
0,740137	0,891766	**CAS3708S**	C/C
0,540335	0,600511	**CAS936S**	C/C
0,784606	1,185268	**CAS3518S**	C/C
0,55095	0,536358	**CAS3920S**	C/C
0,675453	0,714113	**L3140**	C/C
0,946183	1,765474	**L2327**	C/C
0,864442	1,386737	**L2482**	C/C
0,435004	0,585085	**L2485**	C/C

0,659421	0,870243	**L2486**	C/C
1,07539	1,197146	**A3319**	C/C
1,118479	1,323407	**A3337**	C/C
0,686941	0,696776	**A3344**	C/C
0,512995	0,502227	**L1847**	C/T
0,615793	0,712172	**A1976**	C/T
0,353376	0,358709	**L1499**	C/T
0,730115	0,783972	**L1848**	C/T
0,780405	0,797109	**L2030**	C/T
0,712602	0,893218	**L2494**	C/T
0,363301	0,43558	**L1667**	C/T
0,461143	0,586436	**L2161**	C/T
0,399931	0,620986	**L2156**	C/T
1,196668	1,061354	**L2237**	C/T
0,789289	0,71327	**L3121**	C/T
0,686032	0,747345	**L3133**	C/T
0,71934	0,885543	**A3288**	C/T
0,917512	0,995577	**A3323**	C/T
1,405557	1,974453	**L2186**	T/T
0,827256	0,943757	**A3115**	T/T
1,423405	1,546142	**A3287**	T/T

Legend: Homozygosity for the ancestral/normal allele (C/C); heterozygous genotype (C/T); homozygosity for the polymorphic allele (T/T).

Appendix XIII: Results from the previous studies, in terms of cell proliferation, migration in LoVo cell line, performed by the Gastroenterology group from UIPM-IPOLFG to evaluate the response of different CRC subtypes to different drugs and to combinatory treatments with different nutraceuticals (Santos et al., 2018)

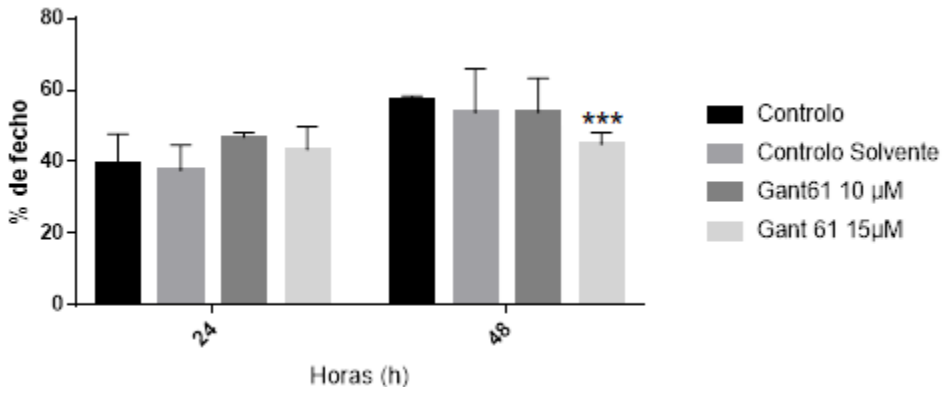


Figure XIII.1- Wound-healing assay in LoVo cell line, performed by the Gastroenterology group from UIPM-IPOLFG. The graph represents the assay 24h and 48h after treatment with GANT61. For each condition the percentage of close ("% de fecho") was calculated in comparison with the initial area at 0h. The control ("Controlo") represents cells without treatment. The solvent control ("Controlo

solvente") represents the cells cultured in solvent. ***p<0.0001, statistical test made with ANOVA two-way.

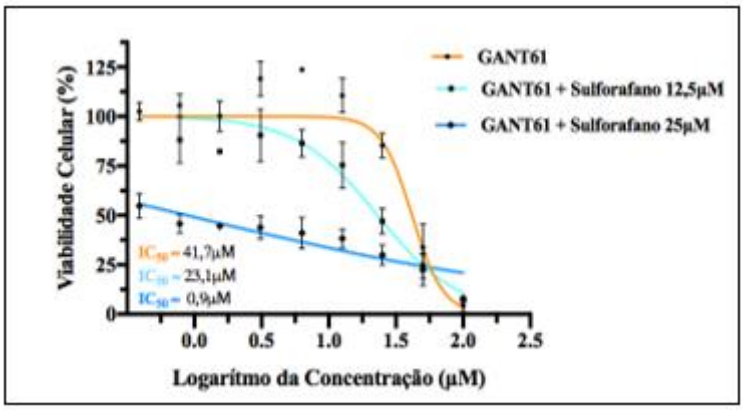


Figure XIII.2- Graphic representation of the cellular viability in percentage by the logarithm of the concentration (µM) of LoVo cell line treated with GANT61, GANT61 and sulforaphane (12.5µM) and GANT61 and sulforaphane (25µM), performed by the Gastroenterology group from UIPM-IPOLFG. It is also represented the respective IC_{50} of each treatment.