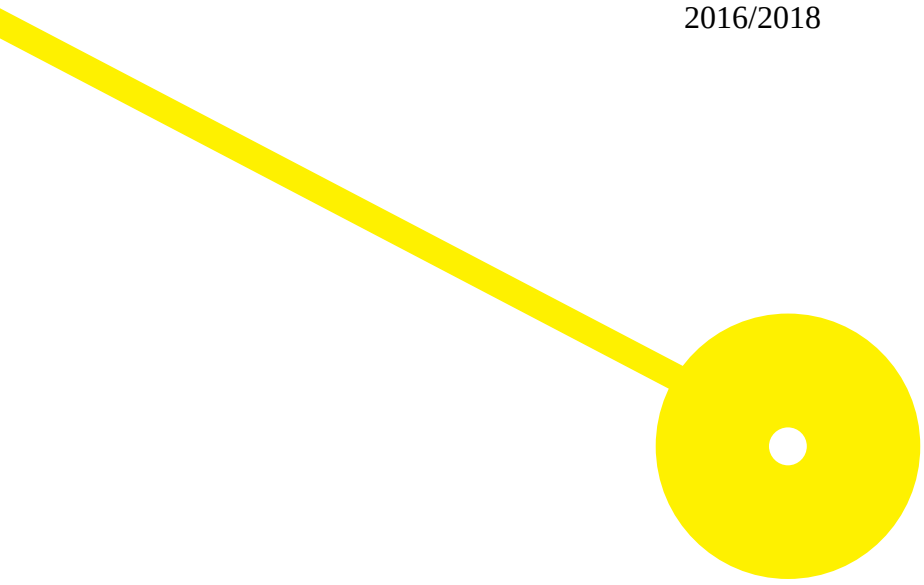
ESCOLA
SUPERIOR
DE SAÚDE
POLITÉCNICO
DO PORTO

M BIOQUÍMICA EM SAÚDE
BIOTECNOLOGIA

New molecular prognostic and predictive biomarkers in stage II colorectal cancer

Sara Maria Azevedo Ribeirinho Soares

2016/2018



Escola Superior de Saúde do Porto
Instituto Politécnico do Porto

Sara Maria Azevedo Ribeirinho Soares

New molecular prognostic and predictive biomarkers in stage II colorectal cancer

Projeto de Dissertação submetido à Escola Superior de Saúde para o cumprimento dos requisitos necessários à obtenção de grau de Mestre em Bioquímica em Saúde ramo de Biotecnologia, realizada sob a orientação científica de Raquel Almeida, Principal Investigator, Differentiation and Cancer, Instituto de Patologia e Imunologia Molecular da Universidade do Porto, sob a coorientação de Patrícia Mesquita, Assistant Researcher, Differentiation and Cancer, Instituto de Patologia e Imunologia Molecular da Universidade do Porto e sob a orientação institucional de Cristina Prudêncio, Professora Coordenadora com Agregação, CBQ.

2016/2018

Dreams are often most profound when they seem the most crazy.
— Sigmund Freud

Agradecimentos

Quero começar por agradecer ao Professor Manuel Sobrinho Simões. Obrigada por me apresentar, no final de 2010, uma Ciência pela qual eu me apaixonei e que me levou a perseguir o sonho de ser Cientista. Infelizmente, essa visão sobre o que é a Ciência não é partilhada por todos, e é, ainda hoje passados 8 anos, uma utopia.

Professor, agradeço por ter sido sempre tão humano comigo e, apesar de ser uma pessoa extremamente solicitada, ter estado sempre disponível para responder a todas as minhas dúvidas e para me ajudar ao longo destes anos. Foi também o Professor quem me apresentou a Doutora Raquel Almeida, logo no primeiro ano de faculdade, que aceitou muito gentilmente receber-me no seu grupo e mostrar-me como é fazer investigação durante um pequeno estágio no verão de 2014. Porque tinha adorado o pouco que já tinha conhecido deste grupo, passados 4 anos, para fazer a Tese de Mestrado, decidi pedir novamente para voltar.

Muitíssimo obrigada Doutora Raquel por ter estado sempre disponível a receber-me. Foi espetacular comigo, agradeço muito a sua dedicação. Obrigada por ter confiado em mim para poder aprender no seu grupo. Espero ter correspondido às suas espectativas e continuar a merecer a sua confiança.

Quando regressei fiquei muito contente ao rever algumas caras. No entanto vinha ainda a conhecer algumas das pessoas que mais me iriam ajudar neste ano de tese, nomeadamente a Doutora Patrícia Mesquita a minha co-coordenadora.

Nada foi mais gratificante do que poder aprender tanto consigo todos durante estes meses. Não foi só ciência que me ensinou. Foi também uma grande amiga. Adorei todos os momentos que partilhámos, principalmente sinto que cresci muito ao longo deste ano com um exemplo de uma excelente profissional e magnífica pessoa ao meu lado. Obrigada por toda a dedicação, foi incansável, o seu esforço foi inestimável. O que eu mais lhe desejo é que um dia as suas filhas, seja qual for a profissão que escolham, tenham a oportunidade de ter alguém dentro da área que elas escolherem que esteja disposta a ajudá-las tanto como a Doutora Patrícia me ajudou a mim. Há coisas que nem os nossos pais podem fazer. Às vezes realmente é preciso ter sorte, e eu tenho tido muita.

Quero agradecer muito também à Filipa Freire, contribuíste muito para a minha aprendizagem e estiveste sempre disposta a ajudar-me. Quero agradecer à Ana Luísa, por seres tão boa companhia novamente agora como há 5 anos, por seres uma eterna alma jovem e a mãezinha do grupo. Obrigada por tudo o que me ensinaste. Quero agradecer ao Alexandre Dias, por ser uma das pessoas mais humildes e generosas que tive o prazer de

conhecer, obrigada por toda a amizade. Quero agradecer à Diana Pádua por me inspirar como uma cientista brilhante. Quero agradecer imenso à Sara Ricardo, para além de toda a ajuda com o meu trabalho, por contagiar esperança que estamos a fazer a coisa certa por perseguir os nossos sonhos, mesmo que sejam fazer Ciência. Quero agradecer também toda a gentileza com que me trataram, com que sempre estiveram prontos para me ajudar e fizeram com que o local de trabalho fosse impecável: Ricardo Coelho, Bruno Pereira, Nair Lopes. Mostraram-se sempre extremamente dispostos para me ajudar.

Agradeço também a todo o resto do pessoal do i3S por me terem integrado sempre com tanta simpatia no vosso instituto.

Agradeço muito também à Doutora Cristina Pudêncio e à Doutora Mónica Vieira por serem professoras espetaculares, terem sido o meu contacto com a faculdade durante este ano de tese, e sobretudo todo o carinho.

Quero agradecer à Professora Doutora Fátima Carneiro e ao Banco de Tecido e Tumores do Hospital de São João por possibilitar a recolha das amostras de tecidos humanos e dos dados clinico-patológicos incluídos neste estudo. Agradeço ainda à Professora Doutora Leonor David pela ajuda na identificação das células tumorais, que permitiu a análise da mutação do BRAF e ao Doutor Valdemar Máximo pelo auxílio na interpretação dos resultados da sequenciação. Ainda agradeço à Professora Doutora Carmen Jerónimo, ao Professor Doutor Rui Henrique e à Dra. Paula Lopes do IPO-Porto, por tudo o que diz respeito à análise de microssatélites.

Por fim, quero agradecer também aqueles que já vinham de antes e fazem parte de mim: Mãe e Pai, o meu amor por vocês é incondicional e não existe tela branca no mundo onde eu possa descrever tudo o que fazem por mim. Maninho, és o meu maior orgulho continua sempre a ser a pessoa perfeita que és. João obrigada por além de namorado seres o meu melhor amigo.

Felicito todos os que têm coragem de continuar a fazer Ciência em Portugal!

É só mais uma etapa que chega ao fim, corremos para abrir a próxima e nem podemos saborear o que realmente daqui levamos e que não se pode substituir.

Abstract

Colorectal cancer (CRC) remains a serious health concern, being the third most commonly diagnosed cancer and the fourth leading cause of cancer-related death in Europe. In Portugal, CRC is both the most frequently diagnosed and the one that causes more cancer-related deaths. The incidence of this disease increased in the last thirty years and it is expected to rise by 60% to more than 2.2 million new cases and 1.1 million cancer deaths by 2030.

Despite advances in diagnosis and treatment, up to 20% of stage II CRC patients subjected to tumor resection with curative intent develop disease recurrence. The therapeutic approach after surgery is yet not consensual, nor effective, and depends to a large extent on disease staging. It would be desirable to have both prognostic biomarkers, that stratify better the patients and thus help to identify who should be treated, and predictive biomarkers, that foretold the likelihood of the benefit of administering a specific chemotherapeutic drug to cancer patients.

With this project, we intend to improve scientific insight into new molecular parameters that could help to distinguish specific subgroups of CRC stage II patients and foresee which can effectively benefit from current adjuvant therapy treatments. For this purpose, in our 230 stage II CRC patient cohort, we investigated by immunohistochemistry the biomarker potential of 3 transcription factors - CDX2, SOX2, SOX9 - and also microsatellite status, by analysis of expression of the mismatch repair proteins. We have also assessed $BRAF^{V600E}$ mutation status by Sanger Sequencing.

In our CRC series, low CDX2 expression and *de novo* SOX2 expression significantly correlated with less tumor differentiation. SOX2 expression showed prognostic value but when considering only cases negative for SOX9. In this case, SOX2 expression resulted in worst disease-free survival. Moreover, in patients treated with adjuvant chemotherapy, *de novo* SOX2 expression significantly correlated with worse patient outcome ($P<0.01$). For further validation of SOX2 as a predictive biomarker of resistance to therapy, we tested the viability of a CRC cell line treated with 5-FU, after up- and down-regulation of SOX2 expression but unfortunately, we did not observe significant differences. Despite this, previous results from our research group in a different intestinal cellular model had already suggested that 5-FU therapy resistance is at least partially mediated by SOX2. For that reason, our observations in the retrospective study deserve to be further exploited in the near future.

Resumo

O cancro colorectal permanece uma enorme preocupação de saúde, sendo o terceiro cancro mais diagnosticado e o quarto responsável pelo maior número de mortes por cancro na Europa. Em Portugal, o cancro colorectal é simultaneamente o mais diagnosticado e o mais letal. A incidência desta doença aumentou nos últimos 30 anos e espera-se que aumente cerca de 60%, com mais de 2,2 milhões de novos casos e 1,1 milhões de mortes em todo o mundo até 2030.

Apesar dos avanços no diagnóstico e no tratamento, cerca de 20% dos pacientes, diagnosticados com cancro colorectal no estadio II, apesar de sujeitos a cirurgia com intenção curativa, relapsam. Atualmente, a decisão sobre a administração de quimioterapia adjuvante após a cirurgia não é nem consensual, nem eficaz, e baseia-se sobretudo no estadiamento da doença. Seria uma mais valia a existência de biomarcadores de prognóstico para estratificar os pacientes e ajudar a decidir quem deve ser tratado, e biomarcadores preditivos, que previssem o benefício efetivo da administração de determinado quimioterápico.

Com este projeto o nosso objetivo é ampliar o conhecimento científico atual relativamente a potenciais marcadores moleculares que consigam estratificar os pacientes e antecipar quais irão efetivamente beneficiar da quimioterapia adjuvante atual. Para isso investigámos, por imunohistoquímica, na nossa série de 230 pacientes com cancro colorectal no estadio II, o potencial biomarcador de 3 fatores de transcrição - CDX2, SOX9 e SOX2 e da instabilidade de microssatélites, por análise de expressão das proteínas de *mismatch repair*. Para além disso pesquisámos a presença da mutação $BRAF^{V600E}$ por sequenciação de Sanger.

Na nossa série de carcinomas colorectais, a baixa expressão de CDX2 e a expressão *de novo* de SOX2 correlacionaram significativamente com uma menor diferenciação do tumor. A expressão de SOX2 demonstrou valor prognóstico apenas para os casos negativos para o SOX9. Nestes casos, a expressão de SOX2 resultou numa maior probabilidade de recidiva. Para além disso, em pacientes tratados com quimioterapia adjuvante, a expressão *de novo* de SOX2 correlacionou significativamente com um pior prognóstico dos doentes ($p<0.01$). Para validar o SOX2 como biomarcador preditivo de resposta à terapia, testámos a viabilidade de uma linha celular de carcinoma colorectal após silenciamento e sobre-expressão de SOX2, mas infelizmente não se observaram diferenças significativas. No entanto, resultados obtidos anteriormente pelo nosso grupo de investigação num modelo celular intestinal diferente já tinham sugerido que a

resistência ao 5-FU seria, pelo menos parcialmente, mediada pelo SOX2. Por essa razão, as nossas observações neste estudo retrospetivo merecem continuar a ser investigadas num futuro próximo.

Table of contents

List of Figures

List of Tables

List of Abbreviations

18S: 18S ribosomal RNA

5-FU: fluorouracil

ACT: adjuvant chemotherapy

APC: adenomatous polyposis coli gene

BCA: bicinchoninic acid

BRAF: B-Raf proto-oncogene, serine/threonine kinase

cDNA: complementary DNA

CDX2: caudal type homeobox 2 transcription factor

CIMP: CpG Island methylator phenotype

CIN: Chromosomal Instability

CRC: colorectal cancer

Ct: threshold cycle

DAB: 3,3′-diaminobenzidine

DFS: disease-free survival

DMEM: dulbecco's modified eagle's medium

dMMR: deficient mismatch repair

DMSO: Dimethyl sulfoxide

DNA: deoxyribonucleic acid

ECL: enhanced chemiluminescence

EDTA: ethylenediaminetetraacetic acid

Er2: epitope retrieval solution 2 BOND

FACS: fluorescence-activated cell sorting

FBS: fetal bovine serum

FFPE: formalin-fixed, paraffin-embedded

GADPH: glyceraldehyde-3-phosphate dehydrogenase

HE: haematoxylin and eosin

IC_{50}: half maximal inhibitory concentration

IgG: immunoglobulin G

KRAS: KRAS proto-oncogene, GTPase

MAPK: mitogen-activated protein kinase

MLH1: mutL homolog 1

MMR: mismatch repair genes

mRNA: messenger RNA

MSH2: mutS homolog 2

MSH6: mutS homolog 6

MSI: microsatellite instability

MSI-H: microsatellite instable high

MSI-L: microsatellite instable low

MSS: microsatellite stable

MUTYH: mutY DNA glycosylase gene

Na_3VO_4: sodium orthovanadate

NaCl: sodium chloride

ND: not defined

NP-40: Nonidet P-40

NS: not significant

OS: overall survival

PB: Presto Blue

PBS: phosphate-buffered saline

PCR: polymerase chain reaction

PI: propidium iodide

PMS2: PMS1 homolog 2, mismatch repair system component

PMSF: phenylmethylsulfonyl fluoride

RIPA: radioimmunoprecipitation assay

RNA: ribonucleic acid

SC: scrambled

SD: standard deviation

SDS: sodium dodecyl sulfate

siRNA: small interfering RNA

SOX2: Sex-determining region Y-box 2

SOX9: Sex-determining region Y-box 9

SRB: sulforhodamine

TBS: Tris-buffered saline

TCA: trichoroacetic acid

TMA: tissue microarray

TNM: tumor-node-metastasis

Tris-HCl: tris(hydroxymethyl)aminomethane

WT: wild-type

I. Introduction

1.1. Colorectal Cancer

1.1.1. Incidence and mortality

Colorectal cancer (CRC) remains a serious health concern, being the third most commonly diagnosed cancer and the fourth leading cause of cancer-related deaths worldwide (see **Figure 1**) (Ferlay *et al.*, 2013). In Portugal, CRC is both the most frequently diagnosed and the most lethal (see **Figure 2**) (Ferlay *et al.*, 2013).

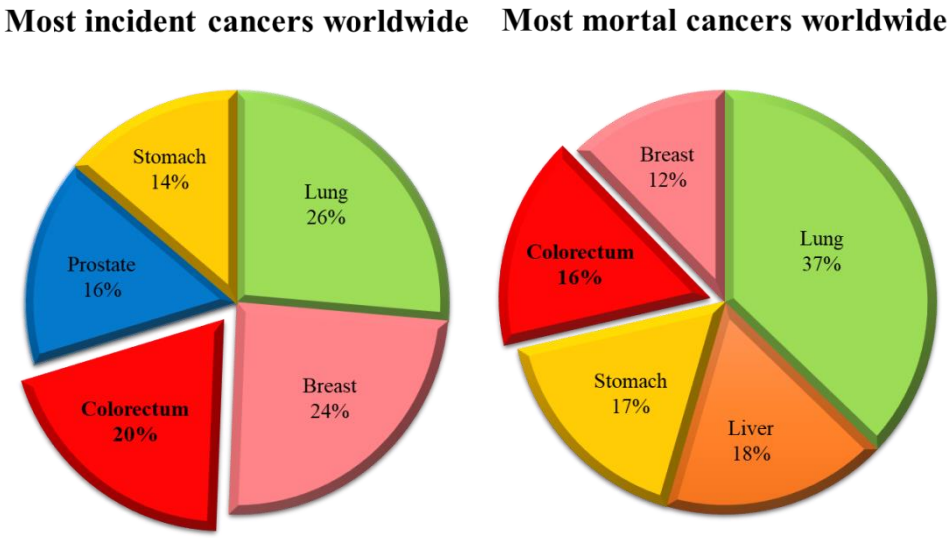


Figure 1. Five most incident and most mortal cancers worldwide; both sexes. Data from the GLOBOCAN 2012 (Ferlay *et al.*, 2013).

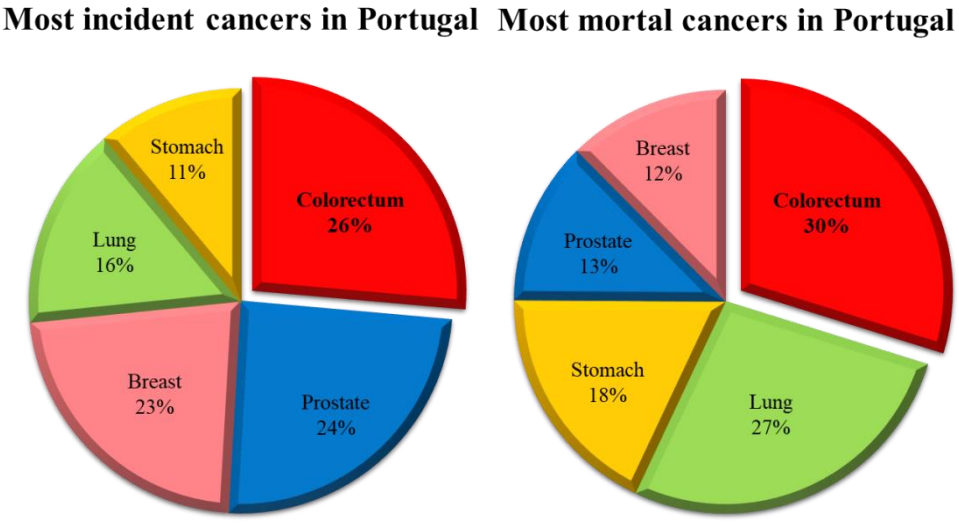


Figure 2. Five most incident and most mortal cancers in Portugal; both sexes. Data from the GLOBOCAN 2012 (Ferlay *et al.*, 2013).

The incidence of CRC increased in the last thirty years, strongly linked to changes in lifestyle and increased exposure to carcinogens, and it is expected to rise by 60% to more than 2.2 million new cases and 1.1 million cancer deaths by 2030 (Ferlay *et al.*, 2013). The American Cancer Society recommend that people at average risk of CRC start regular screening at the age of 45. For CRC screening, people are considered to be at average risk if they do not have: family or personal history of CRC or else confirmed or suspected hereditary CRC syndromes; personal history of adenomatous polyps, inflammatory bowel disease or previous treatment with radiation to the abdomen or pelvic area to treat a prior cancer.

1.1.2. Colorectal cancer aetiology

The majority of CRC, approximately three-quarters, are sporadic (Kuipers *et al.*, 2015). The most common syndrome of hereditary CRC is the Lynch syndrome which is caused by a mutation in one of the DNA mismatch repair genes. The second most common CRC syndrome is familial adenomatous polyposis which is caused by mutations in the adenomatous polyposis coli (APC) gene, known to control the activity of the Wnt/β-catenin signalling pathway. Other hereditary syndromes include polyposis associated with mutations in the mutY DNA glycosylase (MUTYH) gene, Peutz-Jeghers syndrome, serrated polyposis and juvenile polyposis (Kuipers *et al.*, 2015; Vasen *et al.*, 2015).

The environmental and genetic factors that cause CRC promote genomic instability, which is characterized by various genetic and epigenetic alterations leading to stimulation of oncogenes and inactivation of tumor suppressor genes (Cuyle *et al.*, 2017). Accumulation of mutations in critical genes involved in regulating cellular proliferation, differentiation and death provide neoplastic cells with a survival advantage over the surrounding normal intestinal epithelium (Punt *et al.*, 2017) causing abnormal expansion of the malignant tissue into high grade dysplasia adenomatous polyps which have full potential to transform into invasive carcinomas with additional genetic aberrations (Markowitz *et al.*, 2009; Punt *et al.*, 2017).

This traditional adenoma-carcinoma sequence (depicted in **Figure 3**) is thought to be responsible for up to 60% of sporadic CRC (Punt *et al.*, 2017; Tong *et al.*, 2017) and emerging evidence supports that some may evolve via alternate pathways, namely

from serrated polyps (Hardiman 2018). Serrated polyps represent 5 to 10% of all polyps. These arise by molecular and histological events distinct from classical tubular adenomas. These polyps have the potential to transform through the following sequence: hyperplastic polyp to sessile serrated polyp to adenocarcinoma (Rex *et al.*, 2012; Kuipers *et al.*, 2015).

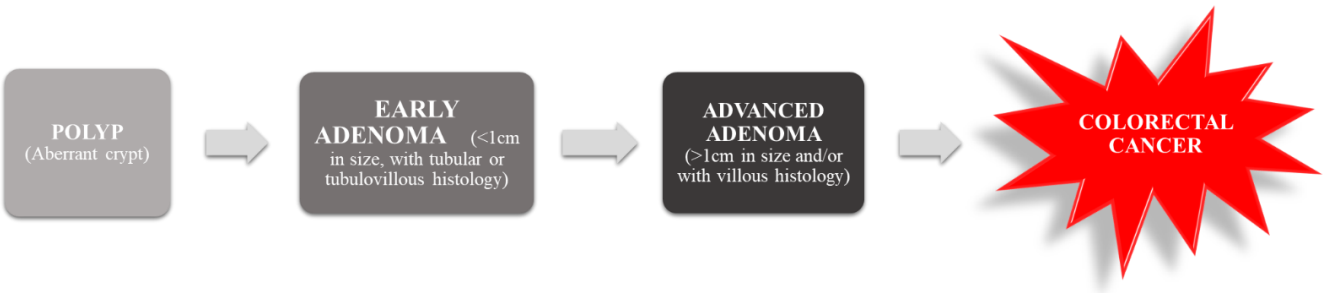


Figure 3. Traditional adenoma to carcinoma sequence.

The order in which mutations accumulate during CRC progression is not random and the association between these and specific histopathological disease stages has been central in CRC research for many years (Punt *et al.*, 2017). Currently, three distinct molecular pathways of CRC pathogenesis are well-described namely:

- Chromosomal Instability (CIN) pathway: the majority of CRCs develop through this pathway; key changes in CIN cancers include widespread alteration in chromosome number and frequent detectable losses at molecular level of portions of chromosomes causing genomic instability.
- Microsatellite Instability (MSI) pathway: subset of 10-15% of CRCs that exhibit aberrations in microsatellite repeat sequences; the carcinoma progression in MSI is faster than in MSS.
- CpG Island methylator phenotype (CIMP) pathway: associated with MSI tumors and *BRAF*^{V600E} mutation; is the pathway of serrated polyposis.

However, many other are under investigation (PDQ Cancer Genetics Editorial Board. Genetics of Colorectal Cancer (PDQ®): Health Professional Version, 2018).

1.1.3. Colorectal cancer staging

Despite the knowledge that CRC is extremely heterogenous, like many other cancers, the tumor-node-metastasis (TNM) staging system (**Figure 4**) is still used to classify CRC.

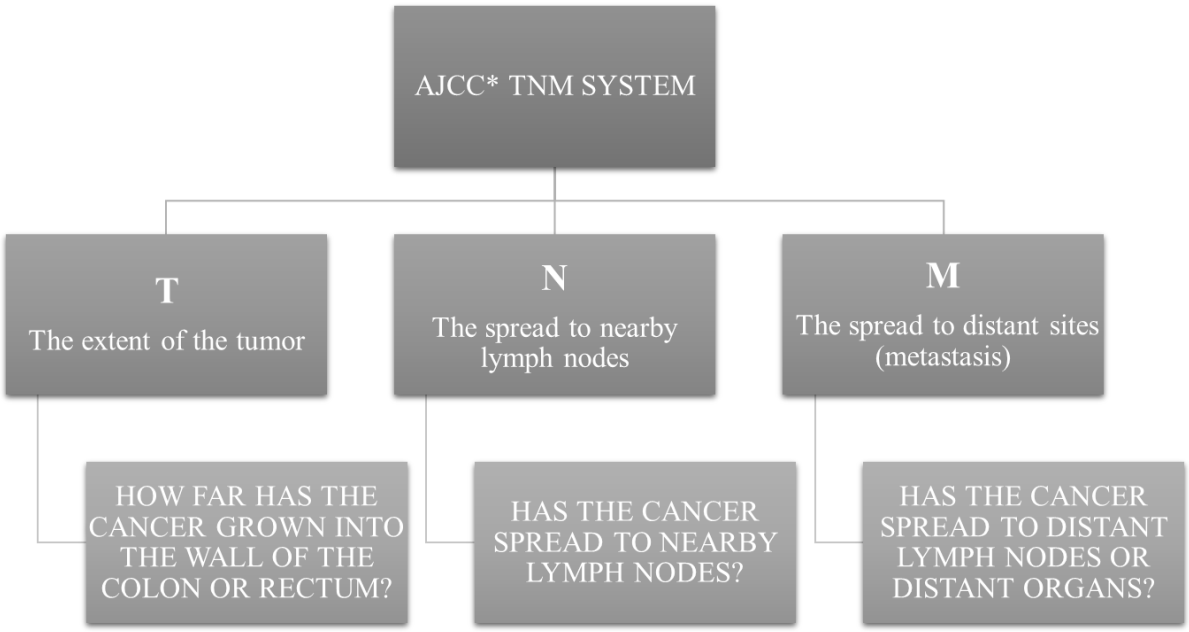


Figure 4. American Joint Committee on Cancer Tumor Node Metastasis staging system. Data from American Cancer Society, 2018.

In a CRC patient, T (extent of the tumor), N (spread to nearby lymph nodes) and M (spread to distant sites or metastasis) categories are usually determined right after resective surgery by examining the resected tissue. This is likely to be more accurate than clinical staging, which considers the results of a physical exam, biopsies and imaging tests done before surgery. Once the values for T, N, and M have been determined, they are combined to assign an overall stage. It is important also to note that TNM stage is determined soon after a cancer is diagnosed and does not change, over time, with disease progression. Although information about the current extent of the cancer is added and of course, the treatment is adjusted as needed, the cancer is always referred to by the stage attributed when diagnosed (American Cancer Society, 2018).

Lymph nodes are small, bean-shaped collections of immune cells. Many types of cancer, including CRC, often spread to nearby lymph nodes before reaching other parts of the body. Although still being firmly debated, the lymph node status is still the strongest predictor and prognosticator in TMN staging system and it is many times used to decide whether to give adjuvant chemotherapy in stage CRC, even though it fails to accurately predict disease recurrence in a considerable number of patients (Veen *et al.*, 2013; Lea *et al.*, 2014; Watson *et al.*, 2016). This happens particularly in Stage II CRC in which there is no consensus about when adjuvant chemotherapy is beneficial and when should it be recommended, leading to an undertreated Stage II subgroup (**Table 1**) which is still to be molecularly defined.

Table 1. American Joint Committee on Cancer characterization of stage II Colorectal Cancer. Data from American Cancer Society, 2018.

	AJCC* Stage	Stage grouping	Stage description
STAGE II	IIA	T3	The cancer has grown into the outermost layers of the colon or rectum but has not gone through them (T3). It has not reached nearby organs. It has not spread to nearby lymph nodes (N0) or to distant sites (M0).
		N0	
		M0	
	IIB	T4a	The cancer has grown through the wall of the colon or rectum but has not grown into other nearby tissues or organs (T4a). It has not yet spread to nearby lymph nodes (N0) or to distant sites (M0).
		N0	
		M0	
	IIC	T4b	The cancer has grown through the wall of the colon or rectum and is attached to or has grown into other nearby tissues or organs (T4b). It has not yet spread to nearby lymph nodes (N0) or to distant sites (M0).
		N0	
		M0	

Sometimes other factors may be considered when staging, such as cell type, tumor location and tumor grade. The grade of a cancer is a measure of differentiation. In low-grade (well-differentiated) cancers, the cancer cells look similar to normal tissue cells and

tend to grow slowly. However, in high grade (poorly-differentiated) cancers, the cancer cells look very abnormal, tend to grow quickly and are often indicative of a worse prognosis. Even when the grade does not affect stage it will probably still affect prognosis and response to therapy (Edge *et al.*, 2010; American Cancer Society, 2018).

1.1.4. Colorectal cancer treatment

For non-metastasized CRC, surgery is the main curative treatment. However, for rectal carcinomas, chemotherapy is recommended in some cases (Kuipers *et al.*, 2015). Since the 1990s, fluorouracil (5-FU) based postoperative chemotherapy, initially fluoropyrimidines and more recently combinations with oxaliplatin (Dienstman *et al.*, 2015), has been used to reduce the risk of tumor recurrence and improve survival for patients with resected CRC cancer. Although disease-free survival (DFS) among patients with stage III CRC has increased significantly owing the introduction of new adjuvant chemotherapy regimens, this has not been extended to earlier stage CRC patients (O'Connor *et al.*, 2011, Dalerba *et al.*, 2016). Despite advances in diagnosis and treatment, about 15% to 30% of patients with stage II disease develop recurrent loco-regional disease or distant metastases within 5 years and their overall survival (OS) is around 70% (see **Figure 5** and **Figure 6**) (O'Connel *et al.*, 2004; Punt *et al.*, 2017).

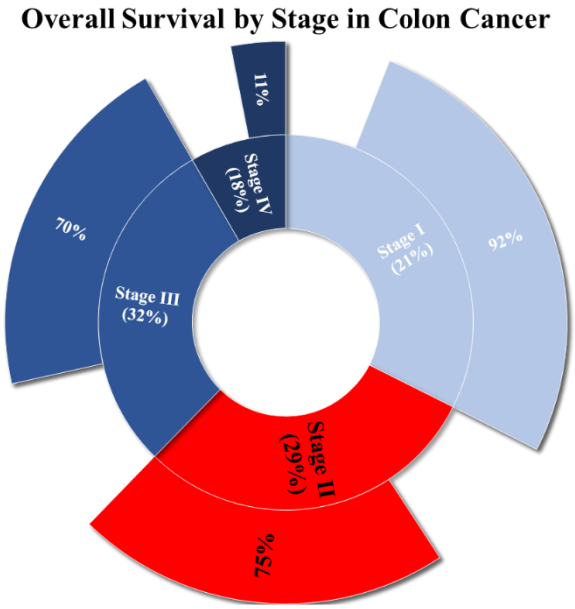


Figure 5. Overall Survival by Stage in Colon Cancer. Data from American Cancer Society, 2018.

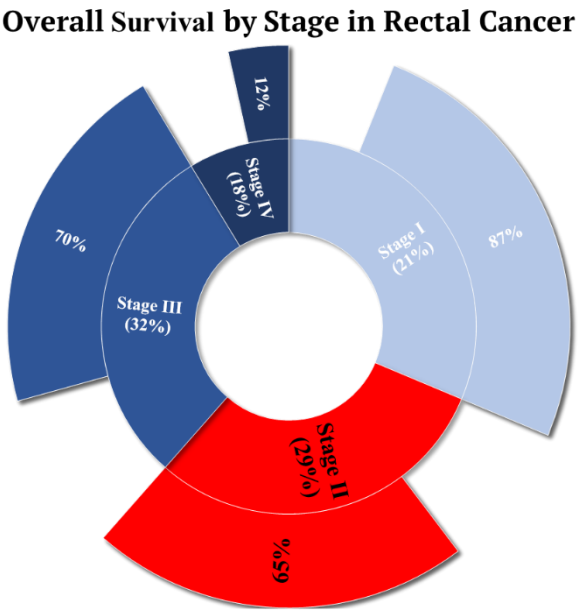


Figure 6. Overall Survival by Stage in Rectal Cancer. Data from American Cancer Society, 2018.

Post-operative treatment is always recommended for stage III CRC patients, however, the absolute survival benefit in stage II disease is only about 2% to 5% with a single 5-FU agent, remaining under debate if these patients have sufficient benefit from adjuvant chemotherapy to justify its associated toxicity, inconvenience and costs (Dienstman *et al.*, 2015). Currently, the decision of giving adjuvant treatment in stage II CRC is recommended to be discussed with patients with one or more high risk features: T4 primary tumors; poorly differentiated histology (except if associated with mismatch repair deficient tumors; presence of lymphovascular and/or perineural invasion; perforation and/or obstruction; close, indeterminate or positive resection margins or less than 12 lymph nodes in the surgical resection specimen. These patients are currently selected based only in histopathologic characteristics and are believed to derive more benefit from adjuvant therapy than patients at low risk of relapse (Cuyle *et al.*, 2017).

1.1.5. Stage II colorectal cancer

Stage II represents almost a third of all CRC patients (see **Figure 7**). Unfortunately, despite being an early stage, in which patients are expected to have better outcomes, stage II patient OS is relatively low and too close to stage III, emphasizing the need to better understand this subset of patients so we can enlarge their quality of life. Interestingly, in rectal cancer, OS for stage II is even lower than stage III (see **Figure 6**).

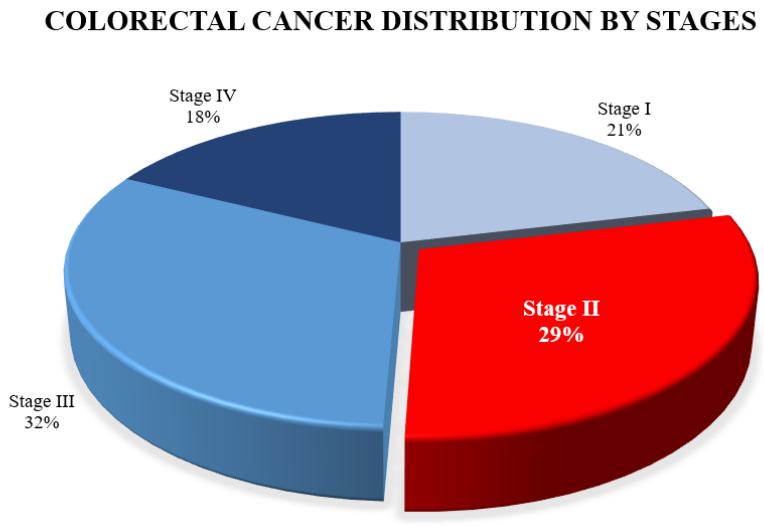


Figure 7. Colorectal Cancer distribution by stages. Data from American Cancer Society, 2018.

Moreover, the fact that adjuvant therapy regimens were shown to improve too little the outcome of CRC stage II patients, plus that chemotherapy is administered in nearly 20% to 59% for stage II CRC patients (Grant *et al.*, 2018) we are led to the assumption that there is an extensive risk of non-effective overtreatment. In fact, in the QUASAR clinical trial, stage II patients were randomized to either 5-FU based therapy or observation and the results demonstrated only 3% improvement in outcome for the 5-FU-treated patients (Marshall 2010; Watson *et al.*, 2016).

Furthermore, even current guidelines fail alone to accurately subclassify which patients will effectively benefit from adjuvant chemotherapy (O'Connor *et al.*, 2011; Dienstmann *et al.*, 2015; Watson *et al.*, 2016). Actually, it has been shown that stage II patients receiving adjuvant treatment with any high-risk variable did not have better OS than patients with any high-risk variable not receiving adjuvant chemotherapy (O'Connor *et al.*, 2011).

It is obvious the urgent need of validated biomarkers to help overcome the current challenge in clinical decision management in stage II CRC. The goal of personalized medicine is to provide each patient, based on his individual biomarker characteristics beyond the functional diagnosis of his disease, with the right treatment and dose at the right time. Therefore, it would be desirable to have both prognostic biomarkers to help stratify patients and thus help to identify who should be treated, and predictive biomarkers

that foretold the likelihood of the benefit of administering a specific chemotherapeutic drug to cancer patients. In the last years there has been intense investigations for new biomarkers. The expectations regarding its potential to upgrade personalized medicine are incredibly high, however the pace has been much slower than hoped.

1.1.6. Biomarkers in colorectal cancer

1.1.6.1. Microsatellite instability (MSI)

Microsatellite instability (MSI) is a molecular marker of deficient mismatch repair (dMMR). Mutations in MMR genes are very common in cancer and lead to a deficient elimination of single-base nucleotide sequences (called the "microsatellites") that occur during DNA replication. This allows the accumulation of somatic mutations, therefore inducing genomic instability that contributes to tumor development, growth and dissemination (Cuyle *et al.*, 2017).

The dMMR can be investigated by testing the loss of expression of at least one MMR protein by immunohistochemistry, or analysing mutations in genes using a PCR-based assay with a panel of microsatellite markers. Both tests were shown to be highly concordant (Sinicrope *et al.*, 2010; Garber 2017). However, neither test is completely sensitive, because an MMR gene may be qualitatively, but not quantitatively, changed by mutation and thus stain even though it is functionally disabled, while PCR testing for microsatellites may miss the defect because the signal is absent due to intratumoral heterogeneity (Garber 2017; Punt *et al.*, 2017).

MSI-High (MSI-H) is a hallmark of Lynch syndrome that, although highly penetrant, accounts for less than 5% of all CRCs. MMR testing is part of the guidelines to identify the patients which are likely to carry a germline mutation in one of the known MMR genes, most commonly *MLH1* or *MSH2*. The majority of MSI-H CRCs, however, are sporadic non-Lynch syndrome cases that result from epigenetic inactivation of the *MLH1* gene promoter by DNA hypermethylation (Sinicrope *et al.*, 2010*).*

MSI prevalence is higher in earlier stages of CRC and decreases in advanced disease, being two times higher in stage II than stage III CRC (Mouradov *et al.*, 2013). Moreover, MSI tumors are more common in elderly patients and in the colon when compared to rectum, being particularly frequent in right-sided colon primary tumors (Grant *et al.*, 2018). In fact, Sinicrope *et al.*, 2010 found positive prognostic relevance of MSI only for right-sided stage II CRC.

In stage II CRC, MSI is found in nearly 20% of patients and it has been strongly associated with better OS when compared to MSS (microsatellite stable) patients (Merok *et al.,* 2013; de Cuba *et al.,* 2016). Furthermore, MSI status is frequently associated with BRAF mutation and it is reported to override the poor prognosis associated with this mutation. In fact, patients with tumors MSI and BRAF-mutated have good prognosis.

Other studies have however indicated that MSI CRC patients are less sensitive to 5-FU based chemotherapy and drive no benefit from it (Ribic *et al.,* 2003; Sinicrope *et al.,* 2011; Mouradov *et al.,* 2013; Chen *et al.,* 2016; Puccini *et al.,* 2017). In fact, it is becoming increasingly common to test all newly diagnosed cases of CRC for MSI, as it serves not only as a screening marker for Lynch Syndrome but also because latest guidelines recommend to not administrate adjuvant 5-FU based therapy to MMR-deficient (MSI) patients (Zarkavelis *et al.,* 2017).

Nonetheless, in two recent meta-analysis focused strictly on stage II CRC patients, MSI status was reported to lack significant prognostic and predictive relevance (Gkekas *et al.,* 2017; Romiti *et al.,* 2017).

1.1.6.2. *BRAF*^{V600E} mutation

B-Raf proto-oncogene, serine/threonine kinase (BRAF) is one of the main RAF family genes, a downstream effector molecule of the KRAS proto-oncogene GTPase and it is involved in the progression of several malignancies including CRC (Davies *et al.,* 2002; Molaei *et al.,* 2016).

The V600E missense mutation in *BRAF* accounts for up to 90% of all mutations in human cancers (Vogelaar *et al.,* 2015) and it is responsible for the constitutive activation of the mitogen-activated protein kinase (MAPK) pathway, known to be one of the most critical pathways in the regulation of cancer cell proliferation and survival (Molaei *et al.,* 2016). This oncogenic mutation is the most frequent in *BRAF* gene and consists in a transversion between thymine and adenine within codon 600, leading to a substitution of a valine by glutamic acid at protein level (Davies *et al.,* 2002; Molaei *et al.,* 2016). *BRAF*^{V600E} mutation is considered as a driver in CRC serrated pathway and polyps are considered its precursor lesions, thus defining this mutation as an early event in CRC (Barras *et al.,* 2017; Sanz-Garcia *et al.,* 2017). Despite being highly frequent in serrated CRC tumors, *BRAF*^{V600E} has been shown to inefficiently drive tumorigenesis in mouse models and to trigger stem cell loss (Tong *et al.,* 2017).

There are many evidences in the literature for $BRAF^{V600E}$ as an indicator of poor outcome in CRC, however patient categories included in these studies were very heterogenous and mostly focused on late stage patients. Moreover, in a recent study that grouped differently early- and late-stage CRC patients it was found that only in late-stage tumors the mutated *BRAF* showed a trend to have worse prognosis, when compared with *BRAF* wild-type patients (Chen *et al.,* 2016). Nevertheless, in some studies no prognostic role of $BRAF^{V600E}$ mutation has been demonstrated (French *et al.,* 2008; Mouradov *et al.,* 2013; Andre *et al.,* 2015; Vogelaar *et al.,* 2015). Interestingly, $BRAF^{V600E}$ reported frequencies in CRC patients vary a lot between studies. This may be due to the fact that rectal cancer patients do not usually harbour this mutation and studies that include a higher percentage of rectal cancers may have lower $BRAF^{V600E}$ mutation frequencies (Tamas *et al.,* 2015).

Since *BRAF*-mutated CRC is often associated with other high-risk clinicopathological features such as later age at diagnosis and poorly differentiated tumors, it is very difficult to distinguish the actual prognostic influence of the *BRAF* mutation from the interaction with the other poor prognostic characteristics (Andre *et al.,* 2015; Vogelaar *et al.,* 2015; Cuyle *et al.,* 2017). There is still much controversy in BRAF potential as a biomarker, also due to its paradoxically strong association with MSI, which is an indicator of good prognosis (Mouradov *et al.,* 2013; Sanz-Garcia *et al.,* 2017). $BRAF^{V600E}$ mutation is about 8 times more prevalent in MSI-High as compared with MSS tumors (Dienstmann *et al.,* 2017). It has been suggested has BRAF mutation only predicts outcome in MSI cases. Interestingly, this association is lost in late stage CRC, reinforcing the idea that BRAF mutation has stronger independent prognostic value in later stages of CRC (Chen *et al.,* 2016).

Furthermore, it is important to take into account the relevant molecular heterogeneity between $BRAF^{V600E}$ mutated tumors that might translate into clinical differences, both in terms of prognosis and sensitivity to chemotherapeutic agents (Guinney e*t al.,* 2015; Cuyle *et al.,* 2017; Barras *et al.,* 2017; Cremolini *et al.,* 2015).

1.1.6.3. Caudal type homeobox 2 (CDX2)

The caudal type homeobox 2 transcription factor (CDX2) is a major regulator of intestine-specific genes, essential for the balance between cell differentiation and proliferation (Zheng *et al.,* 2015) and have a central role in the maintenance of the intestinal homeostasis (Lundberg *et al.,* 2016).

CDX2 is expressed at high levels in the normal colorectum epithelium, but loss or decrease of expression is seen in a subset of CRCs (Lundberg *et al.*, 2016) and has been associated with an increased likelihood of aggressive features such as advanced stage and $BRAF^{V600E}$ mutation (Baba *et al.*, 2009; De Sousa e Melo *et al.*, 2013; Bae *et al.*, 2015; Dalerba *et al.*, 2016; Bruun et al, 2018).

CDX2 expression is inversely correlated with stem cell properties (Huang *et al.*, 2017). Accordingly, in CRC, low CDX2 expression has been associated in some studies with poor differentiation and poor prognosis. CDX2 expression is predictive of a better response to adjuvant chemotherapy (Dalerba *et al.*, 2016; Bruun *et al.*, 2018). However, other studies found no prognostic value for CDX2 expression (Olsen *et al.*, 2016; Dawson *et al.*, 2014). There is still much controversial data emphasizing the need of further studies.

In vivo studies showed an important role for CDX2 as a tumor suppressor gene in the formation and the development of tumors, however its antitumor mechanisms remain to be fully elucidated (Takakura *et al.*, 2010; Zheng *et al.*, 2015).

1.1.6.4. Sex-determining region Y-box 2 (SOX2)

Sex-determining region Y-box 2 (SOX2) is a member of the large SOX gene family comprising transcription factors known to be important in the regulation of developmental processes and cell type specification (Sarkar *et al.*, 2013; Lundberg *et al.*, 2014).

In CRC, SOX2 *de novo* expression has been associated with poorly differentiated plus more invasive tumors and poor OS, especially in $BRAF^{V600E}$ mutated cases (Lundberg *et al.*, 2014). However, this prognostic value is stage-dependent and was only observed in subsets of patients who did not receive adjuvant therapy (Du *et al.*, 2016). *In vitro,* SOX2 has previously been shown to be expressed at higher levels in drug-resistant cells when compared to the parental colon cancer cell line (Yang *et al.*, 2013).

The fact that SOX2 expression is correlated with poorly differentiated tumors seems concordant with the fact that SOX2 is a cancer stem cell marker (Ben-Porath *et al.*, 2008; Lundberg *et al.*, 2014). SOX2 plays a critical role is cancer stem cells self-renewal, affecting tumorigenesis, prognosis and chemoresistance and has been shown to induce *in vitro* a cancer stem cell state in CRC (Lundberg *et al.*, 2016).

1.1.6.5. Sex-determining region Y-box 9 (SOX9)

Sex-determining region Y-box 9 (SOX9) is a transcription factor expressed in the highly proliferative compartment of the healthy intestinal epithelium at the bottom of the crypts and in Paneth cells (Blache *et al.,* 2004). Current literature pertaining SOX9 role in CRC is controversial as it is described to behave both as a tumor suppressor and as an oncogene (reviewed in Prévostel *et al.,* 2017). SOX9 was described to be both regulated (Blache *et al.,* 2004) and to inhibit the oncogenic Wnt/β-catenin signalling pathway (Bastide *et al.,* 2007; Topol *et al.*, 2009; Prévostel *et al.,* 2014). Moreover, SOX9 conditional knock-out in the mouse intestinal epithelium resulted in increased proliferation and decreased differentiation (Bastide *et al.,* 2007; Mori-Akiyama *et al.,* 2007), while SOX9 overexpression supressed proliferation in the intestine of a transgenic mouse model (Formeister *et al.,* 2009). Concordantly, SOX9 overexpression in colorectal cancer cells is sufficient to inhibit cell proliferation (Shi *et al.,* 2013; Prévostel *et al.,* 2014) whereas SOX9 knock-down increases proliferation of human colorectal cancer cells (Shi *et al.,* 2013). These results point to a tumor suppressor role of SOX9 in CRC. Yet other studies support an oncogenic role for SOX9 by reporting that SOX9 knock-down results in decrease of proliferation and tumor growth capacity (Matheu *et al.,* 2012; Carrasco-Garcia *et al.,* 2016) whereas SOX9 overexpression increases tumorigenic potential of CRC cells grafted in mice (Lu *et al.,* 2008).

Correspondingly, the relevance of SOX9 expression as a possible prognostic biomarker is still paradoxical. There are reports correlating both high levels (Lu *et al.,* 2008) and low levels (Espersen *et al.,* 2016) of SOX9 with poor prognosis in CRC and a study enrolling a large number of patients that shows no association with prognosis (Bruun *et al.,* 2014).

II. Aim

With this project we intend to improve scientific insight into new molecular parameters that could stratify CRC stage II patients in subgroups with different expected prognosis and response to current, 5-FU based, adjuvant chemotherapy treatments.

For this purpose, we aim to evaluate not only the prognostic relevance but also the predictive value of a panel of molecules of interest in stage II colorectal carcinoma. More specifically, we will analyse microsatellite instability (MSI) and $BRAF^{V600E}$ mutation status and the expression of CDX2, SOX2 and SOX9 transcription factors in a series of 230 Portuguese patients diagnosed with stage II colorectal adenocarcinoma.

Because we found a significant predictive value for SOX2 expression in our retrospective study we further intend to test the viability of a CRC cell line, treated with 5-FU, after up- and down-regulation of SOX2 expression.

The specific objectives of this project were:

- To characterize MSI status and $BRAF^{V600E}$ mutation, and to detect the expression of CDX2, SOX2 and SOX9 proteins in a stage II CRC Portuguese cohort establishing the prevalence of these 5 putative biomarkers in our series.
- To analyse the correlation between our observations and clinicopathological features of the tumors, disease-free survival and overall survival, in order to assess the relevance of these molecules as predictive and/or prognostic CRC biomarkers.
- To modulate SOX2 expression in colon carcinoma cell lines and evaluate their viability after treating them with 5-FU, in order to infer about the value of SOX2 in 5-FU based chemotherapy resistance.

III. Materials and Methods

3.1. Patients

A retrospective study was performed including 230 patients diagnosed with stage II colorectal adenocarcinoma subjected to curative surgery in Centro Hospitalar S. João, Porto, Portugal, between January 2002 and December 2010. More patients were initially enrolled but were excluded, namely if they: 1) were less than 18 years old; 2) were lost to follow-up; 3) were incorrectly staged; 4) died of post-operative complications; 5) had more than one type of cancer; 6) had a relapse during the first 6 months after surgery. Patients for whom clinical information could not be collected were also excluded, as well as patients without or with insufficient tumor tissue available. Clinicopathological features collected included date of diagnosis, age at diagnosis, gender, clinical presentation at diagnosis with obstruction or intestinal perforation, tumor location, TNM stage, histological type, grade of differentiation, number of ganglia removed at surgery, vascular, lymphatic or perineural invasion, administration of adjuvant chemotherapy, chemotherapy scheme, date and site of relapse, date of death and cause of death (colorectal cancer-related or -unrelated). The tumor tissue was previously processed as part of the diagnostic routine following resective surgery. For immunohistochemistry analyses, 2mm diameter cores from the formalin-fixed, paraffin-embedded (FFPE) tissue were organized in tissue microarrays (TMA).

3.1.1. DNA extraction from paraffin-embedded tissues

Haematoxylin and eosin (HE) staining was performed to guarantee that specimens tested contained >20% cancer cells, and areas enriched in malignant cells were identified by a pathologist before DNA extraction. Two 10 μm slides for each sample were deparaffinized and then incubated for 10 minutes in absolute ethanol. Marked areas were macrodissected with a surgical blade from the two tissue sections. Genomic DNA was extracted with Cell Lysis solution (Citomed, Lisbon, Portugal) and digested with proteinase K 20mg/mL (Thermo Fisher Scientific, Massachusetts, USA) overnight at 55ºC with agitation. Proteins were then precipitated with Protein Precipitation solution (Citomed, Lisbon, Portugal) and discarded following centrifugation at 16,000 g for 3 min at 4ºC. Isopropanol and glycogen were added to the genomic fraction in order to precipitate the DNA. After a centrifugation step at 16,000 g for 3 min, supernatant was carefully discarded, and the pellet washed with ethanol. Pellets were rehydrated with autoclaved deionized water and stored at -20 °C until use. The concentration of the

extracted DNA was assessed using a Nano-Drop 1000 instrument (Thermo Fisher Scientific, Massachusetts, USA).

3.1.1.1. Characterization of *BRAF*^{V600E} mutation

Primary tumors were assessed for the presence of the *BRAF*^{V600E} mutation in genomic DNA extracted from the paraffin-embedded tissues. DNA was amplified with the Taq PCR Master Mix Kit (Qiagen, Hilden, Germany) using the forward primer, 5'-GGAAAGCATCTCACCTCATCC-3', and the reverse, 5'-AACTCAGCAGCATCTCAGGGC-3', designed for the exon 15 of the *BRAF* gene (Sigma-Aldrich, Missouri, EUA). Sterilized water was included as template negative control. PCR amplification was performed as following: an initial activation step at 95ºC for 15 min, three denaturation cycles at 95 ºC for 30s, a first 8-cycle stage, including denaturation at 95 ºC for 30 s, annealing with touchdown temperature of 65ºC to 57ºC for 90 s and extension at 72ºC for 1 min, then an additional 32-cycle stage, including denaturation at 95 ºC for 30 s, annealing at 60ºC for 30 s and extension at 72ºC for 1 min, and a final step of extension for 10 min at 72ºC. PCR products were analysed in 2% agarose gel and stained with GelRed (Intron Biotechnology, South Korea) in order to confirm the presence of the expected 200 bp fragment.

PCR products were purified using the ExoSAP-IT Express PCR Product Cleanup reagent (Applied Biosystems, California, EUA) and sequencing reactions were run using the BigDye Terminator v3.1 cycle sequencing Kit (Applied Biosystems, California, EUA) according to manufacturer's instructions. Sequencing reaction products using both forward and reverse primers were purified with Sephadex (GE Healthcare, Illinois, EUA) and mixed with formamide. Sanger sequencing of all PCR products was subsequently conducted on an Applied Biosystems 3500 Genetic Analyzer (Thermo Fisher Scientific, Massachusetts, USA) and sequences were analysed with Applied Biosystems Quality Check software (Thermo Fisher Cloud). Tumors with the *BRAF*^{V600E} mutation were classified as mutant *BRAF* (*versus* wild-type).

3.1.2. Immunohistochemical analysis of CDX2, SOX2 and SOX9 expression on TMA

3µm thick sections of the TMAs were subjected to immunohistochemistry for CDX2, SOX2 and SOX9, following standard methodologies and Camilo *et al.*, 2014.

Briefly, after deparaffination in xylene for 10 min and rehydration, heat-induced epitope retrieval was carried out in an IHC-Tek Epitope Retrieval Steamer Set for 40 min with 10 mM citrate buffer, pH 6.0 (CDX2) or 10 mM, pH 8.0 EDTA (SOX2 and SOX9). Incubation with primary antibodies for CDX2 (1:50 dilution, CDX2-88 clone, Biogenex, California, USA), SOX2 (1:50 dilution, SP76 clone, Cell Marque, California, USA), and SOX9 (1:6000 dilution, AB5535, Millipore, Merck group, Darmstadt, Germany) was performed overnight, at 4°C. Sections were then incubated with a biotin-labelled rabbit anti-mouse secondary antibody, followed by the avidin/biotin-peroxidase detection system (Vectastain ABC kit, Vector Laboratories, California, USA). Detection was done using the Dako REAL™ Envision™ Detection System Peroxidase/DAB+ (DAKO, Glostrup, Denmark) according to the manufacturer's instructions. Detection of expression was performed with 3,3′-diaminobenzidine (DAB) (Sigma-Aldrich, Missouri, USA) and tissue sections were counterstained with Gill's haematoxylin (Leica Microsystems, Bucks, UK), then dehydrated, clarified and mounted. Normal colonic mucosa was used as a positive control for CDX2 and SOX9 expression and normal gastric mucosa for SOX2 expression.

3.1.3. Immunohistochemical analysis of MLH1, MSH2, MSH6 and PMS2 expression on TMA

Immunohistochemistry was carried out using antibodies for MLH1 (1:50 dilution, G168-728 clone, BD Pharmingen, New Jersey, EUA), MSH2 (Pre-diluted, 25D12 clone, Leica Biosystems, Wetzlar, Germany), MSH6 (1:500, PU29 clone, Leica Biosystems, Wetzlar, Germany), and PMS2 (1:100, MOR4G clone, Leica Biosystems, Wetzlar, Germany), and the Leica Polymer Refine Detection kit on a Leica Bond-III Automated immunohistochemistry stainer (Leica Microsystems, Wetzlar, Germany). The antigen retrieval for these four proteins was carried out for 20 min in Er2 (similar to EDTA pH=8.0).

3.2. Functional assays in colon carcinoma cell lines

3.2.1. Cell line culture

Human colorectal carcinoma cell line - SW620 (ATCC) - was grown in Dulbecco's minimal essential medium (DMEM) (Thermo Fisher Scientific, Massachusetts, EUA) supplemented with 10% inactive fetal bovine serum (FBS) (Invitrogen, California, EUA) and 1% antibiotics (100 U/ml penicillin and 100 μg/ml

streptomycin) (Thermo Fisher Scientific, Massachusetts, EUA), and maintained at 37°C in a humidified 5% CO_2 incubator.

3.2.2. Determination of the IC_{50} for 5-FU

First, $2x10^4$ cells from SW620 cell line were seeded in a 96-well plate. For IC_{50} determination, six different 5-FU (Sigma-Aldrich, Germany) concentrations, ranging from 1,06 to 34 μg/mL, in subsequent multiples of 2, were used to treat cells. Cells treated with only DMSO (Sigma-Aldrich, Germany) were used as control for treatment efficacy. After 48h incubation, cells were washed once with PBS 1x and then 50μL of PrestoBlue Viability Reagent 1x (Invitrogen, California, USA) was added in the dark to each well. The plates were incubated at 37ºC for 45 min in a humidified 5% CO_2 incubator. The fluorescence of all wells was read both at 560nm (excitation) and 590nm (emission) for normalization.

3.2.3. Overexpression and knock-down of SOX2

$1x10^5$ cells were seeded in a 24 well plate. 24h later, transient transfection was performed using Lipofectamine 2000 (Thermo Fisher Scientific, Massachusetts, USA) with a mix of three small interfering RNAs (siRNAs) targeting SOX2 in a total concentration of 99.6 nM and with a non-targeting scrambled siRNA (Thermo Fisher Scientific, Massachusetts, USA), used as a negative control (Thermo Fisher Scientific, Massachusetts, USA), in a 1:1 ratio. The siRNAs and the Lipofectamine were diluted in Opti-MEM (Thermo Fisher Scientific, Massachusetts, USA) prior to the utilization. The cells were incubated at 37°C in a CO_2 incubator and the efficiency of the siRNA silencing was evaluated 72h after transfection by real-time PCR and Western-blot.

For SOX2 overexpression, cells were equally seeded in a 24-well plate and after 24h were transfected with 1 μg of a human SOX2 expression vector containing two FLAG tags at the N-terminus or the corresponding empty vector in a ratio of 1:1.5 relatively to Lipofectamine 2000 reagent, previously diluted in Opti-MEM medium (Thermo Fisher Scientific, Massachusetts, EUA). Cells were incubated at 37°C in a CO_2 incubator and the efficiency of the SOX2 overexpression was evaluated 72h after transfection by real-time PCR and Western blot.

3.2.4. 5-FU treatment and viability assays

24h after transfection, cells were treated with a 5-FU dose corresponding to the IC_{50} - 7.7 μg/mL - for SW620.

After 48h, cells were washed once with PBS 1x and then 50uL of PrestoBlue Viability Reagent 1x (Invitrogen, California, EUA), diluted in culture medium, was added in the dark to each well with cells and to another three additional wells with no cells, to use as background subtraction. The reaction was allowed to proceed for 45 min at 37ºC in a CO_2 incubator. The fluorescence of all wells was read both at 560nm (excitation) and 590nm (emission) for normalization.

In addition, Sulforhodamine B (Sigma-Aldrich, Germany) viability assay was performed for comparison. For that, 50μL of PBS 1x were added to each well and then cells were fixed by adding 25 μL of Trichloroacetic Acid (TCA) 50%. Plates were incubated for 1h at 4ºC and then rinsed several times with deionised water and left to air dry.

When completely dried, 50 μL of Sulforhodamine B solution 0.4% was added to each well and cells were stained for 30 min before quickly rinsing the plates 3 times with 1% acetic acid. Plates were left to dry and after no moisture was visible, 100 μL of Tris Base Solution 10 mM were added to each well and the absorbance was measured at a wavelength of 510 nm. The results are expressed as means ± SD of representative triplicates.

3.2.5. Annexin V/PI assay

Apoptosis was evaluated using the Annexin V Apoptosis detection Kit (eBioscience, SanDiego CA, USA). Briefly, cells were trypsinized and resuspended in binding buffer (previously diluted in water, according to manufacturer's instructions). Cells were incubated with Annexin V, in the dark, for 10 minutes at room temperature and then with propidium iodide. Apoptosis was quantified using the flow cytometer BD Accuri C6 and respective software (BD Biosciences, San Jose CA, USA).

3.2.6. Protein extraction and Western-Blot

To analyse SOX2 expression in the SW620 normal and transient transfected cells after 72h, cells were washed with cold PBS buffer and incubated on ice for 30 min with cold RIPA buffer - 50 mM Tris-HCl pH=7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40,

0.1% SDS (Sigma-Aldrich, California, EUA) supplemented with Complete protease inhibitor cocktail (Roche Applied Science, Indianapolis, USA), 1 mM PMSF and 1 mM Na_3VO_4. After the incubation period, cells were scraped, and the mixture was centrifuged at 14000 rpm for 15min at 4°C.

The soluble proteins concentration was estimated using Pierce BCA protein Assay Kit (Thermo Fisher Scientific, Massachusetts, USA) according to manufacturer's instructions, but only 20 µg of total protein extract were used for Western blot analysis.

Proteins were separated in a 10% sodium dodecyl sulphate–polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane (Amersham, GE Healthcare, UK). The membrane was stained with *Ponceau Red* (Sigma), to monitor transfer efficiency. Next, the membrane was incubated for 1 hour with the blocking solution, in this case, 5% non-fat milk in TBS (GRISP, Porto, Portugal) -1% Tween-20 (Sigma-Aldrich, Missouri, EUA), and then blotted overnight at 4°C with the SOX2 primary antibody (1:250 dilution, Cell Marque, SP76 clone, rabbit, 34 kDa) in 5% non-fat milk in TBS-1% Tween-20. On the following day, the membrane was washed three times with TBS-1% Tween-20 to remove the unbound primary antibody. Then, the membrane was blocked for 1 hour with the secondary antibody anti-rabbit IgG (1:2,000 dilution, goat, Santa Cruz Biotechnology) in 5% non-fat milk in TBS-1% Tween-20 and washed three times TBS-1% Tween-20. The signal detection was performed using ECL detection kit (GE Healthcare, Illinois, EUA). The loading control used was GAPDH (glyceraldehyde-3-phosphate dehydrogenase). Membrane was first incubated overnight with the primary antibody for GADPH (1:10,000 dilution, mouse, Santa Cruz Biotechnology), then, after 3 consecutive washes, incubated for 1h with the secondary antibody anti-mouse IgG (1:10,000 dilution, goat, Santa Cruz Biotechnology), then washed again, and finally revealed using ECL detection kit.

3.2.7. RNA extraction and Real-time PCR

Cells from each well were lysed in 300 µL of Lysis Buffer, containing 1/100 β-mercaptoethanol, for approximately 30 minutes until cells were totally detached from the well. Total RNA was extracted using the Purelink RNA Mini Kit (Thermo Fisher Scientific, Massachusetts, USA) according to manufacturer's instructions. This protocol includes a PureLink DNase Treatment, in which DNA is removed from RNA that is bound to a Spin Cartridge. After extraction, the purified RNA was stored at -80°C, until RNA concentration was measured using a Nano-Drop 1000 instrument (Thermo Fisher

Scientific, Massachusetts, USA). RNA was reverse transcribed to cDNA using the SuperScript VILO cDNA Synthesis Kit (Thermo Fisher Scientific, Massachusetts, USA), according to manufacturer's instructions. SOX2 and 18S genes were amplified with SYBR Green (Thermo Fisher Scientific, Massachusetts, USA) using the following primers (SOX2: 5′-AACGGCTCGCCCACCTACAGC-3′, 5′-AGTGGGAGGAAGAGGTAACC-3′; 18S: 5′-CGCGCGCTAGAGGTGAAATTC-3′; 5′-CATTCTTGGCAAATGCTTTCG-3′; Camilo *et al.*, 2012) in a fluorescence reader ABI Prism 7500. 10 ng of cDNA were used as template for each real-time PCR with 10 µL *SYBR Green* and 0.3 µM of specific forward and reverse primers, in a final volume of 20 µL. The following real-time PCR protocol has been applied: a denaturation step at 95 ºC for 10 min, a 40-cycle stage, including denaturation at 95 ºC for 15 s and annealing at 60ºC for 1 min, and a melting curve program (60–95ºC) with continuous fluorescence measurement. The amount of SOX2 mRNA was determined using the threshold cycle (Ct) values and ΔΔCt method (Livak *et al.*, 2001). The levels of 18S were used for normalization of target gene abundance and relative mRNA levels were calculated. Reactions containing water instead of template were included as negative controls. The results are expressed as means ± SD using triplicates.

3.3. Statistical Analysis

This study followed the REMARK guidelines to report biomarkers (McShane *et al.*, 2005). In order to assess the significance (*P*-values) of differences in clinicopathological features across different groups in our cohort, we have used different statistical tests. The t student test was used when comparing with age. The Qui-square (χ^2) test was used for all the other variables. The Kaplan-Meier method was used to generate 5-year DFS and OS plots and its significance was assessed by the log-rank test. Differences were considered statistically significant when *P* value <0.05. Statistical analysis was performed in IBM SPSS Statistics version 24. In the functional assays in colon carcinoma cell lines each experiment was carried out in triplicates at least two times, and data was expressed as means ± SD. Statistical analysis was performed using Student's *t*-test. A $P< 0.05$ was considered as significantly different.

IV. Results

4.1 Cohort characterization

Table 2 shows the clinicopathological features of all 230 patients diagnosed with stage II CRC included in the study.

Table 2. Clinicopathological data for a series of 230 patients with stage II colorectal carcinoma.

	Frequency (n)	*Percentage (%)*
Patients	230	
Age (years)		
Media	68.2 (±11.6)	
Range	23-92	
Gender		
Female	93	40.4
Male	137	59.6
Histopathological grade		
G1	4	1.7
G2	160	69.6
G3	8	3.5
ND	58	25.2
Tumor Location		
Proximal colon	75	32.6
Distal colon	105	45.7
Rectum	47	20.4
ND	3	1.3
Location of Relapse		
Peritoneum	4	1.7
Lung	3	1.3
Liver	20	8.7
Lymph nodes	3	1.3
Local	12	5.2
ND	9	3.9
Chemotherapy		
Neo-Adjuvant		
Yes	13	5.7
No	217	94.3
Adjuvant		
Yes	35	15.2
No	189	82.2
ND	6	2.6
Resection margins		
R0	179	77.8
R1/R2	2	0.9
ND	49	21.3
Microsatellite instability		
MSI	92	40.0
MSS	120	52.2
ND	18	7.8
BRAF ^{V600E}		
Wt	187	81.3
Mut	30	13.0
ND	13	5.7
CDX2		
Yes	33	14.3
No	197	85.7
SOX2		
Yes	36	15.7
No	194	84.3
SOX9		
Yes	205	89.1
No	25	10.9

The median age in this retrospective series was 68.2 years (range, 23-92). Among the 230 patients, 137 (59.6%) were men and 93 (40.4%) were women. Concerning tumor grade, 4 tumors (1.7%) were well-differentiated (G1), 160 (69.6%) moderately differentiated and 8 (3.5%) were poorly differentiated. Seventy-five (32.6%) tumors were located proximally (right colon), 105 (45.7%) were located distally (left colon) and 47 (20.4%) were in the rectum. Most relapses occurred in the liver (8.7%) and most patients (77.8%) had the resection margin free of cancer cells. Regarding treatment, 13 (5.7%) tumors (all rectal) received neo-adjuvant therapy and only 35 (15.2%) patients received adjuvant therapy.

In this patient cohort, 80.8% of patients were alive 5 years after diagnosis and 20.6% of the patients had relapse (see **Figure 8**).

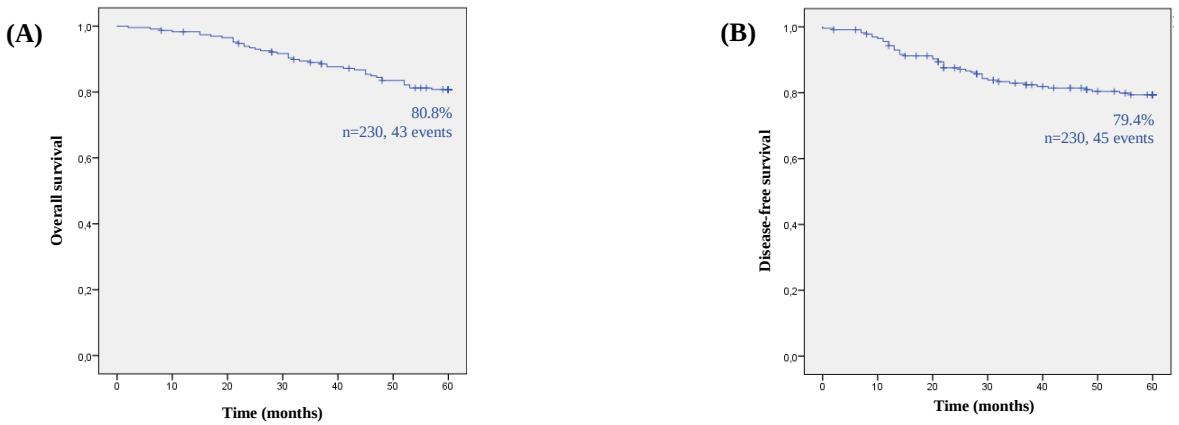


Figure 8. Kaplan-Meyer survival analysis showing OS **(A)** and DFS **(B)** for 230 stage II colorectal cancer patients 5 years after diagnosis.

In this stage II CRC series, adjuvant chemotherapy significantly improves patient OS but not DFS (**Figure 9**).

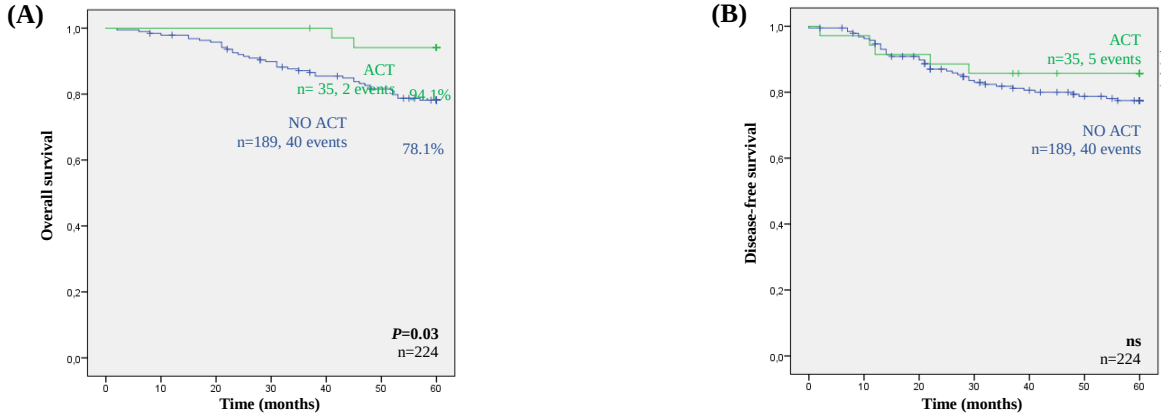


Figure 9. Kaplan-Meyer survival analysis showing OS (A) and DFS (B) for 224 stage II colorectal cancer patients according to different treatment options – treated or not with adjuvant chemotherapy. *Note*. ACT stands for Adjuvant Chemotherapy.

4.2. Evaluation of putative biomarkers of CRC

4.2.1. Microsatellite Instability (MSI)

4.2.1.1 Immunohistochemical analysis of MMR (mismatch repair) protein expression on TMA

Microsatellite instability was analysed using immunohistochemistry of MMR proteins: MLH1 (**Figure 10**), MSH2 (**Figure 11**), MSH6 (**Figure 12**) and PMS2 (**Figure 13**) to infer MSI status of 212 patients. The rest of the cases were unevaluable due to insufficient number of epithelial tumor cells or loss of tissue on the TMA slide.

MLH1 expression was observed in 131/212 (61.8%) tumors and 81/212 (38.2%) showed loss of expression of this protein (**Figure 10**).

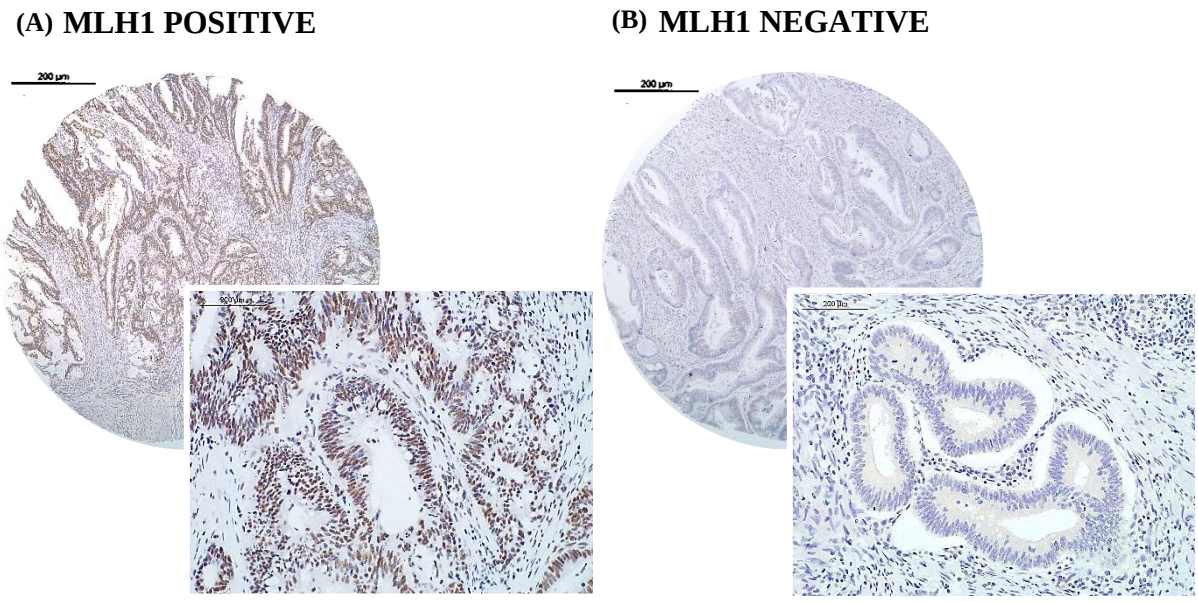


Figure 10. Analysis of MLH1 protein expression in 212 stage II colorectal carcinomas by immunohistochemistry. Positive **(A)** and negative **(B)** stain.

Regarding MSH2 expression, 198/212 (93.4%) patients were positive and 14/212 (6.6%) were negative (**Figure 11**).

(A) MSH2 POSITIVE **(B) MSH2 NEGATIVE**

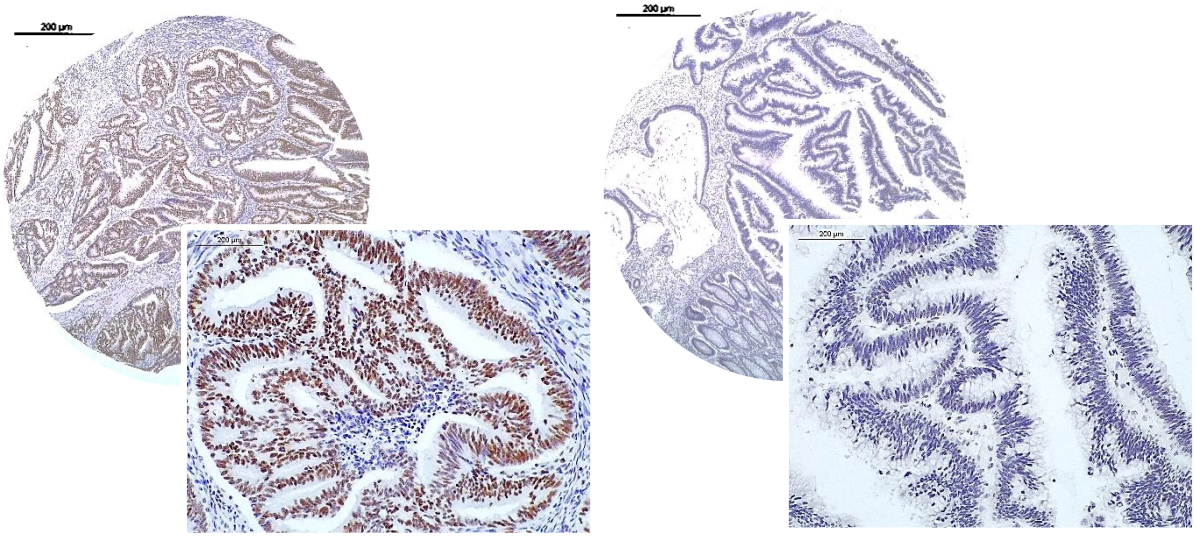


Figure 11. Analysis of MSH2 protein expression in 212 stage II colorectal carcinomas by immunohistochemistry. Positive **(A)** and negative **(B)** stain.

MSH6 expression was found in 170/212 (80.2%) of tumors and absent in the other 42/212 (19.8%) (**Figure 12**).

(A) MSH6 POSITIVE **(B) MSH6 NEGATIVE**

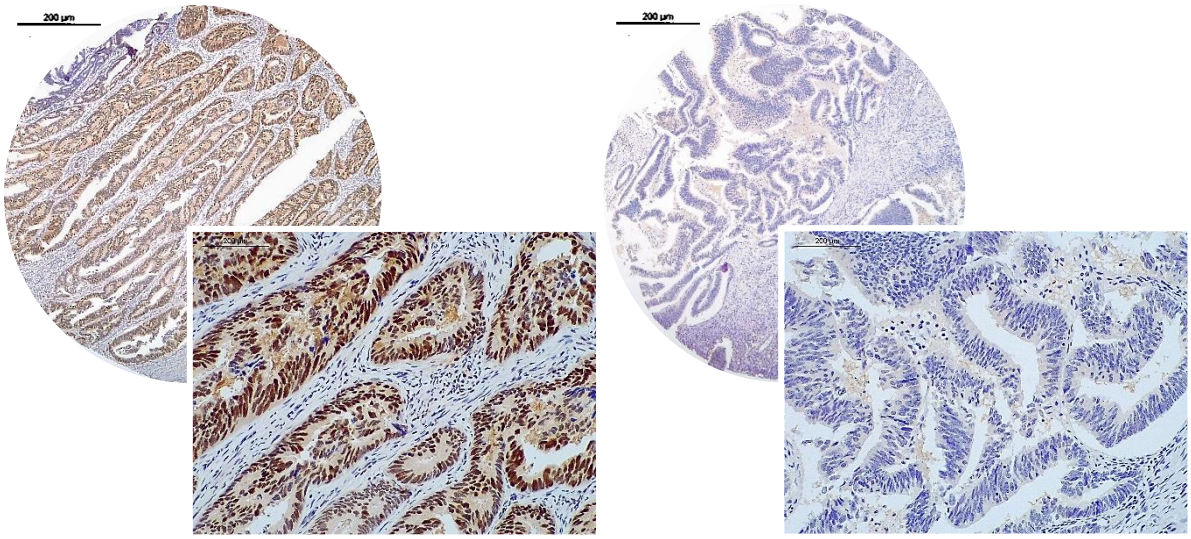


Figure 12. Analysis of MSH6 protein expression in 212 stage II colorectal carcinomas by immunohistochemistry. Positive **(A)** and negative **(B)** stain.

Finally, concerning *PMS2* expression, 174/212 (82.1%) of the tumors were positive whereas 38/212 (17.9%) were negative (**Figure 13**).

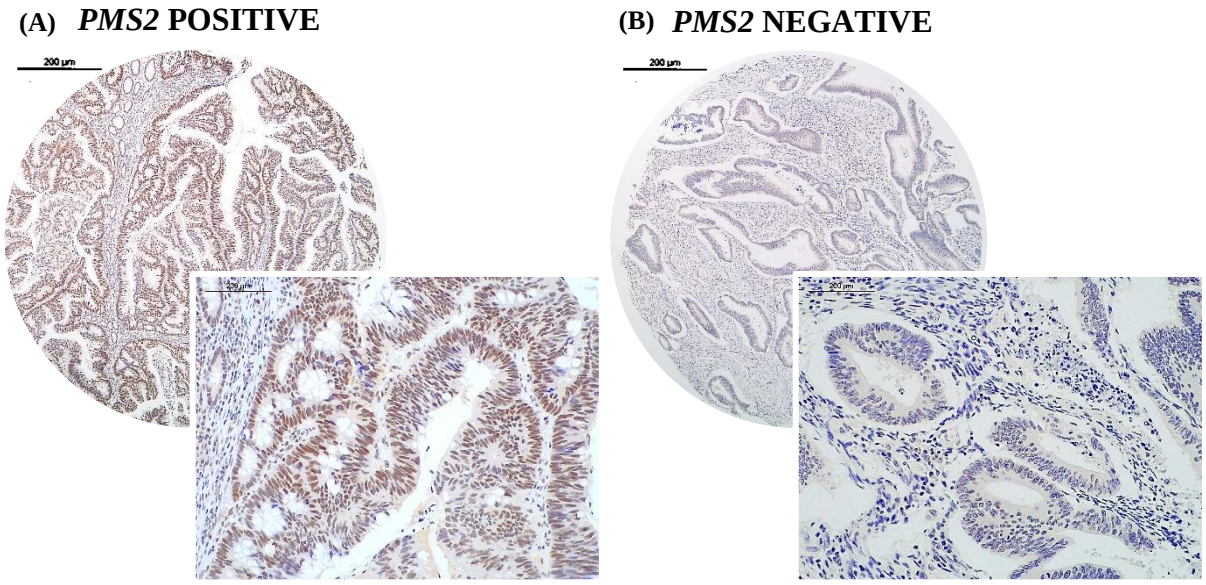


Figure 13. Analysis of *PMS2* protein expression in 212 stage II colorectal carcinomas by immunohistochemistry. Positive **(A)** and negative **(B)** stain.

In summary, 120 (52.2%) patients were MSS based on the fact that they have not lost the expression of any of the four proteins, and 92 (40.0%) patients were MSI, being 39 (17%) MSI-Low since they have lost the expression of only one MMR gene, and 53 (23.0%) were MSI-High since they have lost the expression of at least two MMR genes.

4.2.1.2. Correlation between MSI status and patient outcome

In our patient cohort MSI did not show any relevance as prognostic (see **Figure 14; Supplementary Figure 1**) or predictive (see **Figure 15**) biomarker.

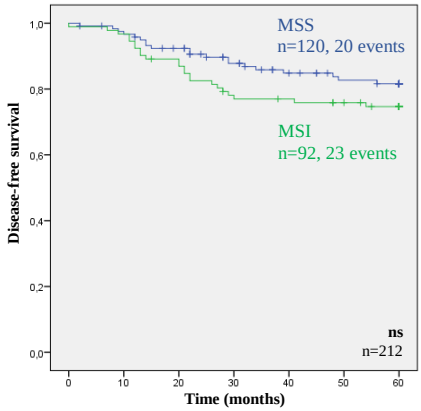


Figure 14. Kaplan-Meyer survival analysis showing the probability of DFS for 212 patients according to MSI status.

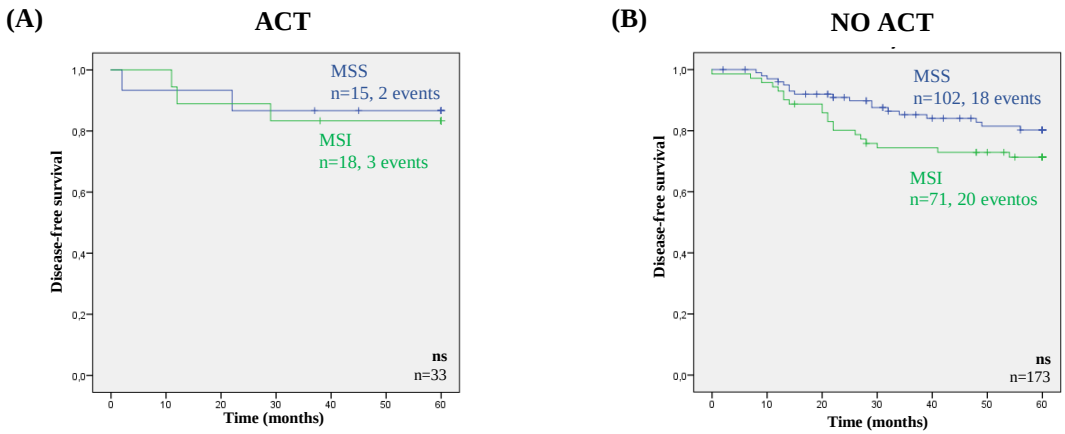


Figure 15. Kaplan-Meyer survival analysis showing the probability of DFS for stage II CRC patients when treated with adjuvant chemotherapy **(A)** or not **(B)** according to MSI status. *Note*. ACT stands for Adjuvant Chemotherapy.

4.2.2. *BRAF*^{V600E} mutation

4.2.2.1. *BRAF*^{V600E} mutation analysis by sequencing

BRAF^{V600E} mutation status was accessed in 217 patients. In the rest of the tumors the material was insufficient for DNA extraction. In our cohort, *BRAF*^{V600E} mutation was present in 30 (13.0%) tumors *versus* 187 (81.3%) that were wild type (see **Figure 16**).

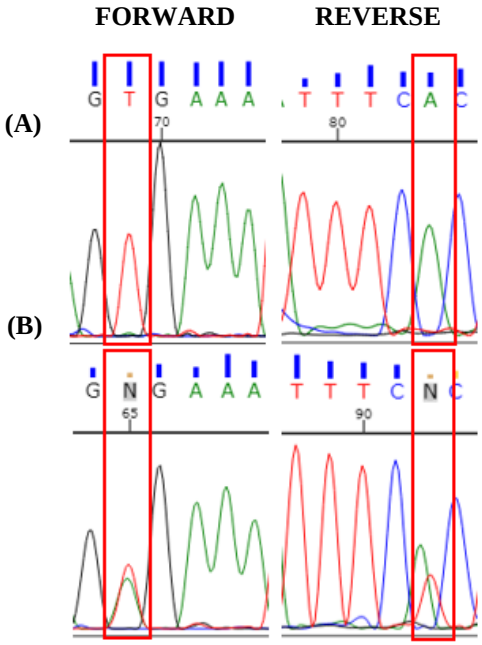


Figure 16. BRAF sequencing electropherograms in respect to *BRAF*^{V600E} mutation: **(A)** wild-type; **(B)** mutated.

4.2.2.2 Correlation between *BRAF*^{V600E} mutation status and patient outcome

In this series, *BRAF*^{V600E} mutation did not show any relevance as prognostic (see **Figure 17; Supplementary Figure 2**) or predictive (see **Figure 18**) biomarker.

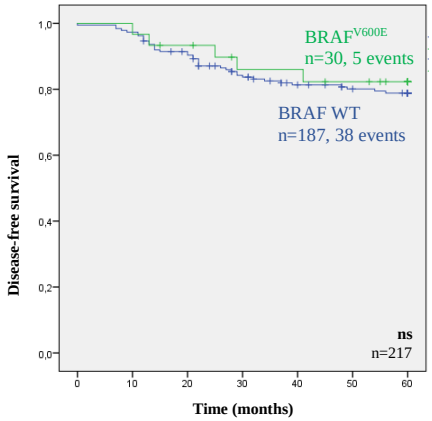


Figure 17. Kaplan-Meyer survival analysis showing the probability of DFS for 217 patients according to *BRAF*^{V600E} status.

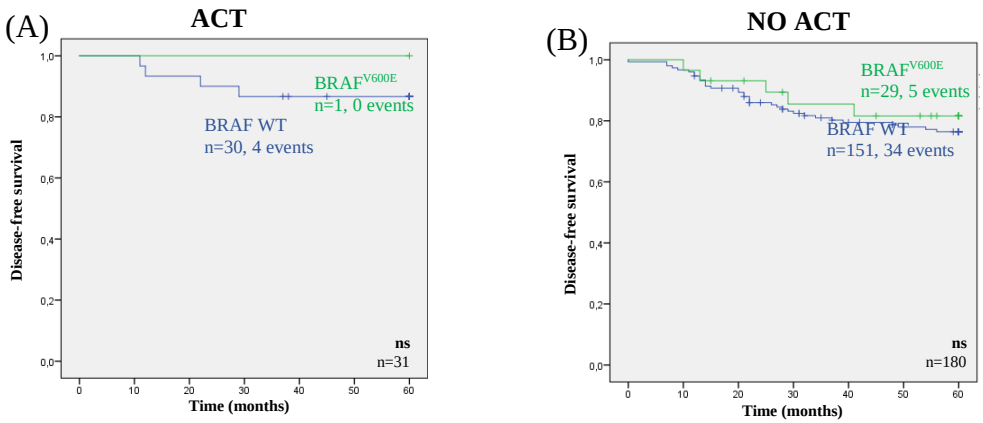


Figure 18. Kaplan-Meyer survival analysis showing the probability of DFS for stage II CRC patients when treated with adjuvant chemotherapy **(A)** or not **(B)** according to BRAF^{V600E} mutation *Notes*. ACT stands for Adjuvant Chemotherapy.

4.2.3. CDX2 expression

4.2.3.1. Immunohistochemical analysis of CDX2 protein expression on TMA

Analysis of CDX2 expression was performed by immunohistochemistry in all 230 tumors. Expression in the normal intestine epithelium is shown in **Figure 19**. CDX2 loss of expression was observed in 33 (14.3%) of patients compared to the other 197 (85.7%) that showed a strong staining (**Figure 20**).

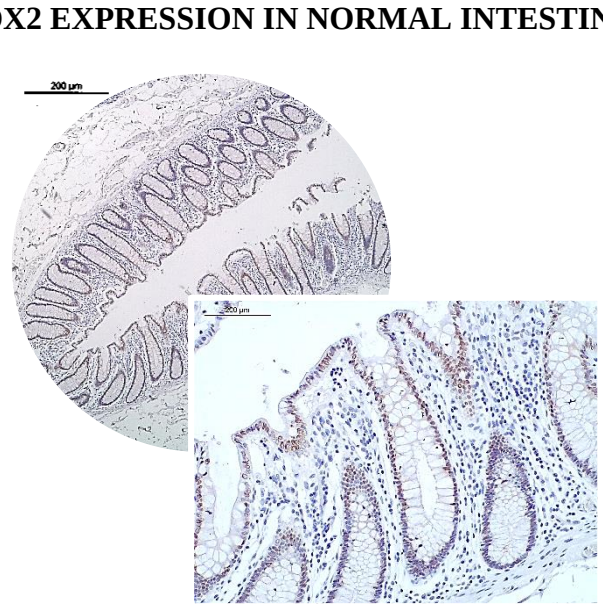


Figure 19. Analysis of CDX2 expression in normal intestine epithelium by immunohistochemistry.

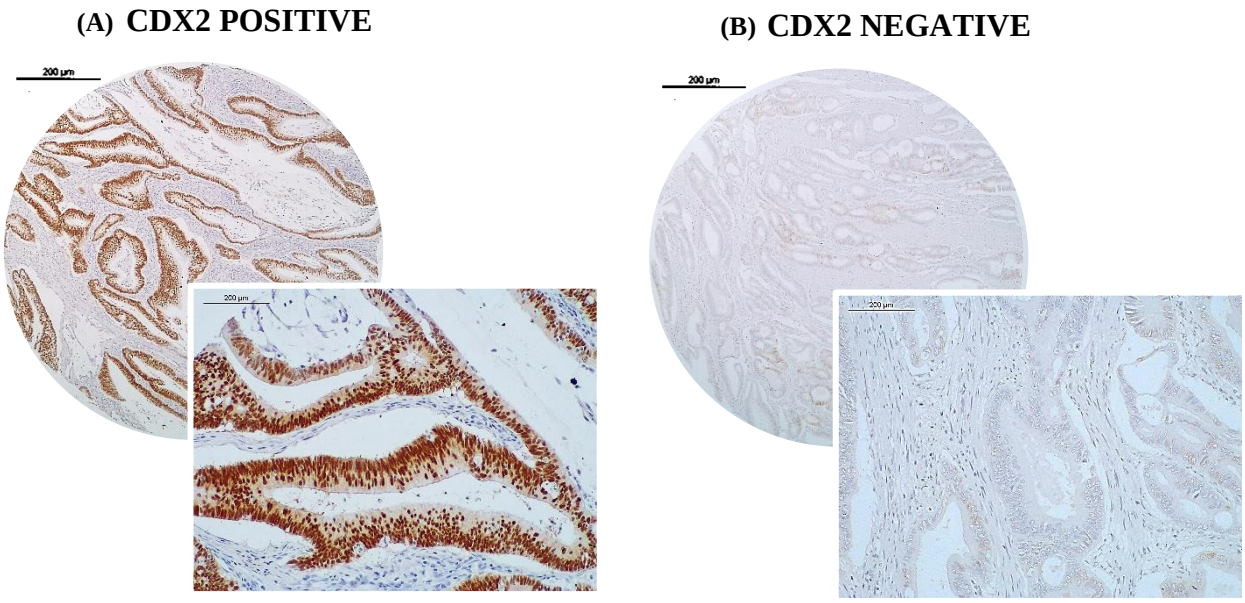


Figure 20. Analysis of CDX2 protein expression in 230 stage II colorectal carcinomas by immunohistochemistry. Positive **(A)** and negative **(B)** stain.

4.2.3.2. Correlation between CDX2 expression and patient outcome

In our patient cohort CDX2 loss of expression did not show any relevance as prognostic (see **Figure 21; Supplementary Figure 3**) or predictive (see **Figure 22**) biomarker.

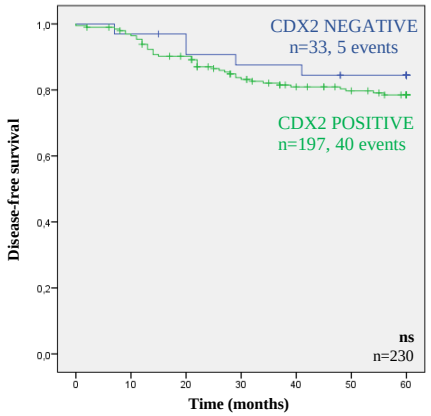


Figure 21. Kaplan-Meyer survival analysis showing the probability of DFS for 230 patients according to CDX2 expression.

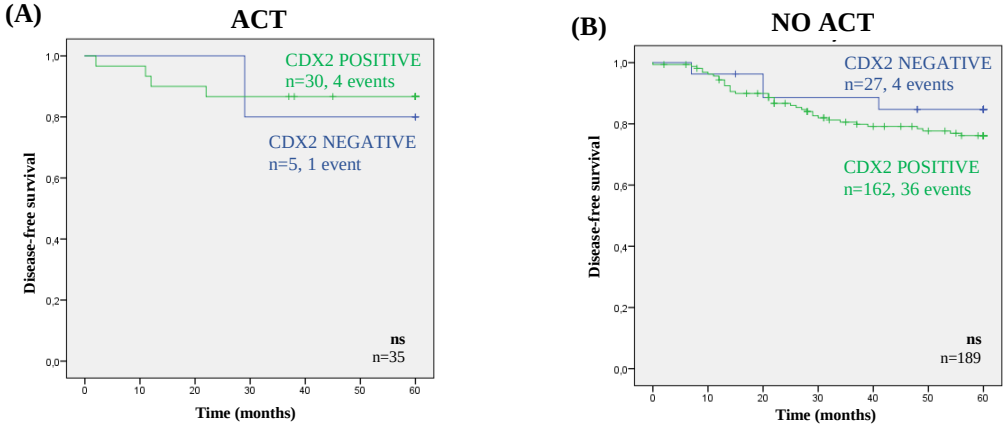


Figure 22. Kaplan-Meyer survival analysis showing the probability of DFS for stage II CRC patients when treated with adjuvant chemotherapy **(A)** or not **(B)** according to CDX2 expression. *Note*. ACT stands for Adjuvant Chemotherapy.

4.2.4. SOX2 expression

4.2.4.1. Immunohistochemical analysis of SOX2 protein expression on TMA

Analysis of SOX2 expression was performed by immunohistochemistry in all 230 tumors. SOX2 is not expressed in the normal intestine epithelium as shown **Figure 23.** SOX2 *de novo* expression was observed in 36 (15.7%) patients compared to the other 194 (84.3%) that showed no staining (see **Figure 24**).

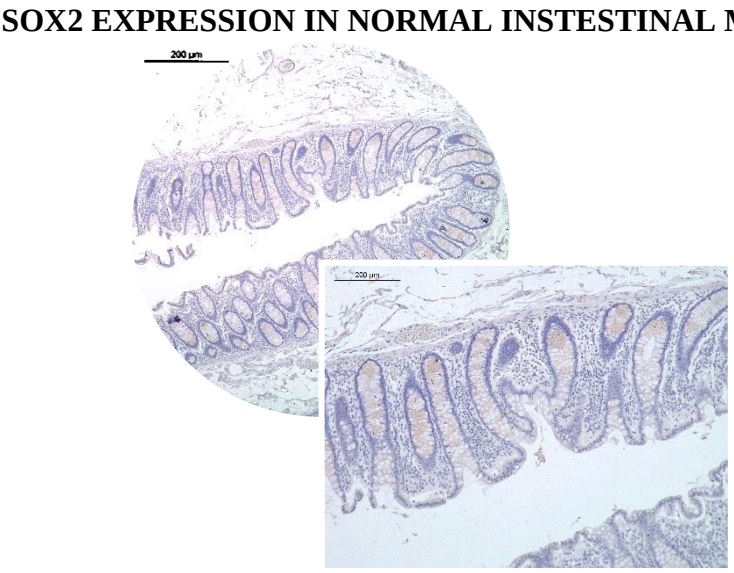


Figure 23. Analysis of SOX2 expression in normal intestine epithelium by immunohistochemistry.

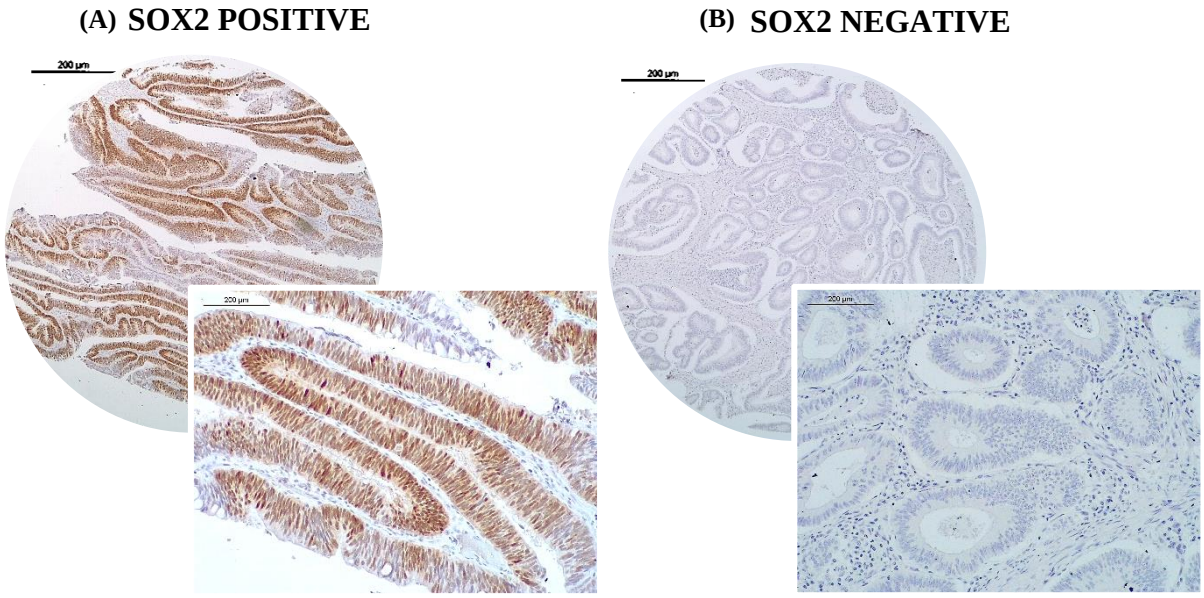


Figure 24. Analysis of SOX2 protein expression in 230 stage II colorectal carcinomas by immunohistochemistry. Positive **(A)** and negative **(B)** stain.

4.2.4.2. Correlation between SOX2 expression and patient outcome

In our patient cohort, SOX2 *de novo* expression did not show any relevance as prognostic biomarker (**Figure 25, Supplementary Figure 4**).

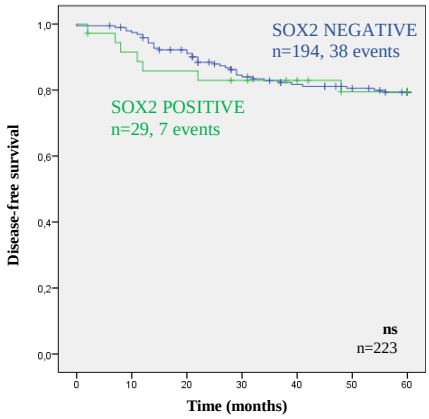


Figure 25. Kaplan-Meyer survival analysis showing the probability of DFS for 230 patients according to SOX2 expression.

However, SOX2 positive patients treated with adjuvant chemotherapy had significantly higher probability of relapse after 5 years than SOX2 negative patients treated with adjuvant chemotherapy (**Figure 26**).

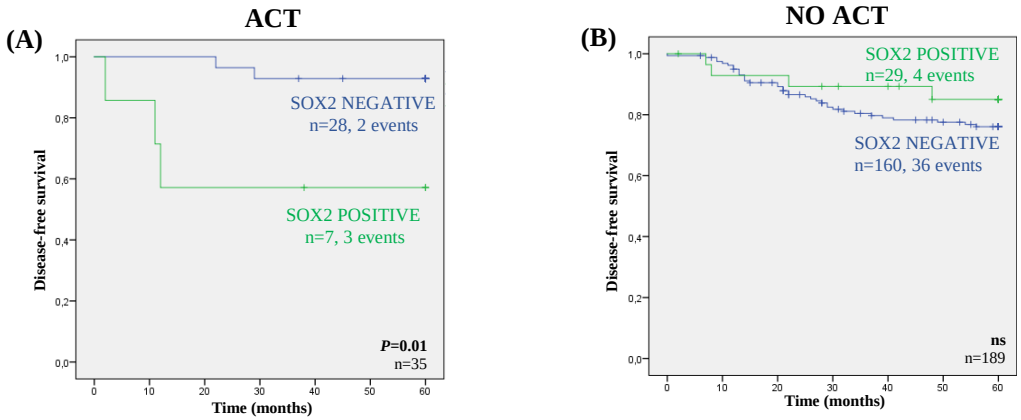


Figure 26. Kaplan-Meyer survival analysis showing the probability of DFS for stage II CRC patients when treated with adjuvant chemotherapy **(A)** or not **(B)** according to SOX2 expression. *Note.* ACT stands for Adjuvant Chemotherapy.

Furthermore, SOX2 positive patients respond poorly to adjuvant chemotherapy having a tendency, borderline significant (*P*=0.06) of lower DFS than SOX2 positive patients not treated with adjuvant treatment (see **Figure 27**). On the other way around, SOX2 negative patients treated with adjuvant chemotherapy have a significantly higher probability of DFS survival that SOX2 negative patients that did not receive adjuvant therapy. (see **Figure 27**).

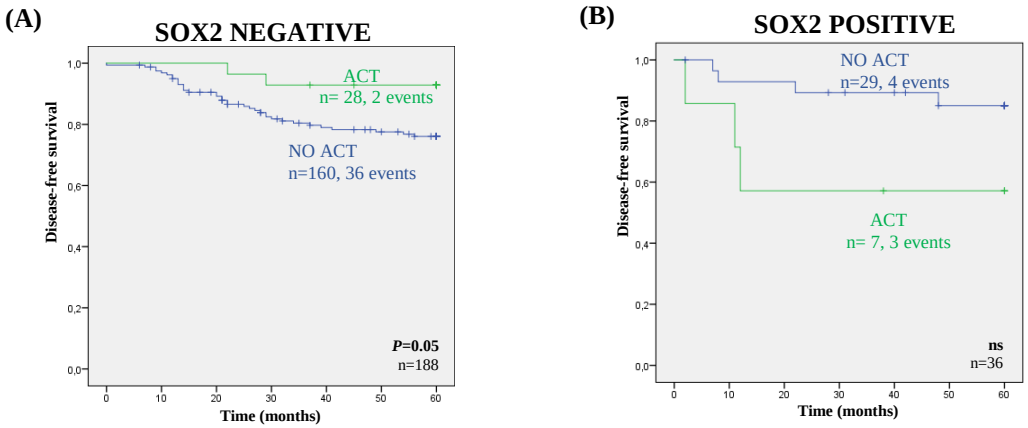


Figure 27. Kaplan-Meyer survival analysis showing the probability of DFS for SOX2 negative **(A)** and SOX2 positive **(B)** stage II CRC patients according to having received or not adjuvant chemotherapy. *Note.* ACT stands for Adjuvant Chemotherapy.

4.2.5. SOX9 expression

4.2.5.1. Immunohistochemical analysis of SOX9 protein expression on TMA

Analysis of SOX9 expression was performed by immunohistochemistry in all 230 tumors. SOX9 expression pattern in the normal intestine epithelium is shown in **Figure 28**. SOX9 loss of expression was observed in 25 (10.9%) patients compared to the other 205 (89.1%) that showed a strong staining (see **Figure 29**).

SOX9 EXPRESSION IN NORMAL INTESTINAL MUCOSA

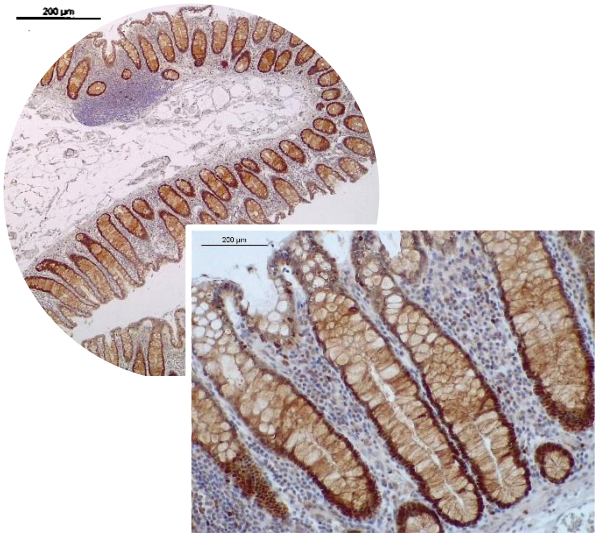


Figure 28. Analysis of SOX9 expression in normal intestine epithelium by immunohistochemistry.

(A) SOX9 POSITIVE **(B) SOX9 NEGATIVE**

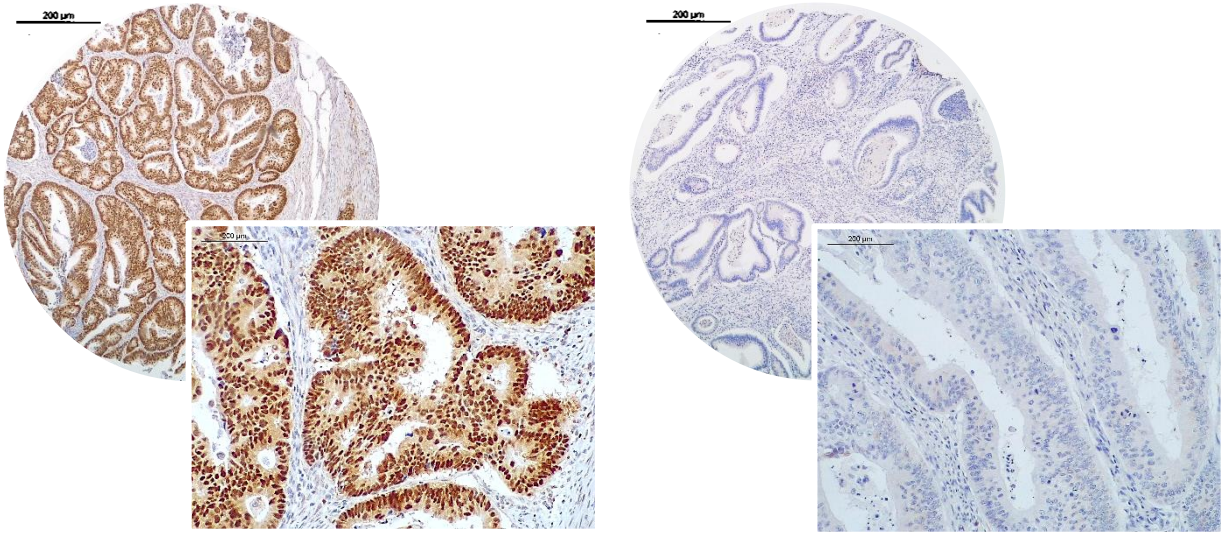


Figure 29. Analysis of SOX9 protein expression in 230 stage II colorectal carcinomas by immunohistochemistry. Positive **(A)** and negative **(B)** stain.

4.2.5.2. Correlation between SOX9 expression and patient outcome

In our patient cohort, SOX9 loss of expression did not show any relevance as a prognostic (see **Figure 30; supplementary Figure 5**) or predictive (see **Figure 31**) biomarker.

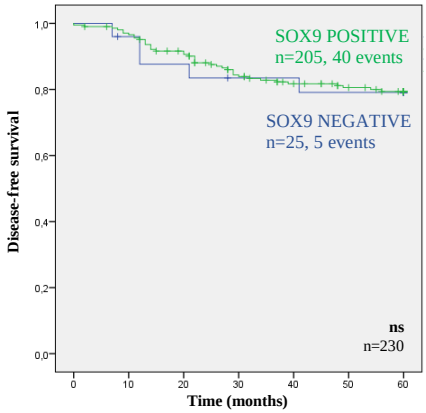


Figure 30. Kaplan-Meyer survival analysis showing the probability of DFS for 230 patients according to SOX9 expression.

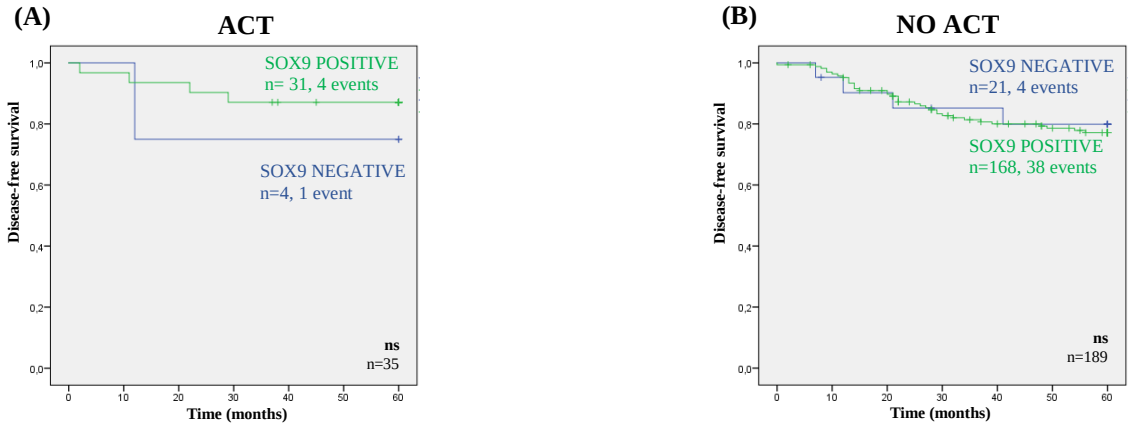


Figure 31. Kaplan-Meyer survival analysis showing the probability of DFS for stage II CRC patients when treated with adjuvant chemotherapy **(A)** or not **(B)** according to SOX9 expression. *Notes*. ACT stands for Adjuvant Chemotherapy.

Interestingly, when analysing the expression of more than one molecule and patient outcome, we noticed that in patients with SOX9 negative tumors, those that are SOX2 positive have a significantly lower DFS when compared to SOX2 negative (see **Figure 32**).

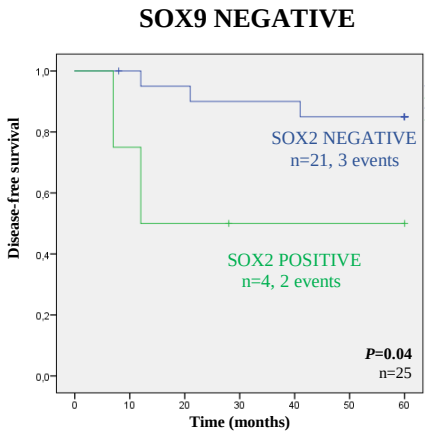


Figure 32. Kaplan-Meyer survival analysis showing the probability of DFS for 25 SOX9 negative patients, according to SOX2 expression.

4.2.6 Clinicopathological data association with the different putative biomarkers

In **Table 3** and **Table 4** we correlated the clinicopathological data for all patients and the putative biomarkers studied.

Table 3. Clinicopathological data association with MSI and *BRAF*^{V600E} status in 230 patients with stage II colorectal carcinoma.

	Microsatellite Instability			**BRAF^{V600E}**		
	MSS(%)	MSI(%)	***P***	WT(%)	MUT(%)	***P***
Age (years)						
Media	69.0±11.3	66.6±12.3	0.13	67.6±11.6	72.9±8.8	**0.02**
Range	35-91	23-92		23-92	55-87	
Gender						
Female	35.8	64.2	0.16	78.2	21.8	**0.01**
Male	60.6	39.4		91.5	8.5	
Histopathological grade						
G1	50.0	50.0	0.29	100.0	0.0	0.18
G2	55.4	44.6		84.4	15.6	
G3	20.0	80.0		62.5	37.5	
Tumor Location						
Proximal colon	50.0	50.0	0.29	77.8	22.2	0.05
Distal colon	58.8	41.2		86.3	13.7	
Rectum	64.4	35.6		100.0	0.0	
Microsatellite instability						
MSI				85.7	14.3	1.00
MSS				85.4	14.6	
***BRAF*^{V600E}**						
Wt	55.8	44.2	1.00			
Mut	55.2	44.8				
CDX2						
Positive	61.7	38.3	**<0.001**	87.6	12.4	0.17
Negative	24.1	75.9		78.1	21.9	
SOX2						
Positive	65.6	34.4	0.33	90.3	9.7	0.58
Negative	55.0	45.0		85.5	14.5	
SOX9						
Positive	58.7	41.3	0.08	85.7	14.3	0.75
Negative	39.1	60.9		90.5	9.5	

Notes. P values (statistical significance threshold<0.05) were obtained using Student's t test for the continuous variable and Qui-square (χ^2) test for categorical variables; comparisons with P<0.05 are indicated in bold face.

Table 4. Clinicopathological data association with CDX2, SOX2 and SOX9 expression in all patients included

	CDX2			**SOX2**			**SOX9**		
	+	-	***P***	+	-	***P***	+	-	***P***
Age (years)									
Media	68.3±11.6	67.7±11.8	0.80	68.9±11.5	68.1±11.6	0.68	68.2±11.3	68.1±13.8	0.96
Range	23-92	35-87		37-87	23-92		23-91	35-92	
Gender									
Female	20.4	79.6	**0.04**	16.1	83.9	0.86	87.1	12.9	0.52
Male	89.8	10.2		15.3	84.7		90.5	9.5	
Histopathological grade									
G1	100.0	0.0	**<0.001**	25.0	75.0	**0.01**	100.0	0.0	0.78
G2	86.2	13.78		12.5	87.5		90.0	10.0	
G3	25.0	75.0		50.0	50.0		87.5	12.5	
Tumor Location									
Proximal colon	78.7	21.3	0.07	17.3	82.7	0.11	85.3	14.7	0.30
Distal colon	86.7	13.3		10.5	89.5		92.4	7.6	
Rectum	93.6	6.4		23.4	76.6		87.2	12.8	
Microsatellite instability									
MSI	76.1	23.9	**<0.001**	12.0	88.0	0.33	84.8	15.2	0.08
MSS	94.2	5.8		17.5	82.5		92.5	7.5	
***BRAF*^{V600E}**									
Wt	86.6	13.4	0.17	15.0	85.0	0.58	89.8	10.2	0.75
Mut	76.7	23.3		10.0	90.0		93.3	6.7	
CDX2									
Positive				15.2	84.8	0.61	91.4	8.6	**0.01**
Negative				18.2	81.8		75.8	24.2	
SOX2									
Positive	83.3	16.7	0.61				88.9	11.1	1.00
Negative	86.1	13.9					89.2	10.8	
SOX9									
Positive	87.8	12.2	**0.01**	15.6	84.4	1.00			
Negative	68.0	32.0		16.0	84.0				

Notes. P values (statistical significance threshold<0.05) were obtained using Student's t test for the continuous variable and Qui-square (χ^2) test for categorical variables; comparisons with P<0.05 are indicated in bold face.

MSI did not correlate with any of the clinicopathological parameters, which included age at diagnosis, gender, histopathological grade and tumor location. However, microsatellite stable tumors were significantly associated with CDX2 positive expression (P<0.001).

BRAF^{V600E} mutation correlated with older age at time of diagnosis. Patients harboring the mutation are diagnosed about 5 years later than patients that are wild type for this mutation (72.9±8.8 *versus* 67.6±11.6; P=0.02). Also, this mutation was significantly more common in female than in male patients (P=0.01) and borderline significant in proximal colon (P=0.05). This mutation was not found in the rectal tumors.

CDX2 loss of expression was more common in females (P=0.04) and it correlated strongly with poorly differentiated tumors (P<0.001). CDX2 expression was not associated with other clinicopathological parameters. Moreover, CDX2 loss of expression was significantly correlated with MSI status and was also significantly associated with SOX9 negative expression.

SOX2 expression was only associated with poorly differentiated tumors (P=0.01).

Finally, SOX9 expression did not correlate with any of the clinicopathological parameters studied. However, it did correlate with CDX2 expression, as previously described.

4.3. Functional assays in SW620 colon carcinoma cell line

Based on the results obtained in the patient series that suggested that SOX2 could be a predictive biomarker of resistance to chemotherapy and on previous results of the group in a modified colon cancer cell line (see **supplementary Figure 6, 7 and 8**), we decided to evaluate the viability of a colon cancer cell line - SW620 - treated with 5-FU, after up- and down-regulation of SOX2 expression.

4.3.1. Determination of the IC_{50} of 5-FU in the SW620 colon cancer cell line

With the purpose of evaluating the cytotoxicity of 5-FU in SW620 colon cancer cell line, the IC_{50} (half maximal inhibitory concentration) value was assessed by two different viability assays: a resazurin-based method (Presto Blue) that measures metabolically active cells, and Sulforhodamine assay that quantifies protein content (**Table 5**).

Table 5. Value of IC_{50} in SW620 colon cancer cell line.

Cell Line	IC_{50} (µg/mL)
SW620	7.72

4.3.2. SOX2 silencing and overexpression

4.3.2.1. Western-blot analysis of SOX2 expression in normal and transiently transfected cells

To confirm that we successfully silenced and overexpressed SOX2 in SW620 colon cancer cell lines we performed a Western-blot to analyse SOX2 protein expression in normal and transiently transfected cells, 72h after transfection (**Figure 33**).

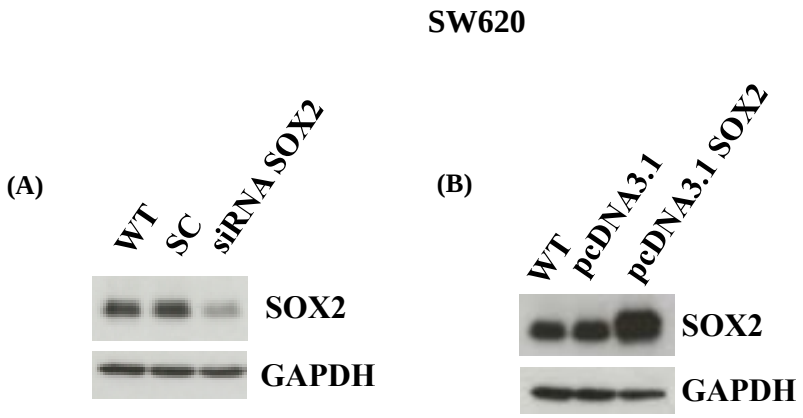


Figure 33. Western-blot analysis of SOX2 protein expression in SW620 wild-type and transfected cells. **(A)** Detection of SOX2 protein expression in wild-type SW620 cells as in negative control of the silencing, a scrambled siRNA (SC), and also in transiently transfected cells with siRNA for SOX2 (siRNA SOX2). **(B)** Detection of SOX2 protein expression in wild-type SW620 cells as in negative control of the overexpression, an empty vector (pcDNA 3.1), and also in transiently transfected cells with pcDNA3.1 for SOX2 (pcDNA 3.1 SOX2).

4.3.2.2. Real-Time PCR analysis of SOX2 mRNA expression in normal and transiently transfected cells

Likewise, we extracted mRNA from normal and transient cells to perform Real-Time PCR to analyse SOX2 mRNA expression 72h after transfection (**Figures 34**).

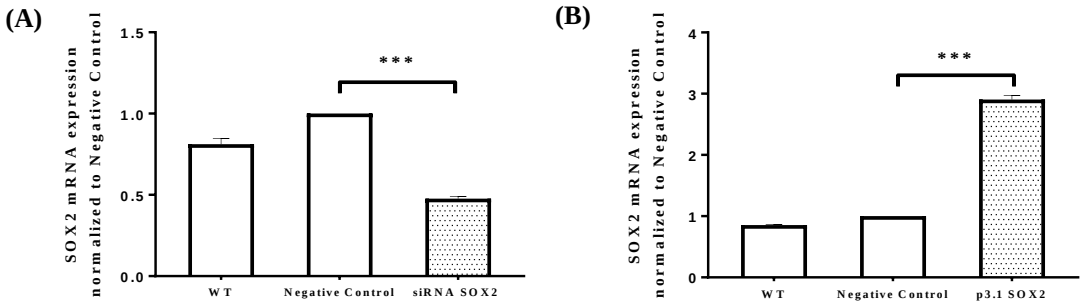


Figure 34. SOX2 mRNA expression comparison between WT and transfected SW620 cells. The values are normalized to SOX2 expression in negative controls. **(A)** silencing with siRNA SOX2 and **(B)** SOX2 overexpression with pcDNA3.1. SOX2 vector. Significant differences (***p<0.001).

We were able to successfully silence and overexpress SOX2 in SW620.

4.3.3. Viability assays in SW620 colon cancer cell line treated with 5-FU

Unfortunately, when we compared the viability of SW620 cells in which SOX2 was silenced or overexpressed, after treating SW620 cells with 5-FU (7.72μg/mL) (**Supplementary Figure 9**), we could not observe significative differences when compared with negative controls (**Figure 35**).

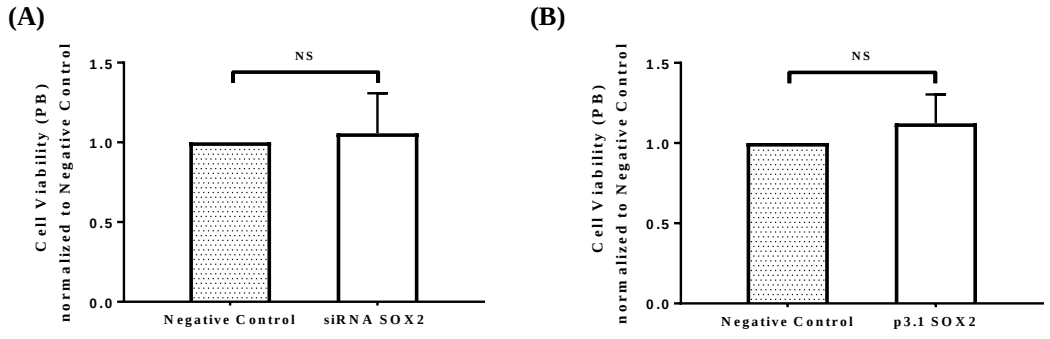


Figure 35. Viability of the SW620 cells treated with 5-FU **(A)** after transient transfection with siRNA SOX2 *versus* negative control (scrambled siRNA), or **(B)** after transient transfection with pcDNA3.1 SOX2 overexpression vector *versus* negative control (empty vector).

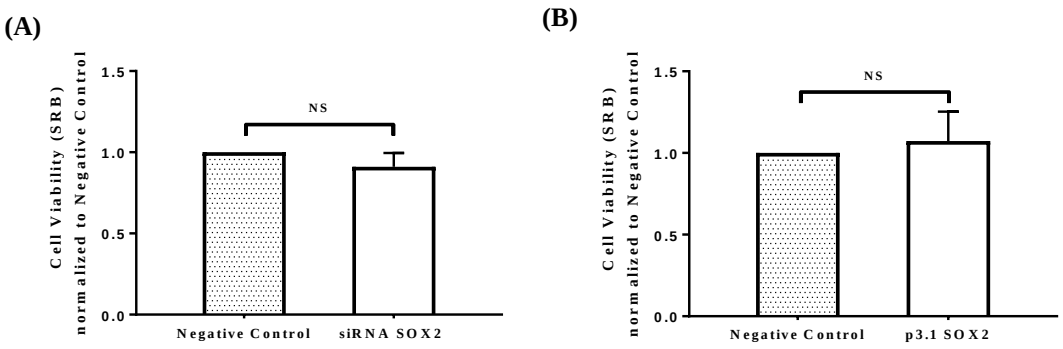


Figure 36. Viability of the SW620 cells treated with 5-FU **(A)** after transient transfection with siRNA SOX2 *versus* negative control (scrambled siRNA), or **(B)** after transient transfection with pcDNA3.1 SOX2 overexpression vector *versus* negative control (empty vector).

V. Discussion

5.1. Cohort characterization

We investigated the prognostic and predictive value of five putative molecular biomarkers - microsatellite instability status, *BRAF*^{V600E} mutation and CDX2, SOX2 and SOX9 expression - in a Portuguese cohort of 230 patients diagnosed with stage II colorectal cancer. In this series, the median patient age at time of diagnosis was 68 years old. This value is just slightly above that reported for the median age at time of diagnosis for all stages, in both men and women, worldwide, that is of 63 years old (American Cancer Society. Colorectal Cancer Facts & Figures, 2017-2019).

In our cohort, 80.8% (186/230) of the patients were alive 5 years after diagnosis. This value is above the 71% OS reported for CRC patients diagnosed in early stages (American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019). In the same period, 20.6% of the patients relapsed which is a high number. Boland *et al.,* 2016 also reported a high percentage of relapse of 20% in their stage II cohort. However, a recent multicentre study with several hundreds of stage II CRC patients reports less than 15% of relapse (Yamano *et al.,* 2018).

In our cohort the percentage of patients treated with adjuvant chemotherapy, in addition to surgery, was of 15.2% (35/230) being lower than the 20% to 59%, reported in different studies (O'Connor *et al.,* 2011; Tsikitis *et al.,* 2014; Grant *et al.,* 2018, Yamano *et al.,* 2018). This difference in administration of adjuvant treatment could partially explain the high percentage of recurrence and emphasizes the existence of an untreated subgroup of patients that is not selected by the current criteria. Chemotherapy seems to have a significant impact in patient OS however this result might be subjected to a bias since the use of adjuvant chemotherapy significantly correlates with younger age at time of diagnosis (57.6±11.6 years in treated *versus* 70.4±10.5; $\boldsymbol{P}<0.001$ in untreated patients). When we compare the DFS of the patients treated or not with adjuvant chemotherapy, no significant differences are observed. This observation corroborates the urgent need of new biomarkers that could enhance the potential of current therapy in order to minimize the risk of recurrence.

5.2. Microsatellite Instability (MSI)

Microsatellite Instability was accessed by evaluating the expression of genes involved in MMR. According to the expression of MLH1, MSH6, MSH2 and PMS2, and

considering positive all the cases with expression regardless of the extension, we found 40% of MSI tumors. This result is higher than expected since most of previous studies report approximately 20% of MSI cases in stage II CRC (reviewed in Boland *et al.*, 2010; Puccini *et al.,* 2017). However, the highest frequency described so far for a stage II CRC series was 57.9% (Moghbeli *et al.,* 2011). MSI is more frequent in colon cancer when compared to the rectum (Hong *et al.,* 2012). A study with only stage II colon cancer cases reported an MSI percentage of 34.1% (Grant *et al.,* 2018). The higher frequency of MSI in this stage II CRC cohort could partially be explained by the fact that it is enriched in colon cancer cases. The expected proportion of rectal/colon cancers is 25% (American Cancer Society. Colorectal Cancer Facts & Figures, 2017-2019). In our cohort the percentage of rectal cancers is only 20.4%.

Like in previous studies MLH1 was the most frequently mutated gene, (Herfarth *et al.,* 1997; Kheirelseid *et al.,* 2013) being its expression lost in 81/92 (88.0%) MSI patients. MSI has been extensively studied as a predictive biomarker in stage II CRC. In fact, clinicians are already recommending to not administrate adjuvant therapy to MSI patients, since they have good prognosis and do not benefit from it. However, in our study MSI status did not correlate with better DFS or OS and it did not have influence in response to therapy. Accordingly, in a clinical trial (QUASAR - Quick and Simple and Reliable) with 1913 patients, (95% of whom had stage II CRC) which were randomly assigned between 5-FU plus folinic acid and no chemotherapy, no evidence was found that patients with dMMR fail to respond to chemotherapy (Hutchins *et al.,* 2011). Moreover, in two recent meta-analysis by Gkekas *et al.,* 2017 and Romiti *et al.,* 2017 it was also not found any value for MSI as a biomarker in stage II colon cancer. In addition, Kim *et al.,* 2015 did not find any correlation between MSI and a better patient outcome but they report a significant correlation between MSS and better DFS, when compared to the MSI patients.

Although very widely described the association between MSI and BRAF, it was not reproduced in our study. A possible explanation, though unexpected, could be that our series is enriched in hereditary forms of CRC. Lynch syndrome, the most common form of hereditary CRC is characterized by MSI phenotype (Thiel *et al.,* 2013). The distinction between hereditary and sporadic MSI CRC is a crucial step in Lynch syndrome diagnoses. Within MSI colorectal cancers, the $BRAF^{V600E}$ mutation was strongly associated with sporadic origin (Capper *et al.,* 2013). Indeed, $BRAF^{V600E}$ mutation in MSI colorectal carcinomas virtually excludes Lynch syndrome (Toon *et al.,* 2013).

In what concerns associations between MSI status and the other putative biomarkers we have studied, the absence of CDX2 expression was much more common in the MSI tumours. CDX2 was already reported to be significantly positively correlated with the expression the proteins involved in DNA repair (Tóth *et al.*, 2018).

5.3. *BRAF^{V600E}* mutation

In our stage II CRC series, BRAF mutation was present in 30 (13%) patients, in accordance with previous reports in which this percentage varies between 4 and 18% (Cuyle *et al.*, 2017). In rectal cancer, BRAF^{V600E} mutation was reported as extremely rare (Tamas *et al.*, 2015) and in our series none of the rectal cancer patients harboured this mutation. Like in previous studies, BRAF^{V600E} strongly correlated with older age patients, female gender and tumour location in the proximal colon (de Cuba *et al.*, 2016; Chen *et al.*, 2016; Puccini *et al.*, 2017).

Since BRAF mutations are associated with older age, its value as a biomarker in early stage CRC is less well clarified (de Cuba *et al.*, 2016). Like previous studies (Chen *et al.*, 2016, French *et al.*, 2008; Mouradov *et al.*, 2013, André *et al.*, 2015; Shen *et al.*, 2016) we did not find any prognostic value for BRAF^{V600E} in our stage II CRC series. Also, our results agree with the former described lack of evidence for a predictive value for BRAF^{V600E} in early stage CRC (Cuyle *et al.*, 2017).

5.4. CDX2 expression

In what concerns the expression of CDX2, we found a downregulation of this intestinal transcription factor in 14.3% of patients. In prior studies loss of CDX2 is found in 4% (Dalerba *et al.*, 2016) to 30% (Baba *et al.*, 2009) CRC patients. Like in previous studies (Dawson *et al.*, 2014; Olsen *et al.*, 2015; Bruun *et al.*, 2018), we did not find prognostic neither predictive value for CDX2 expression.

CDX2 loss of expression was more often found in female patients, as already reported in Zhang *et al.*, 2016. As well, low CDX2 expression correlated with poorly and moderately differentiated tumors in our series. This goes in hand with previous reports that correlate CDX2 loss of expression with poor differentiated tumors (Bakaris *et al.*, 2008; Olsen *et al.*, 2015; Lundberg *et al.*, 2016, Bruun *et al.*, 2018). This would be expected since CDX2 is a major regulator of the intestine-specific genes involved in cell differentiation (Lundberg *et al.*, 2016).

A decreased expression of CDX2 was, like in previous studies, closely linked to MSI (Lundberg *et al.*, 2016; Schirripa *et al.*, 2016; Olsen *et al.*, 2015; Dawnson *et al.*, 2014). However, whether loss of CDX2 expression plays a particularly active role in tumour progression in MSI tumours remains to be elucidated (Olsen *et al.*, 2015).

5.5. SOX2 expression

SOX2 is not expressed in the normal intestinal epithelium. Recently, it was reported that SOX2 is amplified in a variety of cancers (Du *et al.*, 2016; Camilo *et al.*, 2015). We found *de novo* expression of SOX2 in 15.7% of the patients. SOX2 expression has been extensively associated with poor differentiation (Lundberg *et al.*, 2014). In our CRC stage II series, SOX2 correlated also with poor differentiation. SOX2 expression has been reported to be associated with worse patient outcome, however there was significant heterogeneity between studies (Du *et al.*, 2016). In our cohort, SOX2 expression did not show any prognostic value. Moreover, SOX2 overexpression was reported to correlate with poorer OS in patients who did not receive adjuvant chemotherapy, but no differences were found between patients that received adjuvant therapy (Du *et al.*, 2016). In our stage II CRC patient cohort, SOX2 de novo expression significantly correlated with worse DFS survival in patients who received adjuvant chemotherapy. This suggests that SOX2 positive patients could respond worse to therapy.

SOX2 was expected to have an important role in CRC, since it is widely associated with stemness, growth and metastasis (Saigusa *et al.*, 2009; Fang *et al.*, 2010; Neumann *et al.*, 2011; Zheng *et al.*, 2017). The role we found for SOX2 as a predictive biomarker in early CRC is very promising. In a previous study by Yang *et al.*, SOX2 was already reported to be expressed at higher levels in drug-resistant cells when compared to parental colon cancer cells.

For further validation of SOX2 as a predictive biomarker of resistance to therapy, we evaluated the viability of a colon cancer cell line treated with 5-FU, after up- and downregulation of SOX2 expression. According to our results in the retrospective study we expected that SOX2 expression would increase the resistance of the cells to 5-FU, reducing the effectiveness of the 5-FU based adjuvant chemotherapy. So, in cells with SOX2 silencing, we expected less viability when treated with 5-FU. Similarly, when SOX2 is overexpressed, we expected cells to be more resistant and then be more viable after treatment with 5-FU. Unfortunately, we could not observe significant differences. In a different cellular model, with CDX2 knockdown using a genome-editing approach,

SOX2 becomes highly expressed. The cells from the CDX2 knock-out cellular model, which expressed high levels of SOX2, were more resistant to the 5-FU cytotoxicity. To confirm that SOX2 effectively plays a role in this observation we also compared LS174T CDX2-KO cells transfected with siRNA for SOX2 with parental ones. When SOX2 is downregulated by siRNA, cells become significantly more sensitive to 5-FU, leading to the conclusion that in fact the resistance is, at least partially, mediated by SOX2. At this point we argue that the different results could be due to the efficacy of transient transfection which does not guarantee a significant number of transfected cells to observe biological differences.

5.6. SOX9 expression

We found that SOX9 expression was lost in 10.9% of cases. It is very clear the stronger staining of SOX9 in the normal colonic epithelium in the proliferative lower half of the crypts of Lieberkuhn. This expression pattern was already noticed by Blache *et al.*, 2004 and Lu *et al.,* 2008 who observed SOX9 also in the nuclei of the Paneth cells also located at the bottom of the crypts. In our stage II cohort, SOX9 did not have any prognostic neither predictive value. However, SOX9 negativity when associated with SOX2 positivity significantly correlated with worst DFS. SOX9 loss of expression was previously, reported to identify patients with worse prognosis in CRC (Espersen *et al.*, 2016) and SOX2 was also previously associated with worse patient outcome (Lundberg *et al.*, 2014),

SOX9 has been reported to be involved in the repression of differentiation genes including CDX2 (Blache *et al.*, 2004; Subramanian *et al.* 1998; Sillberg *et al.*, 2000) Interestingly, CDX2 expression was also associated with SOX9 expression in our stage II cohort, however, not as expected since CDX2 negative expression was more associated with SOX9 negative expression.

VI. Conclusions

The main goal of this thesis was to evaluate the prognostic and predictive value of a panel of 5 molecules of interest in colorectal carcinogenesis in a Portuguese stage II CRC patient cohort. Namely, we investigated the value of microsatellite instability status, *BRAF*^{V600E} mutation, and CDX2, SOX2 and SOX9 expression.

In this CRC series, loss of CDX2 expression and *de novo* SOX2 expression significantly correlated with less tumour differentiation suggesting that modulation of these transcription factors contribute to tumour progression. SOX2 expression *per se* did not predict patient outcome but in combination with SOX9 loss of expression predicts tumour recurrence, which has not been reported so far. We would like to further investigate the combined effect of these two transcription factors in intestinal cellular models and to understand the molecular mechanisms involved. We could also test the tumor-initiation of these cells by carrying out *in vitro* tumorsphere formation assays and xenografts in nude mice.

Moreover, we identified for the first time SOX2 as a putative biomarker of response to adjuvant chemotherapy in stage II CRC patients. In patients treated with adjuvant chemotherapy, *de novo* SOX2 expression significantly correlated with increased risk of recurrence. This observation suggests that SOX2 might be involved in resistance to 5-FU based adjuvant chemotherapy. The *in vitro* experiments performed in this study, based on transient modulation of SOX2 expression did not allow to reinforce this conclusion. Different experiments are ongoing to modulate SOX2 expression using stable transfection assays to more accurately measure its influence in resistance to chemotherapy.

VI. Bibliography

American Cancer Society (2018). Colorectal cancer. Early Detection, Diagnosis, and Staging. Retrieved from http://www.cancer.org/cancer/cancercauses/geneticsandcancer/index, accessed on 26/07/2018

American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019

André, T., de Gramont, A., Vernerey, D., Chibaudel, B., Bonnetain, F., Tijeras-Raballand, A., … de Gramont, A. (2015). Adjuvant Fluorouracil, Leucovorin, and Oxaliplatin in Stage II to III Colon Cancer: Updated 10-Year Survival and Outcomes According to *BRAF* Mutation and Mismatch Repair Status of the MOSAIC Study. *Journal of Clinical Oncology*, *33*(35), 4176–4187. https://doi.org/10.1200/JCO.2015.63.4238

Baba, Y., Nosho, K., Shima, K., Freed, E., Irahara, N., Philips, J., … Ogino, S. (2009). Relationship of CDX2 Loss with Molecular Features and Prognosis in Colorectal Cancer. *Clinical Cancer Research*, *15*(14), 4665–4673. https://doi.org/10.1158/1078-0432.CCR-09-0401

Bae, J. M., Lee, T. H., Cho, N.-Y., Kim, T.-Y., & Kang, G. H. (2015). Loss of CDX2 expression is associated with poor prognosis in colorectal cancer patients. *World Journal of Gastroenterology*, *21*(5), 1457. https://doi.org/10.3748/wjg.v21.i5.1457

Bakaris, S., Cetinkaya, A., Ezberci, F., & Ekerbicer, H. (2008). Expression of homeodomain protein CDX2 in colorectal adenoma and adenocarcinoma. *Histology and Histopathology*, *23*(9), 1043–7. https://doi.org/10.14670/HH-23.1043

Barras, D., Missiaglia, E., Wirapati, P., Sieber, O. M., Jorissen, R. N., Love, C., … Delorenzi, M. (2017). *BRAF V600E* Mutant Colorectal Cancer Subtypes Based on Gene Expression. *Clinical Cancer Research*, *23*(1), 104–115. https://doi.org/10.1158/1078-0432.CCR-16-0140

Bastide, P., Darido, C., Pannequin, J., Kist, R., Robine, S., Marty-Double, C., … Jay, P. (2007). Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. *The Journal of Cell Biology*, *178*(4), 635–48. https://doi.org/10.1083/jcb.200704152

Blache, P., van de Wetering, M., Duluc, I., Domon, C., Berta, P., Freund, J.-N., … Jay, P. (2004). SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes. *The Journal of Cell Biology*, *166*(1), 37–47. https://doi.org/10.1083/jcb.200311021

Boland, C. R., & Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. *Gastroenterology*, *138*(6), 2073–2087.e3. https://doi.org/10.1053/j.gastro.2009.12.064

Bruun, J., Kolberg, M., Nesland, J. M., Svindland, A., Nesbakken, A., & Lothe, R. A. (2014). Prognostic Significance of β-Catenin, E-Cadherin, and SOX9 in Colorectal Cancer: Results from a Large Population-Representative Series. *Frontiers in Oncology*, *4*, 118. https://doi.org/10.3389/fonc.2014.00118

Camilo, V., Barros, R., Celestino, R., Castro, P., Vieira, J., Teixeira, M. R., … Almeida, R. (2014). Immunohistochemical molecular phenotypes of gastric cancer based on SOX2 and CDX2 predict patient outcome. *BMC Cancer*, *14*, 753. https://doi.org/10.1186/1471-2407-14-753

Camilo, V., Barros, R., Sousa, S., Magalhaes, A. M., Lopes, T., Mario Santos, A., … Almeida, R. (2012). Helicobacter pylori and the BMP pathway regulate CDX2 and SOX2 expression in gastric cells. *Carcinogenesis*, *33*(10), 1985–1992. https://doi.org/10.1093/carcin/bgs233

Camilo, V., Garrido, M., Valente, P., Ricardo, S., Amaral, A. L., Barros, R., … Almeida, R. (2015). Differentiation reprogramming in gastric intestinal metaplasia and dysplasia: role of SOX2 and CDX2. *Histopathology*, *66*(3), 343–350. https://doi.org/10.1111/his.12544

Capper, D., Voigt, A., Bozukova, G., Ahadova, A., Kickingereder, P., von Deimling, A., … Kloor, M. (2013). BRAF V600E-specific immunohistochemistry for the exclusion of Lynch syndrome in MSI-H colorectal cancer. *International Journal of Cancer*, *133*(7), 1624–1630. https://doi.org/10.1002/ijc.28183

Carrasco-Garcia, E., Lopez, L., Aldaz, P., Arevalo, S., Aldaregia, J., Egaña, L., … Matheu, A. (2016). SOX9-regulated cell plasticity in colorectal metastasis is attenuated by rapamycin. *Scientific Reports*, *6*(1), 32350. https://doi.org/10.1038/srep32350

Chen, K.-H., Lin, Y.-L., Liau, J.-Y., Tsai, J.-H., Tseng, L.-H., Lin, L.-I., … Cheng, A.-L. (2016). BRAF mutation may have different prognostic implications in early- and late-stage colorectal cancer. *Medical Oncology (Northwood, London, England)*, *33*(5), 39. https://doi.org/10.1007/s12032-016-0756-6

Cremolini, C., Loupakis, F., Antoniotti, C., Lupi, C., Sensi, E., Lonardi, S., … Falcone, A. (2015). FOLFOXIRI plus bevacizumab versus FOLFIRI plus bevacizumab as first-line treatment of patients with metastatic colorectal cancer: updated overall survival and molecular subgroup analyses of the open-label, phase 3 TRIBE study. *The Lancet Oncology*, *16*(13), 1306–1315. https://doi.org/10.1016/S1470-2045(15)00122-9

Cuyle, P.-J., & Prenen, H. (2017). Current and future biomarkers in the treatment of colorectal cancer. *Acta Clinica Belgica*, *72*(2), 103–115. https://doi.org/10.1080/17843286.2016.1262996

Dalerba, P., Sahoo, D., Paik, S., Guo, X., Yothers, G., Song, N., … Clarke, M. F. (2016). CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *New England Journal of Medicine*, *374*(3), 211–222. https://doi.org/10.1056/NEJMoa1506597

Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., … Futreal, P. A. (2002). Mutations of the BRAF gene in human cancer. *Nature*, *417*(6892), 949–954. https://doi.org/10.1038/nature00766

Dawson, H., Galván, J. A., Helbling, M., Muller, D.-E., Karamitopoulou, E., Koelzer, V. H., … Zlobec, I. (2014). Possible role of Cdx2 in the serrated pathway of colorectal cancer characterized by BRAF mutation, high-level CpG Island methylator phenotype and mismatch repair-deficiency. *International Journal of Cancer*, *134*(10), 2342–2351. https://doi.org/10.1002/ijc.28564

de Cuba, E. M. V., Snaebjornsson, P., Heideman, D. A. M., van Grieken, N. C. T., Bosch, L. J. W., Fijneman, R. J. A., … Meijer, G. A. (2016). Prognostic value of *BRAF* and *KRAS* mutation status in stage II and III microsatellite instable colon cancers. *International Journal of Cancer*, *138*(5), 1139–1145. https://doi.org/10.1002/ijc.29855

De Sousa E Melo, F., Wang, X., Jansen, M., Fessler, E., Trinh, A., de Rooij, L. P. M. H., … Vermeulen, L. (2013). Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nature Medicine*, *19*(5), 614–618. https://doi.org/10.1038/nm.3174

Dienstmann, R., Mason, M. J., Sinicrope, F. A., Phipps, A. I., Tejpar, S., Nesbakken, A., … Guinney, J. (2017). Prediction of overall survival in stage II and III colon cancer beyond TNM system: a retrospective, pooled biomarker study. *Annals of Oncology*, *28*(5), 1023–1031. https://doi.org/10.1093/annonc/mdx052

Dienstmann, R., Salazar, R., & Tabernero, J. (2015). Personalizing Colon Cancer Adjuvant Therapy: Selecting Optimal Treatments for Individual Patients. *Journal of Clinical Oncology*, *33*(16), 1787–1796. https://doi.org/10.1200/JCO.2014.60.0213

Du, X., Wang, L., Chen, X., Li, Y., Li, Y., & Cao, Y. (2016). Prognostic value of Sox2 expression in digestive tract cancers: A meta-analysis. *Journal of Huazhong University of Science and Technology [Medical Sciences]*, *36*(3), 305–312. https://doi.org/10.1007/s11596-016-1584-9

Edge, S. B., & Compton, C. C. (2010). The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Annals of Surgical Oncology*, *17*(6), 1471–1474. https://doi.org/10.1245/s10434-010-0985-4

Fang, X., Yu, W., Li, L., Shao, J., Zhao, N., Chen, Q., … Lin, B. (2010). ChIP-seq and Functional Analysis of the SOX2 Gene in Colorectal Cancers. *OMICS: A Journal of Integrative Biology*, *14*(4), 369–384. https://doi.org/10.1089/omi.2010.0053

Formeister, E. J., Sionas, A. L., Lorance, D. K., Barkley, C. L., Lee, G. H., & Magness, S. T. (2009). Distinct SOX9 levels differentially mark stem/progenitor populations and enteroendocrine cells of the small intestine epithelium. *American Journal of Physiology. Gastrointestinal and Liver Physiology*, *296*(5), G1108-18. https://doi.org/10.1152/ajpgi.00004.2009

French, A. J., Sargent, D. J., Burgart, L. J., Foster, N. R., Kabat, B. F., Goldberg, R., … Thibodeau, S. N. (2008). Prognostic Significance of Defective Mismatch Repair and BRAF V600E in Patients with Colon Cancer. *Clinical Cancer Research*, *14*(11), 3408–3415. https://doi.org/10.1158/1078-

0432.CCR-07-1489

Garber, K. (2017). Oncologists await historic first: a pan-tumor predictive marker, for immunotherapy. *Nature Biotechnology*, *35*(4), 297–298. https://doi.org/10.1038/nbt0417-297a

Gkekas, I., Novotny, J., Pecen, L., Strigård, K., Palmqvist, R., & Gunnarsson, U. (2017). Microsatellite Instability as a Prognostic Factor in Stage II Colon Cancer Patients, a Meta-Analysis of Published Literature. *Anticancer Research*, *37*(12), 6563–6574. https://doi.org/10.21873/anticanres.12113

Grant, M., Haydon, A., Au, L., Wilkins, S., Oliva, K., Segelov, E., … Shapiro, J. (2018). Immunohistochemistry testing for mismatch repair deficiency in Stage 2 colon cancer: A cohort study of two cancer centres. *International Journal of Surgery*, *51*, 71–75. https://doi.org/10.1016/j.ijsu.2018.01.020

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., … Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, *21*(11), 1350–6. https://doi.org/10.1038/nm.3967

Hardiman, K. (2018). Update on Sporadic Colorectal Cancer Genetics. *Clinics in Colon and Rectal Surgery*, *31*(03), 147–152. https://doi.org/10.1055/s-0037-1602234

Herfarth, K. K., Kodner, I. J., Whelan, A. J., Ivanovich, J. L., Bracamontes, J. R., Wells, S. A., & Goodfellow, P. J. (1997). Mutations in MLH1 are more frequent than in MSH2 in sporadic colorectal cancers with microsatellite instability. *Genes, Chromosomes & Cancer*, *18*(1), 42–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8993979

Hong, S. P., Min, B. S., Kim, T. Il, Cheon, J. H., Kim, N. K., Kim, H., & Kim, W. H. (2012). The differential impact of microsatellite instability as a marker of prognosis and tumour response between colon cancer and rectal cancer. *European Journal of Cancer*, *48*(8), 1235–1243. https://doi.org/10.1016/J.EJCA.2011.10.005

Hutchins, G., Southward, K., Handley, K., Magill, L., Beaumont, C., Stahlschmidt, J., … Quirke, P. (2011). Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *29*(10), 1261–70. https://doi.org/10.1200/JCO.2010.30.1366

Kheirelseid, E. A. H., Miller, N., Chang, K. H., Curran, C., Hennessey, E., Sheehan, M., & Kerin, M. J. (2013). Mismatch repair protein expression in colorectal cancer. *Journal of Gastrointestinal Oncology*, *4*(4), 397–408. https://doi.org/10.3978/j.issn.2078-6891.2013.021

Kim, J. E., Hong, Y. S., Kim, H. J., Kim, K., Lee, J.-L., Park, S. J., … Kim, T. W. (2015). Defective Mismatch Repair Status was not Associated with DFS and OS in Stage II Colon Cancer Treated with Adjuvant Chemotherapy. *Annals of Surgical Oncology*, *22*(S3), 630–637. https://doi.org/10.1245/s10434-015-4807-6

Kuipers, E. J., Grady, W. M., Lieberman, D., Seufferlein, T., Sung, J. J., Boelens, P. G., … Watanabe, T. (2015). Colorectal cancer. *Nature Reviews Disease Primers*, *1*, 15065. https://doi.org/10.1038/nrdp.2015.65

Lea, D., Håland, S., Hagland, H. R., & Søreide, K. (2014). Accuracy of TNM staging in colorectal cancer: a review of current culprits, the modern role of morphology and stepping-stones for improvements in the molecular era. *Scandinavian Journal of Gastroenterology*, *49*(10), 1153–1163. https://doi.org/10.3109/00365521.2014.950692

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Lü, B., Fang, Y., Xu, J., Wang, L., Xu, F., Xu, E., … Lai, M. (2008). Analysis of SOX9 Expression in Colorectal Cancer. *American Journal of Clinical Pathology*, *130*(6), 897–904. https://doi.org/10.1309/AJCPW1W8GJBQGCNI

Lundberg, I. V., Edin, S., Eklöf, V., Öberg, Å., Palmqvist, R., & Wikberg, M. L. (2016). SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer. *BMC Cancer*, *16*(1), 471. https://doi.org/10.1186/s12885-016-2509-5

Lundberg, I. V., Löfgren Burström, A., Edin, S., Eklöf, V., Öberg, Å., Stenling, R., … Wikberg, M. L. (2014). SOX2 Expression Is Regulated by BRAF and Contributes to Poor Patient Prognosis in Colorectal Cancer. *PLoS ONE*, *9*(7), e101957. https://doi.org/10.1371/journal.pone.0101957

Marcker Espersen, M. L., Linnemann, D., Christensen, I. J., Alamili, M., Troelsen, J. T., & Høgdall, E. (2016). SOX9 expression predicts relapse of stage II colon cancer patients. *Human Pathology*, *52*, 38–46. https://doi.org/10.1016/j.humpath.2015.12.026

Markowitz, S. D., & Bertagnolli, M. M. (2009). Molecular origins of cancer: Molecular basis of colorectal cancer. *The New England Journal of Medicine*, *361*(25), 2449–60. https://doi.org/10.1056/NEJMra0804588

Matheu, A., Collado, M., Wise, C., Manterola, L., Cekaite, L., Tye, A. J., … Lovell-Badge, R. (2012). Oncogenicity of the Developmental Transcription Factor Sox9. *Cancer Research*, *72*(5), 1301–1315. https://doi.org/10.1158/0008-5472.CAN-11-3660

McShane, L. M., Altman, D. G., Sauerbrei, W., Taube, S. E., Gion, M., Clark, G. M., & Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. (2005). Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK). *JNCI: Journal of the National Cancer Institute*, *97*(16), 1180–1184. https://doi.org/10.1093/jnci/dji237

Merok, M. A., Ahlquist, T., Røyrvik, E. C., Tufteland, K. F., Hektoen, M., Sjo, O. H., … Nesbakken, A. (2013). Microsatellite instability has a positive prognostic impact on stage II colorectal cancer after complete resection: results from a large, consecutive Norwegian series. *Annals of Oncology*, *24*(5), 1274–1282. https://doi.org/10.1093/annonc/mds614

Moghbeli, M., Moaven, O., Dadkhah, E., Farzadnia, M., Roshan, N. M., Asadzadeh-Aghdaee, H., … Abbaszadegan, M. R. (2011). High frequency of microsatellite instability in sporadic colorectal cancer patients in Iran. *Genetics and Molecular Research : GMR*, *10*(4), 3520–9. https://doi.org/10.4238/2011.December.14.4

Molaei, M., Kishani Farahani, R., Maftouh, M., Taleghani, M. Y., Vahdatinia, M., Khatami, F., … Zali, M. R. (2016). Lack of BRAFV600E mutation in stage I and II of colorectal cancer. *Gastroenterology and Hepatology from Bed to Bench*, *9*(2), 94–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27099668

Mori–Akiyama, Y., van den Born, M., van Es, J. H., Hamilton, S. R., Adams, H. P., Zhang, J., … de Crombrugghe, B. (2007). SOX9 Is Required for the Differentiation of Paneth Cells in the Intestinal Epithelium. *Gastroenterology*, *133*(2), 539–546. https://doi.org/10.1053/j.gastro.2007.05.020

Mouradov, D., Domingo, E., Gibbs, P., Jorissen, R. N., Li, S., Soo, P. Y., … Sieber, O. M. (2013). Survival in stage II/III colorectal cancer is independently predicted by chromosomal and microsatellite instability, but not by specific driver mutations. *The American Journal of Gastroenterology*, *108*(11), 1785–1793. https://doi.org/10.1038/ajg.2013.292

Neumann, J., Bahr, F., Horst, D., Kriegl, L., Engel, J., Mejías-Luque, R., … Jung, A. (2011). SOX2 expression correlates with lymph-node metastases and distant spread in right-sided colon cancer. *BMC Cancer*, *11*(1), 518. https://doi.org/10.1186/1471-2407-11-518

O'Connell, J. B., Maggard, M. A., & Ko, C. Y. (2004). Colon Cancer Survival Rates With the New American Joint Committee on Cancer Sixth Edition Staging. *JNCI Journal of the National Cancer Institute*, *96*(19), 1420–1425. https://doi.org/10.1093/jnci/djh275

O'Connor, E. S., Greenblatt, D. Y., LoConte, N. K., Gangnon, R. E., Liou, J.-I., Heise, C. P., & Smith, M. A. (2011). Adjuvant chemotherapy for stage II colon cancer with poor prognostic features. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *29*(25), 3381–8. https://doi.org/10.1200/JCO.2010.34.3426

Olsen, J., Eiholm, S., Kirkeby, L. T., Espersen, M. L. M., Jess, P., Gögenür, I., … Troelsen, J. T. (2016). CDX2 downregulation is associated with poor differentiation and MMR deficiency in colon cancer. *Experimental and Molecular Pathology*, *100*(1), 59–66. https://doi.org/10.1016/j.yexmp.2015.11.009

PDQ Cancer Genetics Editorial Board. Genetics of Colorectal Cancer (PDQ®): Health Professional Version. 2018 Jun 14. In: PDQ Cancer Information Summaries [Internet]. Bethesda (MD): National

Cancer Institute (US); 2002-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK126744/

Prévostel, C., & Blache, P. (2017). The dose-dependent effect of SOX9 and its incidence in colorectal cancer. *European Journal of Cancer*, *86*, 150–157. https://doi.org/10.1016/j.ejca.2017.08.037

Prévostel, C., Rammah-Bouazza, C., Trauchessec, H., Canterel-Thouennon, L., Busson, M., Ychou, M., & Blache, P. (2014). SOX9 is an atypical intestinal tumor suppressor controlling the oncogenic Wnt/ß-catenin signaling. *Oncotarget*, *7*(50), 82228–82243. https://doi.org/10.18632/oncotarget.10573

Puccini, A., Berger, M. D., Zhang, W., & Lenz, H.-J. (2017). What We Know About Stage II and III Colon Cancer: It's Still Not Enough. *Targeted Oncology*, *12*(3), 265–275. https://doi.org/10.1007/s11523-017-0494-5

Punt, C. J. A., Koopman, M., & Vermeulen, L. (2017). From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nature Reviews Clinical Oncology*, *14*(4), 235–246. https://doi.org/10.1038/nrclinonc.2016.171

Rex, D. K., Ahnen, D. J., Baron, J. A., Batts, K. P., Burke, C. A., Burt, R. W., … Church, J. (2012). Serrated Lesions of the Colorectum: Review and Recommendations From an Expert Panel. *The American Journal of Gastroenterology*, *107*(9), 1315–1329. https://doi.org/10.1038/ajg.2012.161

Ribic, C. M., Sargent, D. J., Moore, M. J., Thibodeau, S. N., French, A. J., Goldberg, R. M., … Gallinger, S. (2003). Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. *New England Journal of Medicine*, *349*(3), 247–257. https://doi.org/10.1056/NEJMoa022289

Romiti, A., Rulli, E., Pilozzi, E., Gerardi, C., Roberto, M., Legramandi, L., … Floriani, I. (2017). Exploring the Prognostic Role of Microsatellite Instability in Patients With Stage II Colorectal Cancer: A Systematic Review and Meta-Analysis. *Clinical Colorectal Cancer*, *16*(2), e55–e59. https://doi.org/10.1016/j.clcc.2016.08.007

Saigusa, S., Tanaka, K., Toiyama, Y., Yokoe, T., Okugawa, Y., Ioue, Y., … Kusunoki, M. (2009). Correlation of CD133, OCT4, and SOX2 in Rectal Cancer and Their Association with Distant Recurrence After Chemoradiotherapy. *Annals of Surgical Oncology*, *16*(12), 3488–3498. https://doi.org/10.1245/s10434-009-0617-z

Sanz-Garcia, E., Argiles, G., Elez, E., & Tabernero, J. (2017). BRAF mutant colorectal cancer: prognosis, treatment, and new perspectives. *Annals of Oncology*, *28*(11), 2648–2657. https://doi.org/10.1093/annonc/mdx401

Sarkar, A., & Hochedlinger, K. (2013). The Sox Family of Transcription Factors: Versatile Regulators of Stem and Progenitor Cell Fate. *Cell Stem Cell*, *12*(1), 15–30. https://doi.org/10.1016/j.stem.2012.12.007

Schirripa, M., Loupakis, F., & Lenz, H.-J. (2016). CDX2 as a Prognostic Biomarker in Colon Cancer. *The New England Journal of Medicine*, *374*(22), 2183. https://doi.org/10.1056/NEJMc1602584#SA2

Shen, Y., Han, X., Wang, J., Wang, S., Yang, H., Lu, S.-H., & Shi, Y. (2016). Prognostic impact of mutation profiling in patients with stage II and III colon cancer. *Scientific Reports*, *6*(1), 24310. https://doi.org/10.1038/srep24310

Shi, Z., Chiang, C.-I., Mistretta, T.-A., Major, A., & Mori-Akiyama, Y. (2013). SOX9 directly regulates *IGFBP-4* in the intestinal epithelium. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *305*(1), G74–G83. https://doi.org/10.1152/ajpgi.00086.2013

Silberg, D. G., Swain, G. P., Suh, E. R., & Traber, P. G. (2000). Cdx1 and cdx2 expression during intestinal development. *Gastroenterology*, *119*(4), 961–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11040183

Sinicrope, F. A. (2010). DNA mismatch repair and adjuvant chemotherapy in sporadic colon cancer. *Nature Reviews Clinical Oncology*, *7*(3), 174–177. https://doi.org/10.1038/nrclinonc.2009.235

Sinicrope, F. A., Foster, N. R., Thibodeau, S. N., Marsoni, S., Monges, G., Labianca, R., … Sargent, D. J. (2011). DNA mismatch repair status and colon cancer recurrence and survival in clinical trials of 5-fluorouracil-based adjuvant therapy. *Journal of the National Cancer Institute*, *103*(11), 863–75.

https://doi.org/10.1093/jnci/djr153

Subramanian, V., Meyer, B., & Evans, G. S. (1998). The murine Cdx1 gene product localises to the proliferative compartment in the developing and regenerating intestinal epithelium. *Differentiation*, *64*(1), 11–18. https://doi.org/10.1046/j.1432-0436.1998.6410011.x

Takakura, Y., Hinoi, T., Oue, N., Sasada, T., Kawaguchi, Y., Okajima, M., … Ohdan, H. (2010). Molecular and Cellular Pathobiology CDX2 Regulates Multidrug Resistance 1 Gene Expression in Malignant Intestinal Epithelium. https://doi.org/10.1158/0008-5472.CAN-09-4701

Tamas, K., Walenkamp, A. M. E., de Vries, E. G. E., van Vugt, M. A. T. M., Beets-Tan, R. G., van Etten, B., … Hospers, G. A. P. (2015). Rectal and colon cancer: Not just a different anatomic site. *Cancer Treatment Reviews*, *41*(8), 671–679. https://doi.org/10.1016/j.ctrv.2015.06.007

Thiel, A., Heinonen, M., Kantonen, J., Gylling, A., Lahtinen, L., Korhonen, M., … Ristimäki, A. (2013). BRAF mutation in sporadic colorectal cancer and Lynch syndrome. *Virchows Archiv*, *463*(5), 613–621. https://doi.org/10.1007/s00428-013-1470-9

Tï¿½th, C., Sï¿½kï¿½sd, F., Valicsek, E., Herpel, E., Schirmacher, P., & Tiszlavicz, L. (2018). Loss of CDX2 gene expression is associated with DNA repair proteins and is a crucial member of the Wnt signaling pathway in liver metastasis of colorectal cancer. *Oncology Letters*, *15*(3), 3586–3593. https://doi.org/10.3892/ol.2018.7756

Tong, K., Pellón-Cárdenas, O., Sirihorachai, V. R., Warder, B. N., Kothari, O. A., Perekatt, A. O., … Verzi, M. P. (2017). Degree of Tissue Differentiation Dictates Susceptibility to BRAF-Driven Colorectal Cancer. *Cell Reports*, *21*(13), 3833–3845. https://doi.org/10.1016/j.celrep.2017.11.104

Toon, C. W., Walsh, M. D., Chou, A., Capper, D., Clarkson, A., Sioson, L., … Gill, A. J. (2013). BRAFV600E Immunohistochemistry Facilitates Universal Screening of Colorectal Cancers for Lynch Syndrome. *The American Journal of Surgical Pathology*, *37*(10), 1592–1602. https://doi.org/10.1097/PAS.0b013e31828f233d

Topol, L., Chen, W., Song, H., Day, T. F., & Yang, Y. (2009). Sox9 Inhibits Wnt Signaling by Promoting β-Catenin Phosphorylation in the Nucleus. *Journal of Biological Chemistry*, *284*(5), 3323–3333. https://doi.org/10.1074/jbc.M808048200

Tsikitis, V. L., Larson, D. W., Huebner, M., Lohse, C. M., & Thompson, P. A. (2014). Predictors of recurrence free survival for patients with stage II and III colon cancer. *BMC Cancer*, *14*, 336. https://doi.org/10.1186/1471-2407-14-336

Vasen, H. F. A., Tomlinson, I., & Castells, A. (2015). Clinical management of hereditary colorectal cancer syndromes. *Nature Reviews Gastroenterology & Hepatology*, *12*(2), 88–97. https://doi.org/10.1038/nrgastro.2014.229

Veen, T., Nedrebø, B. S., Stormark, K., Søreide, J. A., Kørner, H., & Søreide, K. (2013). Qualitative and Quantitative Issues of Lymph Nodes as Prognostic Factor in Colon Cancer. *Digestive Surgery*, *30*(1), 1–11. https://doi.org/10.1159/000349923

Vogelaar, F., Van Erning, F., Reimers, M., Van Der Linden, J., Pruijt, J., Van Den Brule, A., & Bosscha, K. (2015). The prognostic value of Microsatellite Instability, KRAS, BRAF and PIK3CA mutations in stage II colon cancer patients. *Molecular Medicine (Cambridge, Mass.)*, *21*(1), 1–26. https://doi.org/10.2119/molmed.2015.00220

Watson, M. M., & Søreide, K. (2016). The Prognostic Yield of Biomarkers Harvested in Chemotherapy-Naive Stage II Colon Cancer: Can We Separate the Wheat from the Chaff? *Molecular Medicine*, *22*(1), 1. https://doi.org/10.2119/molmed.2016.00098

Yamano, T., Yamauchi, S., Tsukamoto, K., Noda, M., Kobayashi, M., Hamanaka, M., … Japanese Study Group for Postoperative Follow-up of Colorectal Cancer. (2018). Evaluation of appropriate follow-up after curative surgery for patients with colorectal cancer using time to recurrence and survival after recurrence: a retrospective multicenter study. *Oncotarget*, *9*(39), 25474–25490. https://doi.org/10.18632/oncotarget.25312

Yang, D., Wang, H., Zhang, J., Li, C., Lu, Z., Liu, J., … Qian, H. (2013). In vitro characterization of stem cell-like properties of drug-resistant colon cancer subline. *Oncology Research*, *21*(1), 51–7.

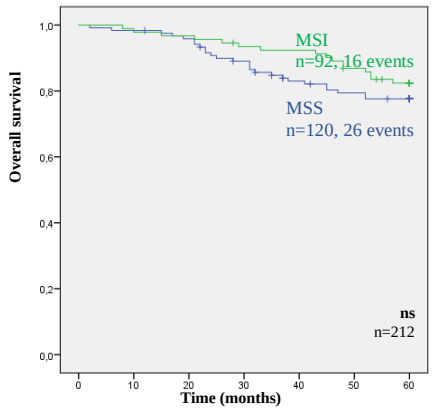
https://doi.org/10.3727/096504013X13793555706768

Zarkavelis, G., Boussios, S., Papadaki, A., Katsanos, K. H., Christodoulou, D. K., & Pentheroudakis, G. (2017). Current and future biomarkers in colorectal cancer. *Annals of Gastroenterology*, *30*(6), 613–621. https://doi.org/10.20524/aog.2017.0191

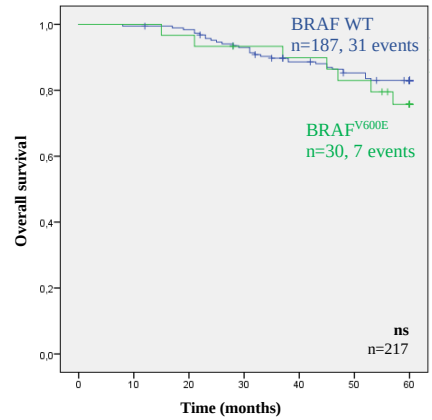
ZHENG, J.-B., QIAO, L.-N., SUN, X.-J., QI, J., REN, H.-L., WEI, G.-B., … JIA, P.-B. (2015). Overexpression of caudal-related homeobox transcription factor 2 inhibits the growth of transplanted colorectal tumors in nude mice. *Molecular Medicine Reports*, *12*(3), 3409–3415. https://doi.org/10.3892/mmr.2015.3838

Zheng, J., Xu, L., Pan, Y., Yu, S., Wang, H., Kennedy, D., & Zhang, Y. (2017). Sox2 modulates motility and enhances progression of colorectal cancer via the Rho-ROCK signaling pathway. *Oncotarget*, *8*(58), 98635–98645. https://doi.org/10.18632/oncotarget.21709

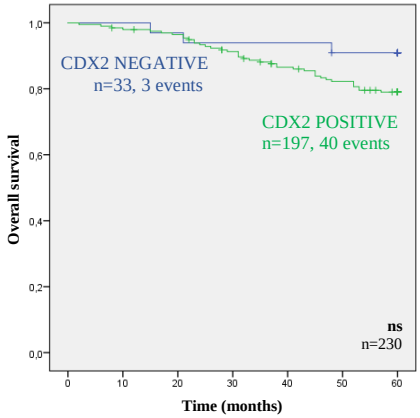
VII. Supplementary data



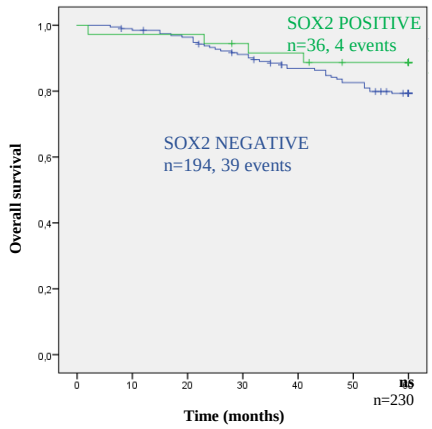
Supplementary Figure 1. Kaplan-Meyer survival analysis showing the probability of OS for 212 patients according to MSI status.



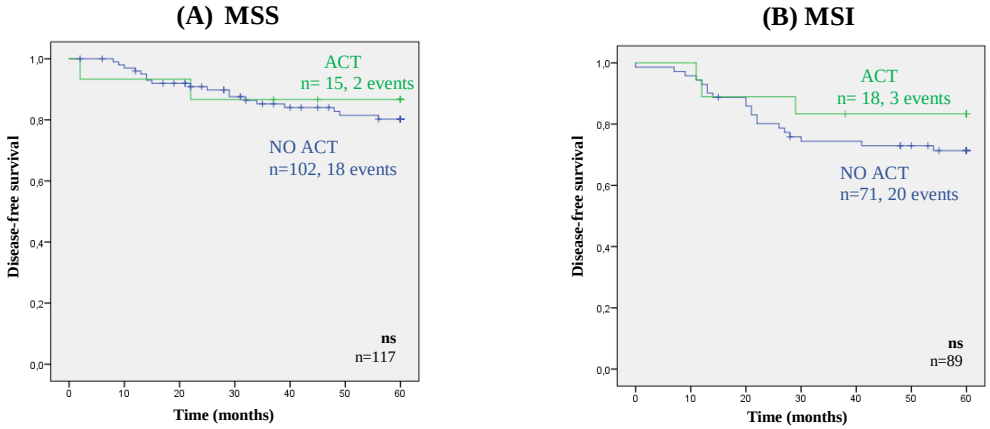
Supplementary Figure 2. Kaplan-Meyer survival analysis showing the probability of OS for 217 patients according to *BRAF*^{V600E} status.



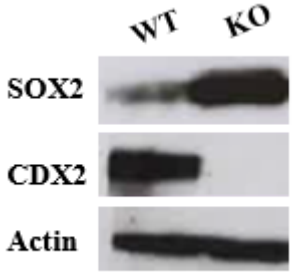
Supplementary Figure 3. Kaplan-Meyer survival analysis showing the probability of OS for 230 patients according to CDX2 expression.



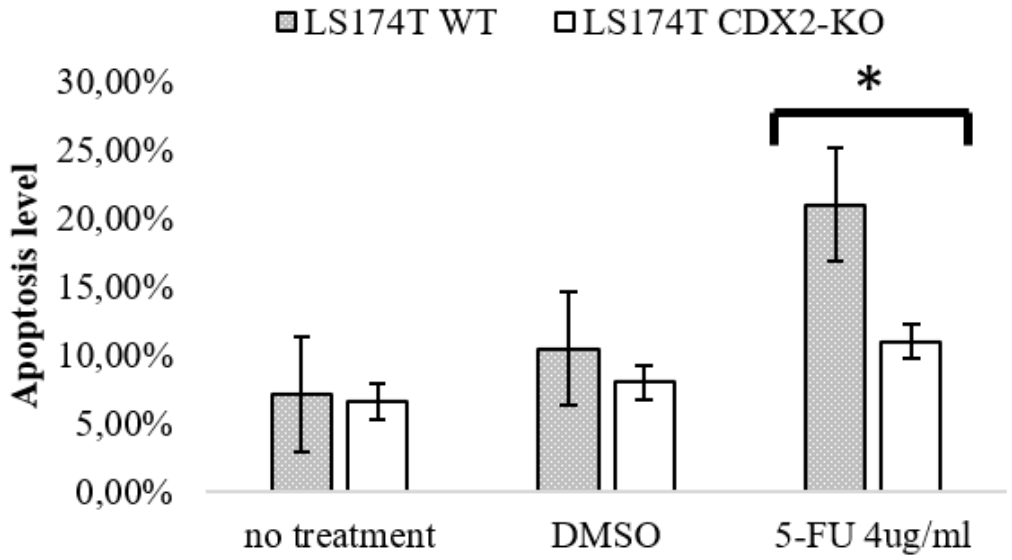
Supplementary Figure 4. Kaplan-Meyer survival analysis showing the probability of OS for 230 patients according to SOX2 expression.



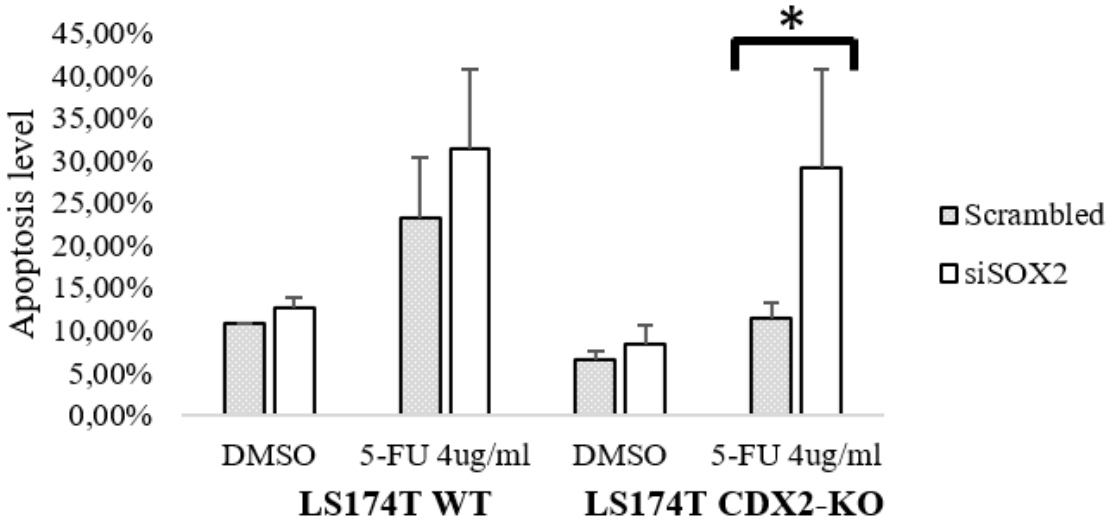
Supplementary Figure 5, Kaplan-Meyer survival analysis showing the probability of DFS for MSS **(A)** and MSI **(B)** stage II CRC patients according to having received or not adjuvant chemotherapy.



Supplementary Figure 6. Western-blot analysis of SOX2 and CDX2 protein expression in LS174T WT and LS174T CDX2-KO (data obtained from a previous study done by the research group).



Supplementary Figure 7. Comparison between the apoptotic levels in LS174T WT and LS174T CDX2-KO cell lines. Cells were incubated 48h with 4µg/mL of 5-FU for 48h then double stained with annexin V and PI. Apoptosis analysis was performed by FACS. Results are mean ± SD. Significant differences (*p<0.05). Data obtained from a previous study done by the research group.



Supplementary Figure 8. Comparison between the apoptotic level in LS174T CDX2-KO normal and transfected cells with siRNA for SOX2. Cells were incubated 48h with 4µg/mL of 5-FU for 48h then double stained with annexin V and PI. Apoptosis analysis was performed by FACS. Results are mean ± SD. Significant differences (*p<0.05). Data obtained from a previous study done by the research group.

Supplementary Figure 9. Microscopical observation (magnification 50x) of SW620 treated with 5-FU versus DMSO, after SOX2 silencing and overexpression.