INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Isolation and characterization of cancer stem cells (CSCs) from colorectal cancer - transcription factors involved in their reprogramming

Inês Catarina Garcia Ribeiro

U.PORTO

Inês Garcia Ribeiro. Isolation and characterization of cancer stem cells (CSCs) from colorectal cancer - transcription factors involved in their reprogramming

M. ICBAS **2020**

U.PORTO

MESTRADO EM ONCOLOGIA

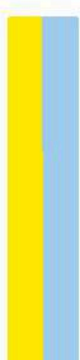
ONCOLOGIA LABORATORIAL

Isolation and characterization of cancer stem cells (CSCs) from colorectal cancer - transcription factors involved in their reprogramming

Inês Garcia Ribeiro

M

2020



Inês Garcia Ribeiro

ISOLATION AND CHARACTERIZATION OF CANCER STEM CELLS (CSCs) FROM COLORECTAL CANCER – TRANSCRIPTION FACTORS INVOLVED IN THEIR REPROGRAMMING

Dissertação de Candidatura ao grau de Mestre em Oncologia - Especialização em Oncologia Laboratorial, submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

Orientadora

Doutora Patrícia do Carmo Bidarra Reis e Mesquita

Investigadora – Differentiation and Cancer Group, Instituto de Investigação e Inovação em Saúde/ Instituto de Patologia e Imunologia Molecular da Universidade do Porto (i3S/Ipatimup)

Co-orientadora

Professora Raquel Maria da Silva Graça Almeida

Professora Afiliada – Faculdade de Medicina da Universidade do Porto

Investigadora Principal – Differentiation and Cancer Group, Instituto de Investigação e Inovação em Saúde/ Instituto de Patologia e Imunologia Molecular da Universidade do Porto (i3S/Ipatimup)

The work presented in this dissertation has been included in a mini review that has already been published: Pádua D, Figueira P, Ribeiro I, Almeida R, Mesquita P. The Relevance of Transcription Factors in Gastric and Colorectal Cancer Stem Cells Identification and Eradication. Front Cell Dev Biol. 2020;8:442.

Financial support

This study has received financial support by Fundação para a Ciência e a Tecnologia (FCT, I.P.) with grant number (PTDC/MED-ONC/29017/2017) – ReProgram: Reconstrução do Programa de Células Estaminais do Cancro. PI: Raquel Almeida. From 2018-07-01 To: 2021-06-30.

“The dictionary is the only place where success comes before work. Hard work is the price we must pay for success. I think you can accomplish anything if you're willing to pay the price.” - *Vince Lombard*

Acknowledgments

Chegada ao fim de uma das etapas mais importantes da minha vida e ao apresentar esta Dissertação de Mestrado, valorizo a importância e a influência que todos os envolventes tiveram ao longo destes dois anos. Assim, cabe-me agradecer àqueles que contribuíram para a minha formação pessoal e académica.

Ao Instituto de Ciências Biomédicas Abel Salazar, na pessoa da Professora Doutora Carmen Jerónimo, diretora do Mestrado em Oncologia, por lutar por cada um dos “seus meninos”, por me receber (mesmo quando o telemóvel não parava de tocar), por me apoiar e por ser uma Mentora importante na minha formação.

Ao instituto de investigação e inovação em saúde da Universidade do Porto, i3S, na pessoa da Professora Doutora Raquel Almeida, minha co-orientadora e Investigadora Principal do Grupo *Differentiation and Cancer* por me ter aceitado nos últimos dois anos neste grupo, pelo seu espírito crítico e rigor científico e por me ter apoiado e incentivado ao longo deste percurso.

À minha orientadora, Doutora Patrícia Mesquita, pelo seu apoio, por me ensinar e por exigir de mim sempre o melhor. Um ano e meio depois, posso dizer que sou uma pessoa mais capaz, mais elucidada e mais habilitada. Obrigada por ter sido fundamental no meu crescimento pessoal e profissional.

Ao grupo Differentiation and Cancer, pela disponibilidade e apoio. À Diana Pádua, por tantas vezes me ter ajudado e incentivado. À Mariana, pela amizade, pela ajuda, pelos sorrisos, pela companhia e por ter estado ao meu lado neste último ano, mostrando-me sempre mais uma razão para continuar. À Rita, pelos olhares que escusaram palavras, pela ajuda, pelos esclarecimentos, por ser uma menina brilhante e por se permitir partilhar isso com quem a rodeia.

À Professora Doutora Maria José Oliveira, por ter semeado em mim valores, curiosidades e sonhos que levarei sempre comigo, por ser uma inspiração e por todo o seu apoio e carinho. Nunca me vou esquecer de si, minha querida Professora.

À Professora Doutora Felisbina Queiroga, por exigir sempre de mim aquilo que sabe que consigo atingir: o que eu quiser. Por ter ensinado aquela menina de 20 anos a lutar afincadamente pelos seus sonhos e por, agora com 23, se certificar que o continuo a fazer.

À Professora Doutora Filomena Adega, por ter sido sempre tão humana comigo e por ter sido, e por continuar a ser, a impulsionadora de grandes sonhos e oportunidades que a vida me tem vindo a dar. Obrigada por me ir mostrando que as minhas capacidades

são do tamanho que eu quiser e por me ter munido de competências que levo para a vida. Estou certa de que os nossos caminhos se continuarão a cruzar.

Às pessoas mais importantes da minha vida: Cecília, Fernando, Miguel e João. Por serem o motivo maior pelo qual luto tanto e me tento sempre superar. Por serem a força que me faz ir e pela qual desejo tanto voltar. Por me ensinarem que a definição de casa é onde eles estiverem.

Aos meus pais, por me mostrarem que vale mesmo a pena lutar por aquilo que queremos e por me fazerem ver que, na vida, não há decisões certas nem erradas e que, no final, cresço sempre com elas. Por serem o melhor exemplo que tenho na vida e por me mimarem tanto. À minha mãe por lutar incansavelmente por mim, por me mostrar que a vida tem os limites que eu lhe quiser impor e por ter o abraço mais bonito do mundo. Ao meu pai, por viver tudo tão intensamente comigo, pela força e pelo carinho e por me fazer ver sempre mais uma razão para continuar. Ao meu irmão, por todas as partilhas, por me ensinar tanto, por exigir o melhor de mim, por ser o meu melhor amigo e por ficar sempre do meu lado.

Ao meu namorado, João, por todas as vezes que secou as minhas lágrimas, por me ajudar a enfrentar os meus medos e por os enfrentar comigo, pela boleia até ao laboratório ao fim-de-semana, pela espera no carro mesmo quando o «demoro no máximo duas horas» se converteram em quatro. Por ter lutado sempre ao meu lado nos últimos seis anos, por não me deixar cair e por ser tão carinhoso comigo. Por ser o meu orgulho e por demonstrar que sou, também, o seu. Por crescer comigo, por me mostrar que as coisas menos boas que nos vão acontecendo na vida nos fazem sempre crescer, por me fazer ver o lado positivo de cada situação e por ter acreditado em mim, mesmo quando eu não o consegui fazer. À Dona Brigite, ao Senhor Fernando e ao Pedro, por serem a minha segunda casa, por me apoiarem e mimarem tanto, pelos desabafos e por terem criado o Homem da minha vida.

A toda a minha família de sangue e do coração, pelo apoio, pela inspiração e por quererem sempre o melhor para mim.

À Rita, por ser ter sido o meu suporte nestes dois últimos anos. Por ser a amiga incrível que é, por me ter ajudado e amparado enquanto cresci, por me fazer ver o melhor lado de cada situação e por exigir sempre o melhor de mim. Por ouvir os meus desabafos e por me ter acompanhado em tudo. Acredito que na vida há coisas que não acontecem por acaso e esta escolha e a nossa amizade são a prova disso.

À Cristina, Joana, Kevin e Diogo, por serem os melhores amigos da vida, por me mostrarem que a distância não significa nada quando uma amizade significa tanto. Por estarem sempre na primeira fila a aplaudir de pé as minhas conquistas e por desejarem para mim, o mesmo que desejam para eles, o melhor.

À Ana, Céline, Francisca, Francisco e Diogo por serem a minha saúde mental. Por serem o melhor legado que podia ter deixado e por me mostrarem que a minha segunda casa precisa tanto de mim como eu dela.

À Ana, pelos desabafos, sermões, risos e lágrimas. Por ter sempre pronta a resposta certa e por me descomplicar tanto. À Céline por estar presente nas decisões mais importantes da minha vida e por estar do meu lado. À Francisca por dar sentido ao meu percurso e por partilhar comigo o seu. É um orgulho ser madrinha de amigas como vocês.

À Adriana e Beatriz por terem sido as melhores colegas de Mestrado. À Adriana, pelos desesperos conjuntos, pelos olhares que escusaram palavras e por fazer das minhas conquistas também as suas. À Beatriz, sempre pronta e disponível, pela amizade e paciência.

Aos meus engenheiros, Catarina, Tiago, Nunes, Manuel, Diogo, Beatriz, Afonso e Ricardo por me receberem tão bem, pelos convívios, pelas partilhas e pela amizade.

Ao JEOCHUA por terem aparecido no momento certo e por aumentarem a minha Fé. À Madalena, Elisabete, David, Carolina, Diogo, Lara e Matilde por me terem mostrado que a Fé não tem limites. À Madalena pelo voto de confiança e pela amizade. À Elisabete por todo o seu apoio, disponibilidade, carinho e por ter sempre as palavras certas. Ao David, por me ter apoiado incondicionalmente na fase mais desafiante da minha vida e pela compreensão. À Matilde, por ser a irmã que nunca tive, por me acompanhar em tudo e pelo coração e valores bonitos que tem. À Lara, pelo caminho inspirador que temos percorrido juntas e por ser um orgulho para mim.

À Bárbara, por continuar ao meu lado em todas as etapas da minha vida e por ser uma inspiração. À Áurea, pelo seu apoio, carinho, preocupação e por aumentar a minha fé. À Chefe Helena, pelos conselhos, desabafos, partilhas e por estar presente nos momentos mais importantes e especiais. Ao Monsenhor Miranda pela compreensão, amizade, carinho e sentido de justiça, por nunca ter duvidado de mim e das minhas capacidades. Está feito, Chefe!

Table of contents

Acknowledgments .... **VIII**
Abbreviations .... **XIV**
Figures Index .... **XVIII**
Tables Index .... **XX**
Resumo .... **XXII**
Summary .... **XXIV**
Chapter 1 | Introduction **1**
1 | Colorectal cancer 3
1.1 | Epidemiology 3
1.2 | Aetiology 4
1.3 | Diagnosis and Staging 6
1.4 | Treatment 8
2 | Cancer Stem Cells (CSCs) 9
2.1| Origin 9
2.2| Evidence 11
2.3| Identification using cell-surface markers 12
2.4| Colorectal CSCs 12
3| Transcription factors 14
3.1| SOX2 and OCT4 14
3.2| SOX family 15
3.3| Colorectal CSCs-related factors 17
3.4| CRC-related factors 19
3.5| SOX2-related factors 20
Chapter 2 | Hypothesis and Aims **21**
1.1 | Hypothesis 23
1.2 | Aims 23
Chapter 3 | Materials and Methods **25**
1| Transcription factors selection 27
2| SORE6-GFP cellular model 28
2.1| Cell culture 28
2.2| SORE6-GFP Fluorescence-activated cell-sorting 29
2.3 | Knockdown and overexpression of SOX2 29
3| Expression studies 30
3.1| Protein extraction, quantification and Western Blot 30
3.2| RNA extraction and Real-time PCR 31

3.3| Statistical analysis 33
4 | Cell Proliferation assay...33
5 | Colonosphere forming assay 34
6 | 5-FU resistance assay...34
7 | Co-immunoprecipitation and mass-spectrometry analysis 35
Chapter 4 | Results 37
1| Isolation and characterization of SW480 SORE6 subpopulations 39
1.1 | SOX2 and OCT4 expression 39
1.2 | Cell proliferation 40
1.3| Colonosphere formation 41
1.4| 5-FU resistance 41
1.5 | Cancer stem cell markers expression 42
2| Transcription factors expression analysis 43
2.1| SOX family 43
2.2| Colorectal CSCs-related...44
2.3| Colorectal cancer - related...46
3| SOX2 co-immunoprecipitation and mass-spectrometry 47
3.1| HMGA1 as a CSC marker 49
Chapter 5 | Discussion and conclusion 51
Chapter 6 | Future perspectives..59
Chapter 7 | References..63

Abbreviations

µm	*micrometers*
5-FU	5-fluorouracil
ABCB5	ATP-binding cassette B5
AJCC	The American Joint Committee on Cancer
ALDH1	aldehyde dehydrogenase 1
APC	adenomatous polyposis coli
ASR	Age-Standardized Rate
BSA	BSA
CAP	capecitabine
cDNA	complementary DNA
CEA	Carcinoembryonic antigen
CFE	colonosphere forming efficiency
CIMP	CpG Island methylator phenotype
CIN	chromosomal instability
c-MYC	MYC Proto-Oncogene, bHLH Transcription Factor
CO2	Carbon dioxide
CRC	Colorectal cancer
CSCs	Cancer Stem Cells
Ct	threshold cycle
DMEM	Dulbecco’s minimal essential media
DMSO	dimethyl sulfoxide
EMT	epithelial-mesenchymal transition
EOCRC	Early Onset Colorectal Cancer
EpCAM	epithelial cell adhesion molecule
ESCs	Embryonic Stem Cells
FACS	fluorescence-activated cell sorting
FAP	familial adenomatous polyposis
FBS	fetal bovine serum
FITC	FITC – Fluorescein isothiocyanate
FOXC2	Forkhead box protein C2
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
gFOBT	Guaiac Fecal Occult Blood Test
GFP	Green fluorescent protein
GLI1	GLI Family Zinc Finger 1

h	hour
HCl	Hydrochloric acid
HMG	High Mobility Group
HMGA1	High-mobility group A1
HMGA2	High-mobility group AT-Hook 2
IBD	Inflammatory Bowel Disease
IC50	half maximal inhibitory concentration
IMPDH	Inosine-5′-monophosphate dehydrogenase
iPS	induced pluripotent stem
IRI	irinotecan
KLF4	Kruppel like factor 4
Lgr5	Leucine-rich repeat-containing G-protein coupled receptor 5
MACS	magnetic-activated cell sorting
min	minutes
MSI	Microsatellite instability
Na3VO4	Sodium orthovanadate
NS	not significant
OCT4	octamer-binding transcription factor 4
OX	oxaliplatin
P	P-value
PBS	phosphate buffered saline
rpm	revolutions per minute
RT	room temperature
s	seconds
SALL4	Spalt Like Transcription Factor 4
SD	standard deviation
SDS-PAGE	dodecyl sulphate–polyacrylamide gel electrophoresis
siRNAs	small interfering RNAs
SOX	SRY related HMG-box
SOX2	Sex determining region Y-box 2
SSC-H	*side scatter height*
STAT3	Signal Transducer and Activator of Transcription 3
TBS	Tris buffered saline
TBS-T	Tris buffered saline with Tween 20
TCF4	T-cell factor 4
TEAD2	TEA Domain Transcription Factor 2

TFs	Transcription Factors
THY1	Thy-1 cell surface antigen
TNM	Tumor Node Metastasis
TSGs	tumor suppressor genes
WB	Western Blot
WT	*Wild Type*
β-actin	Actin, cytoplasmic 1
β-Catenin	Catenin Beta 1
μg	micrograms
μL	microliters

Figures Index

Figure 1| Colorectal cancer epidemiology a | Estimated incidence and mortality rates in 2018, worldwide, for both sexes at all ages **b|** Rates of CRC estimated incidence worldwide 3

Figure 2| Percentage of CRC cases related to sporadic and hereditary factors 5

Figure 3| Diagram representing the adenoma to carcinoma sequence 5

Figure 4| Schematic representation of the cancer stem cell model 10

Figure 5| Schematic representation of cancer stem cells origin 11

Figure 6| a| SW480 cells were transduced with the SORE6-GFP reporter **b|** SW480 cells were transduced with an empty vector 28

Figure 7| FACS analysis with gate selection and cell sorting of two cell subpopulations, SORE6+ and SORE6- cells 39

Figure 8| a| SOX2 expression by WB analysis **b|** SOX2 relative mRNA expression by real-time PCR analysis 40

Figure 9| Proliferation rate obtained by BrdU incorporation 40

Figure 10| Colonosphere-forming ability 41

Figure 11| 5-FU treatment and subsequent cell viability assessment using Presto blue 41

Figure 12| 5-FU treatment and subsequent cell viability assessment using Presto blue, after SOX2 knockdown 42

Figure 13| WB analysis of CSC surface markers CD44v6 and CD49f 42

Figure 14| WB analysis of CSC markers ALDH1 and c-MYC 43

Figure 15| SOX4, SOX8, SOX9, SOX12, SOX13 and SOX18 relative mRNA expression 44

Figure 16| KLF4, c-MYC, GLI1, β-catenin, TCF4, HMGA2, STAT3 and HMGA1 relative mRNA expression 45

Figure 17| FOXC2, TEAD2, CNOT3, HOXA10, MEIS2 and OCT1 relative mRNA expression.....46

Figure 18| WB analysis of transfection with an expression vector for SOX2 (pSOX2) and with an empty vector (pcDNA).....47

Figure 19| WB analysis of HMGA1.....49

Tables Index

Table I | CRC TNM Staging Classification. Data from AJCC Cancer Staging Manual, Eighth Edition 7

Table II | CSCs surface-markers for CRC 13

Table III | Overview of the TFs identified as related to colorectal CSCs or as SOX2-related in CRC or as a SOX2 co-factor 27

Table IV | Primary antibodies and conditions used in immunoblot procedure 31

Table V | Primers used for real-time PCR..32

Table VI | SOX2-associated proteins with an abundance ≥ 2 in SORE6+ in relation to SORE6- cells 48

Table VII | SOX2-associated proteins with an abundance ≤ 0.5 in SORE6+ in relation to SORE6- cells..48

Resumo

O cancro colorretal continua a ser considerado um grave problema de saúde, sendo o terceiro mais comum e a segunda causa de morte por cancro. A eficácia das soluções terapêuticas é limitada pela resistência aos fármacos. Evidências recentes mostraram que existe uma subpopulação de células que tem a capacidade de se autorrenovar - Células Estaminais de Cancro (CECs) – e é capaz de iniciar e suster a tumorigénese. As CECs possuem propriedades que as tornam clinicamente relevantes, uma vez que são responsáveis pela falha das terapias dirigidas ao cancro e pela sua recidiva. No cancro colorretal, o fator de transcrição (FT) SOX2 é expresso em cerca de 20% dos tumores e está associado a um nível mais baixo de sobrevivência dos doentes e a uma pior resposta à quimioterapia. Sugerimos que o SOX2 é um dos FTs envolvido na reprogramação de CECs e, por isso, tem sido usado como um ponto de partida para a identificação do sistema de sinalização e de regulação transcricional crítico para as CECs intestinais. Implementámos uma estratégia análoga à proposta por *Tang et al.* (2015) para separar células de cancro com base na atividade dos FTs SOX2 e OCT4. Para tal, fizemos a transdução da linha de cancro do cólon SW480 com um lentivírus que contém um plasmídeo na qual uma proteína verde florescente (GFP) está sob o controlo de um promotor que inclui locais de ligação para o SOX2 e o OCT4 (SORE6) e separámos as células por *Fluorescence-Activated Cell Sorting (FACS)* de acordo com a expressão da GFP. Foram estabelecidas duas subpopulações, uma muito enriquecida em células que expressam GFP e que exibe características de CECs - com resistência significativamente aumentada a fármacos quimioterápicos e uma maior capacidade para formar colonosferas - e outra muito escassa em células que expressam GFP. Surpreendentemente, o OCT4 não é expresso nas duas subpopulações enquanto que a expressão de SOX2 é comparável entre estas. Assim, tentámos identificar outros FTs que estivessem diferencialmente expressos nas nossas subpopulações e que, sendo capazes de se ligar aos locais de ligação do SOX2 ou diretamente à proteína SOX2, poderiam ser usados para explicar as diferenças fenotípicas encontradas nas células SORE6+ e SORE6-. Através de PCR em tempo real, a expressão de SOX18, STAT3, OCT1 e HMGA1 revelou-se significativamente aumentada nas células SORE6+, quando comparadas com as SORE6-. Adicionalmente, depois da co-imunoprecipitação e da análise por espetrometria de massa, o HMGA1 é uma das proteínas que interage com o SOX2 e que se revelou aumentada nas células SORE6+ *versus* SORE6-. O fator de transcrição HMGA1 tem sido constantemente associado na literatura a características de CECs e é, portanto, um bom candidato para ajudar a perceber as diferenças fenotípicas observadas no nosso modelo. Outros estudos são, agora, necessários para a elucidação do papel do HMGA1 na

reprogramação das CECs colorretais, bem como na interação com o SOX2 e respetiva regulação.

Palavras chave: Cancro colorrectal, Células estaminais de cancro, SOX2, Fatores de transcrição, Co-imunoprecipitação, Espectrometria de massa, HMGA1.

Summary

Colorectal cancer (CRC) remains a serious health concern, being the third most common and the second leading cause of cancer related deaths. The efficacy of therapeutic solutions is limited due to drug resistance. Emerging evidence has shown that a subpopulation of self-renewing cells - Cancer Stem Cells (CSCs) - is capable of initiating and sustaining tumorigenesis. CSCs possess properties that make them clinically relevant since they are responsible for cancer therapy failure and disease recurrence. In CRC, SOX2 transcription factor (TF) is expressed in about 20% of the tumors and is associated with worse patient survival and worse response to chemotherapy. We suggest that SOX2 is one of the TFs involved in CSCs reprogramming so it has been used as a starting point to identify critical signaling and transcriptional regulatory network in intestinal CSCs. We have implemented a strategy proposed by *Tang et al.* (2015) to sort cancer cells based on the activity of SOX2/OCT4 transcription factors. For that, we transduced SW480 colon cancer cell line with a lentivirus containing a Green Fluorescent Protein (GFP) under the control of a promoter containing SOX2/OCT4-responsive elements (SORE6) and sorted the cells by Fluorescence-Activated Cell Sorting (FACS), according to GFP expression. We have established two sub-populations, one that is highly enriched in GFP-expressing cells and exhibits features of CSCs - with both significantly increased resistance to chemotherapeutic drugs and higher capacity to form colonospheres - and the other one highly depleted in GFP-expressing cells. Surprisingly, OCT4 is not expressed in both cell sub-populations whilst the expression of SOX2 is comparable between them. Therefore, we tried to identify the TFs that can also bind to SOX2 binding sites or directly to SOX2 protein that could be expressed differently in our subpopulations and could be used to explain the phenotypical differences between SORE6+ and SORE6- cells. By real-time PCR, SOX18, STAT3, OCT1 and HMGA1 expression was seen significantly increased in SORE6+ cells, when compared to SORE6-. Additionally, after co-immunoprecipitation and mass spectrometry analysis, HMGA1 was one of the SOX2-interacting proteins that appeared increased in SORE6+ *versus* SORE6- cells. HMGA1 has been consistently linked to CSC features in the literature and it is therefore a good candidate to help explaining the different phenotypes observed in our model. Further studies are now necessary to elucidate the role of HMGA1 in colorectal CSCs reprogramming and in SOX2 interaction and regulation.

Keywords: Colorectal cancer, Cancer Stem Cells, SOX2, Transcription factors, Co-immunoprecipitation, Mass-spectrometry, HMGA1.

Chapter 1 | Introduction

1 | Colorectal cancer

1.1 | Epidemiology

Colorectal cancer (CRC) incidence and mortality considerably differ around the world. In 2018, The International Agency for Research on Cancer estimated that this cancer was the 3rd most commonly diagnosed and the 2nd most deadly cancer globally (figure 1a) and is predicted to increase to 2.5 million new cases worldwide in 2035 (1-3). Africa and South-Central Asia exhibit the lowest incidence rates, contrastingly to North America, Australia, New Zealand and Europe which have the highest incidence (figure 1b). These differences stem from a combination of various factors, such as lifestyle, genetics and life expectancy (3, 4).

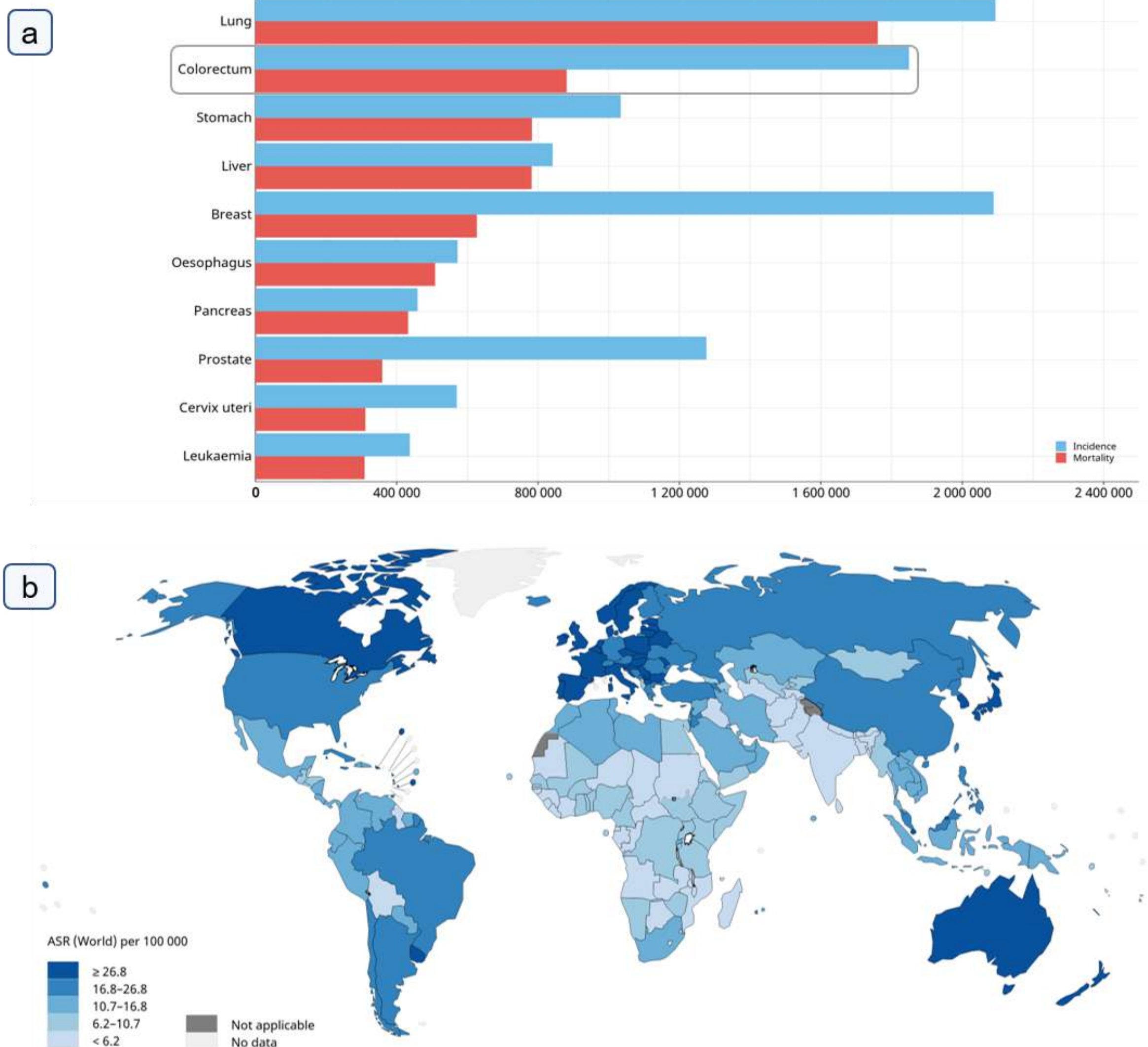


Figure 1| Colorectal cancer epidemiology. a | Estimated incidence and mortality rates in 2018, worldwide, for both sexes at all ages (ordered by mortality). **b|** Rates of CRC estimated incidence worldwide (2018). GLOBOCAN 2018, ASR- Age-Standardized Rate (cases per 100.000 persons per year).

Age-standardization identifies men as more likely to be affected by CRC than women, with an approximately 1.4 fold difference in incidence and nearly 1.5 fold difference for mortality (3).

CRC incidence is associated to both hereditary and environmental risk factors, with males presenting a higher vulnerability and exposure to environmental risk factors including dietary fat intake, visceral fat, diets including red and processed meats, smoking and, among others, alcohol ingestion. The risk of CRC is also increased when other diseases are present such as ulcerative colitis and Crohn's disease (5-9).

CRC development and death rates tend to rise promptly after the age of 50, for both sexes, with this age group representing approximately 90% of cases and deaths. However, there is an heterogeneous population of patients with CRC under 50 years of age - Early Onset Colorectal Cancer (EOCRC), which might be lost if they are not screened before that age (4). EOCRC may have distinct molecular characteristics when compared to older onset CRC and tends to reveal at a more advanced stage. Family history is primarily considered for risk stratification, being an important approach for the identification and early screening of individuals under 50 years (10-12). Despite family history of CRC, hereditary cancer syndromes and Inflammatory Bowel Disease (IBD) predispose individuals to EOCRC (13), the majority of these cases occur sporadically among individuals at average risk (14).

1.2 | Aetiology

The majority of CRC cases arise sporadically (60–65%) as somatic genetic and epigenetic alterations are acquired, mostly because of variable risk factors, whereas approximately 25% have a CRC family history (first-degree relatives with invasive CRC or colorectal adenomas and familial clustering) without any evident genetic cancer syndrome (4, 15, 16). Genetic susceptibility to CRC includes hereditary cancer syndromes comprising around 5% of CRC cases, as hereditary nonpolyposis colorectal cancer (ie, Lynch syndrome) or familial adenomatous polyposis (FAP) (4) (figure 2).

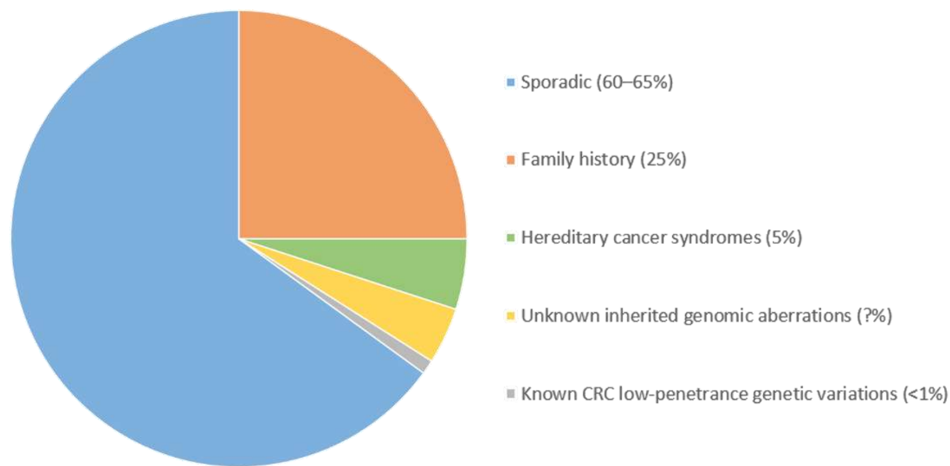


Figure 2| Percentage of CRC cases related to sporadic and hereditary factors. Adapted from (4)

The most common hereditary syndrome is Lynch syndrome, which accounts for 1-4% of all colon cancers and it is linked to germline mutations in MLH1, MSH2, MSH6, PMS1 and PMS2 DNA mismatch repair genes. FAP, the second most common, makes up to 1% of CRC cases and is usually triggered by mutations in the tumor suppressor gene, adenomatous polyposis coli (APC), which modulates the Wnt/β-Catenin signaling pathway (4, 8, 9, 17).

Globally, colorectal cancers arise through three different carcinogenic pathways: the traditional adenoma to carcinoma pathway, the inflammatory pathway and the serrated pathway (4). The adenoma-carcinoma sequence is the standard pathway that is responsible for the majority of CRC cases (18). This sequence comprises the progressive accumulation of both epigenetic and genetic modifications which drive normal cells to transform from small to large adenoma and, lastly, to cancer (figure 3). Inactivation of APC leads to an overactive Wnt/β-catenin signalling pathway, causing adenoma development and dysregulated cell proliferation (19). Subsequent mutations in KRAS oncogene result in adenoma growth. Further, inactivating the TP53 tumor suppressor gene also leads to CRC development (20).

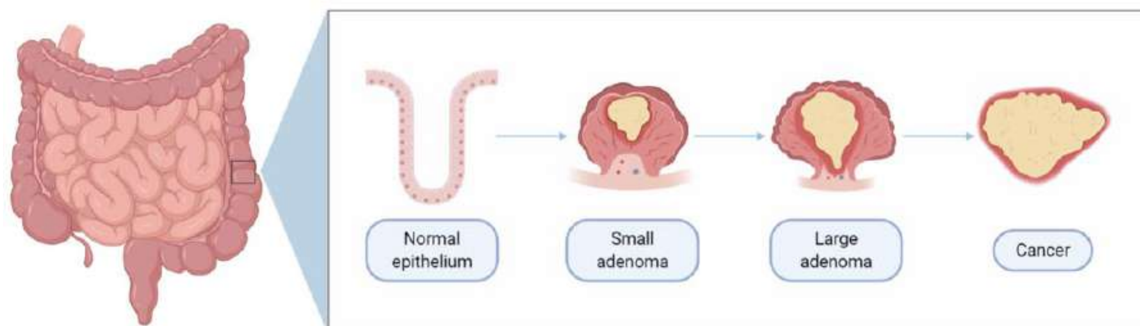


Figure 3| Diagram representing the adenoma to carcinoma sequence. This illustration was created in BioRender website.

In contrast to adenoma-carcinoma sequence, in the inflammatory pathway, TP53 mutations occur earlier, with APC mutations appearing frequently and later in the carcinogenesis process (21). The serrated pathway (mainly sessile serrated adenoma) consists in the modification of normal cells into hyperplastic polyps, to sessile serrated adenoma then, lastly, to CRC. Mutations in the BRAF oncogene leads to uncontrolled cell proliferation and contribute to the development of hyperplastic polyps with constitutive activation of the MAPK pathway (22, 23).

Three principal genetic and epigenetic mechanisms account for sporadic CRC: chromosomal instability (CIN), Microsatellite instability (MSI) and CpG Island methylator phenotype (CIMP) (24-26). CIN is an adenoma-carcinoma sequence model suggesting that oncogenes activation and mutations inactivating tumor suppressor genes (TSGs) induce CRC (4, 27, 28). MSI is comprised of alterations in repetitive DNA length in which certain DNA motifs are repeated several times - microsatellites - resulting from mutations in some DNA mismatch repair genes (29). CpG Island methylator phenotype is a form of epigenetic modification that involves CpG islands hypermethylation in the promoter of TSGs (i.e. MINT1, MINT2, MINT3 and MLH1,) silencing gene expression (30, 31). CIMP promotes subsequent progression into sessile serrated adenoma and CRC (22, 23).

In spite of having distinct characteristics, these mechanisms are not mutually exclusive (32). Remarkably, CIMP and MSI are associated as CIMP is a mechanism that induces high levels of MSI by the inactivation of DNA mismatch repair genes (30, 33) and about 70% of MSI-high CRC cases correspond to CIMP-high (32). During colorectal tumorigenesis, the molecular phenotypes referred are accumulated with CIN being observed in approximately 85%, CIMP-positive in 20% and MSI-high status in 15% of sporadic CRC (34).

1.3 | Diagnosis and Staging

Despite sustained advances in diagnostic and treatment, CRC undoubtedly contributes to numerous cancer-related deaths, remaining one of the most frequently diagnosed cancers (1). Considering that most of CRC cases take place sporadically, they are characterized by a sequence of carcinogenic events involving the progressive accumulation of acquired mutations, lasting 10 to 15 years on average (18, 35, 36). This long evolution interval would allow for early diagnostic, detection and removal of adenomas which could reduce CRC incidence and mortality (11, 36-38).

Despite the possibility of early diagnosis, CRC cases are identified at stage IV, about 20 to 25%, being characterized by distant metastasis and having less than 10% 5-year survival rate. Nevertheless, when surgical resection is possible, the 5-year survival rate increases to 90% for patients which present early localized disease (39, 40).

The American Joint Committee on Cancer (AJCC) Tumor Node Metastasis (TNM) staging classification is commonly used in CRC staging, providing valuable prognostic information and guiding CRC treatment (9, 41) (table I). In a CRC patient, T (extent of the tumor), N (spread to nearby lymph nodes) and M (spread to distant metastasis) are often settled by the examination of the resected tissue straight after resective surgery, whilst in clinical staging pre-surgical imaging tests, physical exams and biopsies are considered. This information is assembled to determine an overall stage of the tumor (American Cancer Society, 2018).

Table I | CRC TNM Staging Classification. Data from AJCC Cancer Staging Manual, Eighth Edition.

Tumor	**TX**	The primary tumor cannot be assessed
	T0	There is no evidence of primary tumor
	Tis	Carcinoma in situ, intramucosal carcinoma (involvement of lamina propria with no extension through muscularis mucosae)
	T1	Tumor invades the submucosa (through the muscularis mucosa but not into the muscularis propria)
	T2	Tumor invades the muscularis propria
	T3	Tumor invades through the muscularis propria into pericolorectal tissues
	T4a	Tumor invades through the visceral peritoneum (including gross perforation of the bowel through tumor and continuous invasion of tumor through areas of inflammation to the surface of the visceral peritoneum)
	T4b	Tumor directly invades or adheres to adjacent organs or structures
Node	**Nx**	Regional lymph nodes cannot be assessed
	N0	No regional lymph node metastasis
	N1a	One regional lymph node is positive
	N1b	Two or three regional lymph nodes are positive
	N1c	No regional lymph nodes are positive, but there are tumor deposits in the subserosa, mesentery, or nonperitonealized pericolic or perirectal/mesorectal tissues.
	N2a	Four to six regional lymph nodes are positive
	N2b	Seven or more regional lymph nodes are positive
Metastasis	**M0**	No distant metastasis by imaging, etc.; no evidence of tumor in distant sites or organs
	M1a	Metastasis to one site or organ
	M1b	Metastasis to two or more sites or organs
	M1c	Metastasis to the peritoneal surface is identified alone or with other site or organ metastases

The most recent AJCC system (January 2018) also combines the pathologic stage determined by examining the tissue removed during a surgery (surgical staging) towards a more precise diagnosis (9, 41).

Colonoscopy is considered the current gold standard screening strategy to diagnose CRC lesions. The guidelines recommend regular screening to those aged more than 45 years with a CRC average risk (11). However, colonoscopies have low patient adherence, not only because the procedure is expensive and invasive but also because it carries risks, namely colonic perforation, hemorrhage, and cardiorespiratory complications. (40). Guaiac Fecal Occult Blood Test (gFOBT), comprised of the hemoglobin peroxidase activity detection is a simple, economical and the most commonly non-invasive CRC screening method used, nevertheless, gFOBT may cause high rates of both false negative and positive results beyond low selectivity and sensitivity rates in the detection of colon polyps. (36, 42).

Therefore, non-invasive, cost-effective, accurate and easily measurable alternative screening tests are necessary in the detection of CRC. Clinical application of biomarkers in CRC should be explored, both for early detection and prognostic stratification, therapy selection and surveillance (43-45). Carcinoembryonic antigen (CEA), a glycoprotein found in 1965 and expressed in embryonic tissue and colorectal malignancies, accounts to be the most widely used prognostic biomarker in CRC clinical practice. Its high levels indicate cancer progression and possibly cancer recurrence after surgery. However, high CEA levels are not accurate, as they can be found in different inflammatory conditions and malignancies, namely IBD, liver disease and pancreatitis (46, 47). Predictive biomarkers are beneficial to the decision-making process for adjuvant and neoadjuvant therapy approaches, importantly optimizing the managing and follow-up of CRC patients, with proteomics representing a powerful strategy for personalized approaches (43, 48). Moreover, the identification of dysregulated proteins and its validation as disease biomarker candidates are often based in quantitative mass spectrometry techniques (49). Mass spectrometry capacity in the detection of low-abundance elements turns this tool effective for the exploration of possible biomarkers (48).

1.4 | Treatment

The mail goal of CRC treatment is to remove completely the tumor and its metastases, which involves surgical intervention (50). Despite all the programs established for prevention of CRC, approximately 25% of CRC cases are diagnosed at an advanced stage with metastases (4, 11, 51). When lesions are unable to be resected, the objective is

to reduce the tumor size as much as possible and inhibit its spread and growth, being radio and chemotherapy the common approaches for these patients, as neoadjuvant or adjuvant treatment (52-54).

Chemotherapy includes not only single-agent therapy, which is based on 5-fluorouracil (5-FU) used in patients at low risk of deterioration and with poor performance, but also combination approaches including one or various drugs like oxaliplatin (OX), irinotecan (IRI) and capecitabine (CAP). Combined therapies such as FOLFOX (5-FU+OX), FOXFIRI (5-FU+IRI), CAPOX (CAP+OX), and CAPIRI (CAP+IRI) persist for first-line treatment (55, 56). The determination of MSI, BRAF and KRAS status has turned to be essential in determining a therapeutic strategy, mainly in patients with metastatic disease (48).

In early stage disease or even in stage III patients who present lymph-node metastases, surgery is the mainstay treatment. Adjuvant chemotherapy helps patients with tumors in stage III and those with high-risk stage II cancer to achieve a better survival. In metastatic CRC, few patients benefit from surgical resection of metastases, as most of them present with metastases which are more advanced and, consequently, unresectable (56, 57). One of the greatest concerns in CRC postoperative follow-up is detecting recurrence. Subsequently to curative purpose surgery, 30 to 40% of these patients exhibit distant metastasis or locoregional recurrence (58). CEA remains the most commonly established protein biomarker in the clinic and is capable of determining prognosis. Patients with both postoperative CEA increment had the worst CRC prognosis (59). New prognostic biomarkers should, therefore, be identified, validated and translated for nodal status, distance metastasis, and post-surgical recurrence prediction (46, 47).

2 | Cancer Stem Cells (CSCs)

2.1| Origin

Due to a self-renewal ability, Cancer Stem Cells (CSCs) are cells with the capacity to originate, by symmetric divisions, daughter cells that maintain a stem cell phenotype and, by asymmetric divisions, more differentiated daughter cells which have the capacity to induce different types of cancer cells inside the tumor (60, 61). The CSC model suggests that the tumor parenchymal cells are hierarchically organized and hypothesizes the existence of a subpopulation of tumor cells which exhibit stem-like characteristics and have

self-renewal ability, initiating and sustaining tumorigenesis, metastasis, cancer relapse and resistance to chemotherapy and radiotherapy (figure 4) (62-66). The bulk of the tumor is composed by a more differentiated, phenotypically diverse although non-tumorigenic offspring (66).

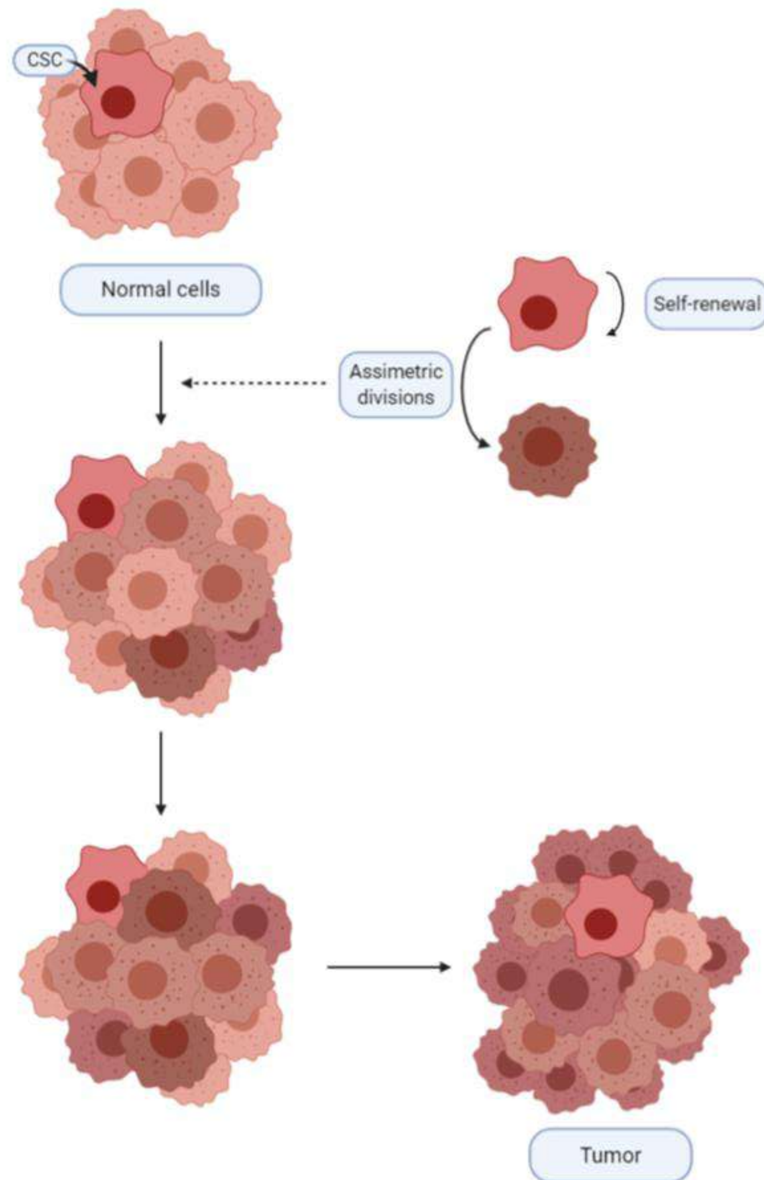


Figure 4| Schematic representation of the cancer stem cell model. CSC (Cancer Stem Cell). This illustration was created in BioRender website.

Although the CSC model enlightens the heterogeneity of cancer, in what concerns to the hierarchical structure and progression, the origin of CSCs is still uncertain and divisive (67, 68). Evidence suggests that the tumor type influences how CSCs may be derived from adult stem cells, mutated adult progenitor cells, or progenitor and differentiated cells that, through dedifferentiation, acquired stem-like characteristics. These cells can be

reprogrammed to regain self-renewal ability, less differentiated characteristics and a malignant phenotype derived from accumulating extra genetic or epigenetic mutations, therefore resulting in CSCs (figure 5) (69-71).

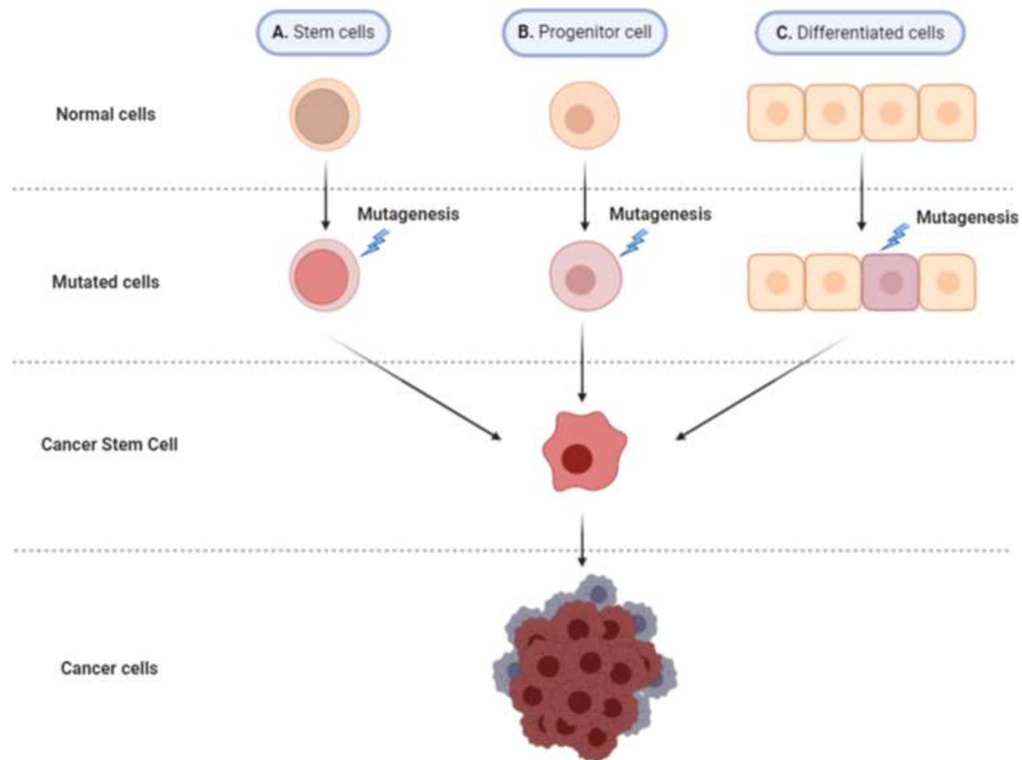


Figure 5 |Schematic representation of cancer stem cells origin. This illustration was created in BioRender website.

2.2| Evidence

CSCs, otherwise branded tumor initiating cells, have been explored for the possible origin, biomarkers, mechanism and possible therapeutic strategies for their targeting (60, 72). The first *in vivo* evidence of CSCs presence in solid tumors was observed in 2003 in immunocompromised mice after the transplantation of human breast cancer cells and it was identified as $CD44^{+}CD24^{-/low}$ cells (73). Regardless, in 2002, CSCs had been already registered *in vitro* through the detection of sphere-forming cells from human brain gliomas (74). Throughout the years, CSCs populations were found in other solid tumors including brain, melanoma, liver, lung, breast, pancreas, ovarian, and colon (73, 75-81).

On one side, CSCs are commonly identified through combinations of cell-surface markers, important for the isolation of CSCs from solid and hematological tumors through fluorescence-activated cell sorting (FACS) analysis. On the other side, this isolation can be made through the conjugation of antibodies with magnetic beads (MACS). With FACS the

purity of isolation is generally higher yet with MACS a higher cell survival is observed (64, 82).

2.3| Identification using cell-surface markers

Various cell surface markers namely CD24 molecule (CD24), CD44 molecule (CD44), prominin 1 (CD133), CD200 molecule (CD200), Thy-1 cell surface antigen (THY1), epithelial cell adhesion molecule (EpCAM) and ATP-binding cassette B5 (ABCB5) have been used in the identification of CSC enriched populations (73, 75, 81, 83). The cell surface markers used to isolate CSCs might vary between cancer types depending on their phenotypes and properties (71). Notably, it is still difficult to monitor CSCs in the tumor *in situ*. Moreover, this approach is limited by the inability of its usage for the assessment of CSC behavior in real-time and at a single cell instead of at a population level. This restricts the use of preclinical models, in which direct observation of CSCs and monitorization of single cells behavior in time and space, would lead to a novel look into CSCs properties and therapeutic response (66).

2.4| Colorectal CSCs

In what concerns CRC, various cell surface-markers such as CD44, Integrin alpha-6 (CD49f), CD133, CD200 and their combinations (i.e. CD44 with EpCAM, CD44 with CD133, CD133 with CD166 antigen (CD166), Leucine-rich repeat-containing G-protein coupled receptor 5 (Lgr5) with EpCAM and finally, Lgr5 with EpCAM and CD44) have been used as targets in CSCs identification and isolation (81, 83-92) (table II).

Furthermore, ALDH1 (aldehyde dehydrogenase 1), an intracellular protein, is used as a CSC marker in many types of cancer such as liver, leukemia, breast, lung, pancreas and colon, among others (84, 85). ALDH1 high expression has been associated with poor differentiation, cancer relapse and metastasis acting as a mediator of drug resistance in colorectal CSCs. Also, in rectal cancer, preoperative radio-chemotherapy induced ALDH1 expression (86, 87).

Table II | CSCs surface-markers for CRC

CSCs cell-surface markers	Condition in CRC	References
CD44	Expression significantly correlated with poor overall survival and a higher expression associated with poor differentiation, lymph node and distant metastasis	(88, 89)
CD44 in combination with **EpCAM**	Better strategy when compared to CD44 in combination with CD133	(83)
CD44 in combination with **CD133**	Enriched in chemoresistant cells and colonospheres allowing CSCs isolation from cell lines and primary tumors. Improvement in discrimination between low and high-risk cases	(89)
CD133	Expression associated with CRC cell differentiation and tumor size and correlated with a significantly worse survival and poorer clinical response to 5-FU-based chemotherapy	(89)
CD133 in combination with **CD166**	Better stratification of CRC risk when compared to CD133 alone	(83)
CD49f	Expression associated with tumor cell invasion and metastasis in colon cancer	(90)
CD166	Expression correlated with pathogenesis, being an early event in colon carcinogenesis and is also used to copurify CSCs	(91)
CD200	CSC marker in colon cancer	(92)
Lgr5	Expression related to tumorigenesis, 5-FU resistance and CRC recurrence. In stage IV, high expression is associated with poor prognosis	(93, 94)
Lgr5 in combination with **EpCAM**	Improve CSCs identification	(95)
Lgr5 in combination with **EpCAM** and **CD44**		(96)

Moreover, some Transcription Factors (TFs) that are overexpressed in stem cells contribute to CSCs self-renewal characteristics and are therefore related with patient prognosis, being a powerful tool in CSCs study and identification (97-99). Together with the development of reporter systems that have the ability to actively monitor certain intracellular markers, they appear a more efficient way to study CSCs than the use of cell-surface markers (66, 85, 100).

3| Transcription factors

3.1| SOX2 and OCT4

SOX2 (Sex determining region Y-box 2), OCT4 (octamer-binding transcription factor 4) and NANOG (Homeobox protein NANOG) are considered the master transcriptional regulators that maintain a stem cell phenotype in Embryonic Stem Cells (ESCs). These three genes are involved in the promotion of stemness by the upregulation of genes implicated in self-renewal and pluripotency whilst suppressing genes involved in cell differentiation (101). Moreover, abnormal expression of SOX2, OCT4 and KLF4 (Kruppel like factor 4) transcription factors in differentiated somatic cells is enough to induce stem-like characteristics and pluripotency, suggesting that the reactivation of stem cell TFs could be a competent mechanism for these cells to acquire the capacity to self-renew (66, 102).

SOX2 is a member of the SOX family of TFs, characterized by a transcriptional activation domain and a High Mobility Group (HMG) domain with DNA binding ability. It is encoded by the SOX2 gene that is located in chromosome 3q (3q26.33). SOX2 plays an important role in pluripotency, cell fate determination and, during development, in cell differentiation in normal tissues such as the lung, esophagus and trachea foregut-derived epithelia. SOX2 is also crucial for the maintenance of ESCs (103-108). In 2006, Takahashi and Yamanaka demonstrated a role for SOX2 in stemness by the introduction of a set of TFs (SOX2 along with OCT3/4, c-Myc and KLF4) in mouse embryonic fibroblasts that were able to reprogram into induced pluripotent stem (iPS) cells (109).

SOX2 expression is found in adult immature cells of several self-renewing epithelial tissues and it has been detected in cells from numerous malignant tissues, some of which effectively rely on SOX2 for their tumor-initiating ability (110, 111). Anomalous expression of SOX2 has been increasingly associated with a CSC phenotype and reported in various types of cancer from different tumor tissues, namely skin, prostate, lung, central nervous system, gastric and colorectal. At the clinical level, in several epithelial cancers, SOX2 has been demonstrated to act as an oncogene and as a powerful marker of poor prognosis (112-120).

In CRC, SOX2 has been associated to cell migration, invasion and metastasis and it was suggested as a regulator of epithelial-mesenchymal transition (EMT). SOX2 *de novo* expression has been related to more invasive and poorly differentiated tumors (121, 122). In 2018, Takeda et al. have shown that $SOX2^{+}$ cells exhibit asymmetric cell division and a higher expression of CSC markers and acquire chemoresistance to 5-FU and oxaliplatin, concluding that colon cancer cells with SOX2 expression behave like CSCs and are

therefore associated with disease recurrence and with poor overall survival and prognosis (115, 122).

OCT4, a homeodomain transcription factor that belongs to the Pit-Oct-Unc family, is expressed in ESCs where it maintains stem-cell characteristics and is also involved in proliferation and differentiation of adult stem cells (123, 124). Ectopic expression of OCT4 in the intestinal epithelium and epidermis blocks differentiation and leads to progenitor cells uncontrolled proliferation (125). OCT4 expression is related with more aggressive and metastatic tumors and it was associated with poor overall prognosis in tumors such as melanoma, medulloblastoma, acute myeloid leukemia, hepatocellular carcinoma, esophageal squamous cell carcinoma, bladder, prostate, glioma, ovarian, gastric and colorectal (126-136).

In CRC, OCT4 expression has been reported in cells undergoing EMT, a crucial step in tumor progression and metastasis which maintains stem cell-like properties and its knockdown inhibited cell migration and invasion (135). OCT4 can directly reprogram adult stem cells to iPS cells and it has also been pinpointed in colorectal CSCs as a prognostic marker (97, 114, 135, 137).

In 2015, Tang et al. developed a stem cell lentiviral-based reporter system in which six tandem repeats of a composite SOX2/OCT4 response element were used to drive the expression of a Green Fluorescent Protein (GFP) reporter. These authors have shown that in more malignant breast cancer cell lines, the mRNA levels of SOX2 and OCT4 were higher, and the SORE6-GFP reporter was capable of identifying a marginal population of cells - CSCs - with higher expression of these TFs, with expected self-renewal and tumor-initiating ability and also higher resistance to chemotherapeutics (66). This reporter system made the direct observation of CSCs plasticity possible, allowing other investigators to use it in other types of cancer (116). In 2020, with the use of SORE6-GFP reporter system, our group was able to isolate gastric CSCs from two gastric cancer cell lines (Kato III and AGS) (116).

3.2| SOX family

The mammalian SOX (SRY related HMG-box) family of TFs comprises 20 proteins and is defined by at least a 50% sequence homology to the SRY HMG-box domain by which they mediate the binding to DNA consensus motifs (138-140). SOX TFs affect diverse tissue systems and processes as they are greatly implicated in developmental disorders and in various types of cancer (107, 140-145).

SOX proteins are important transcriptional regulators which control cell fate decisions and which have the capacity of instructing maintenance, induction or inhibition of cellular states (107, 140). Overexpression of SOX factors can promote pluripotent stem cells differentiation into multiple types of cells and can drive the direct reprogramming of somatic cells. This can be observed when additional TFs are present, reflecting the prevalent interdependency of SOX factors in what concerns to partnerships with tissue-specific transcription factors. Nevertheless, in some cases the overexpression of a SOX factor alone can drive the transcriptional cascades required to completely reprogram somatic cell fate (108, 146-153).

In what concerns clinical decisions, SOX family proteins can be valuable prognostic markers (154). Some transcription factors belonging to the SOX family are upregulated in CRC and their expression have been associated with poor prognosis (155,170)

SOX4 (SRY-Box Transcription Factor 4) is crucial to lymphocyte differentiation and endocardial development and its overexpression in different types of cancer is correlated with angiogenesis and resistance to chemotherapy (155-157). SOX4 expression has been described as significantly increased in CRC resected tissues and highly expressed in CRC cell lines. Its depletion inhibited colon cancer cells migration, invasion and colony formation (namely in SW480 colon cancer cell line) and reversed the EMT process. SOX4 overexpression contributed to CRC development and progression and was correlated with tumor metastasis (158).

SOX8 (SRY-Box Transcription Factor 8) has been suggested as being involved in organ development, tissue specification, stem cell homeostasis and cancer development, being a regulator of cell proliferation through the Wnt/β-catenin signalling pathway. In CRC, SOX8 overexpression was described to be clearly associated with invasion and used to predict poor prognosis (159).

In the gastrointestinal tract, SOX9 (SRY-Box Transcription Factor 9) has been described to be associated with cell fate and, during embryonic development and adulthood, with maintenance of stem cells (160-162). In CRC, SOX9 oncogenic activity has been implicated in tumor proliferation and progression mostly due to the regulation of a CSC pool. High expression of SOX9 was associated with a worst 5-year survival rate and therefore with a poor prognosis (163-165).

SOX12 (SRY-Box Transcription Factor 12) has been associated with a more aggressive CRC phenotype and its overexpression related to CRC cell proliferation, invasion, migration and metastasis (in SW480 cell line), being therefore associated with

malignant progression, disease recurrence and poor patient outcome and prognosis (166, 167).

A significantly higher SOX13 (SRY-Box Transcription Factor 13) expression has been found in CRC tissues compared to normal tissues and its overexpression was correlated with EMT progression, invasion, proliferation, metastasis and therefore tumorigenesis. It was also associated with aggressive clinicopathological features and with poor prognosis (168).

In CRC tissues and cell lines, SOX18 (SRY-Box Transcription Factor 18) expression has been significantly increased and correlated with invasion, lymph node and distant metastasis demonstrating its role in tumor progression. SOX18 was associated with poor prognosis in these patients (169).

3.3| Colorectal CSCs-related factors

KLF4 has been described with a higher expression in undifferentiated cells where it is implicated in the cell cycle and pluripotency. KLF4 has also been described as necessary for the self-renewal capacity maintenance (97, 170). In CRC, KLF4 role is still unclear. Some studies had demonstrated that in colon CSCs KLF4 was overexpressed and its knockdown had an impact in the stemness phenotype and decreased the cells malignancy. However, other studies revealed that KLF4 loss of expression was related to stem-like characteristics such as the formation of colonospheres, uncontrolled cell proliferation, cell growth arrest, self-renewal and pluripotency (97, 170-172).

STAT3 (Signal Transducer and Activator of Transcription 3) plays a crucial role in the regulation of several physiological functions, in inflammation and in tumor proliferation and invasion. It has been described as highly expressed in colorectal CSCs. STAT3 is considered a major oncogenic protein and an important therapeutic target for CRC, being involved in angiogenesis and chemoradiotherapy resistance (173-177).

c-MYC (MYC Proto-Oncogene, bHLH Transcription Factor) has been reported in the regulation of genes that are involved in self-renewal, growth, differentiation and metabolism, which determines the epithelial stem cells characteristics in colon tissues. Although being described as one of the most common activated oncogenes, its overexpression alone is not capable of inducing normal cells transformation into tumor cells (178, 179). In CRC, c-MYC upregulation has been found in approximately 70% of cases, being consistently overexpressed in colon CSC and showing its role in self-renewal and chemoresistance

maintenance. Also, despite several controversial results, high expression of c-MYChas been correlated with an independent poor prognosis (180, 181).

GLI1 (GLI Family Zinc Finger 1) has been described as an essential TF that maintains CRC stem like properties in cancer cells. Moreover, GLI1 knockdown downregulated CD133/SOX9 expression and CRC cells clonogenic ability. GLI1 was associated with metastatic spread of CRC cells and aggressiveness, therefore resulting in a survival decrease, indicating this TF as a potential CSC marker in CRC (182, 183).

SALL4 (Spalt Like Transcription Factor 4) has been associated with ESCs self-renewal and in pluripotency maintenance being silenced or downregulated in differentiated cells. Moreover, SALL4 is involved in tumorigenesis and in tumor initiation and progression. In CRC, SALL4 overexpression is related with tumor metastasis and with a poor prognosis being essential in the maintenance of CSCs properties (184, 185).

β-Catenin (Catenin Beta 1)*, a* transcription factor activated by the canonical Wnt signaling has been involved in the regulation of epithelial stem cell self-renewal and dysregulation of this pathway has been implicated in colon carcinogenesis, playing an important role in regulating colonospheres maintenance and growth. Wnt/*β-Catenin* activation could lead to the conversion of intestinal stem cells into CSCs. β-catenin high nuclear levels in CRC tumors were related with a poor prognosis and it might be used as a late phase CRC biomarker. Alone, β-catenin signaling has revealed necessary for the maintenance of CSC features (186-189). Furthermore β-catenin and TCF4 (T-cell factor 4) interaction has been implicated as a core component of the Wnt/*β-Catenin* pathway in CRC cell growth (190).

TCF4 has been related as an important factor in regulating stemness and drug resistance in CRC and its overexpression promoted stemness in colorectal CSCs, correlating with the expression of CSC markers, as CD133, and the ability to form tumorspheres. TCF4 has been described as an independent prognostic variable of poor clinical outcome in CRC patients, thus suggesting TCF4 as a potential predictive marker in colon cancer cases diagnosed at low stages and a possible target for CRC therapy (191-193).

HMGA gene family comprises HMGA1 (High-mobility group A1) and HMGA2 (High-mobility group AT-Hook 2) that are much expressed through embryonic development, yet in differentiated tissues have lower or undetectable levels (194-196). HMGA1 aberrant expression has been described in most poorly differentiated cancers (also in colon) and its high levels were correlated in diverse tumors with poor outcomes (195, 197-199). Its expression is related to CSCs characteristics. HMGA1 may induce SOX9 and requires the

Wnt signaling to reprogram intestinal epithelial cells and to drive stem cell properties (196, 200-202). Moreover, in human ESCs, HMGA1 is capable of inducing SOX2 (196).

HMGA2 has been associated with multiple biological processes as DNA repair, cell cycle, apoptosis and EMT (203). In CRC, its high expression was associated to tumor progression, poor response to therapy and therefore to a poor prognosis (204). Furthermore, HMGA2 overexpression enhanced resistance to 5-FU chemotherapy suggesting that the observed HMGA2-mediated chemoresistance was mediated through the activation of the Wnt/β-catenin pathway (205).

3.4| CRC-related factors

FOXC2 (Forkhead box protein C2) has been associated with cell proliferation, angiogenesis and metastasis in several tumors (206, 207). FOXC2 high expression has been associated to CRC development and its downregulation to an increase in the apoptotic rate of CRC cells suggesting FOXC2 as an effective therapeutic target (208).

TEAD2 (TEA Domain Transcription Factor 2) role in CRC is still not yet well defined even though it was described as belonging to the intestinal stem cell signature predicting disease relapse (209).

OCT1 (Octamer transcription factor 1), a homologous of the OCT4 TF, is required to the maintenance of stem cells and to embryogenesis (210, 211). It has shown to play an important role in CRC development, cell proliferation, migration, invasion, tumorigenesis and induction of EMT. OCT1 overexpression was also associated with CRC distant metastasis and it is therefore considered a promising target for CRC therapy (212, 213).

CNOT3 (CCR4-NOT Transcription Complex Subunit 3) overexpression was associated to self-renewal and to more aggressive CRC cases, being related to poor clinical outcome (214).

HOXA10 (Homeobox A10) has been implicated in cell proliferation, migration and invasion of CRC cells. Its expression has been described as significantly increased in CRC tumor tissues, being crucial in CRC development and indicating resistance to 5-FU. HOXA10 may be used as a biomarker of poor prognosis (215).

MEIS2 (Meis Homeobox 2) transcription factor has been associated with EMT, migration, invasion and metastasis in CRC and may be used as a CRC biomarker (195) (216, 217). However, it seems that its downregulation is involved in stemness and in resistance to oxaliplatin-based chemotherapy (196).

3.5| SOX2-related factors

Complex gene networks that are required to attain the pluripotency and self-renewal of ESCs are controlled by master transcriptional regulators such as SOX2, which associate with co-repressors and co-activators to accurately control their target genes. These interactions with co-factors provides SOX2 with a better functional flexibility, during development. Furthermore, differences in the master regulator-protein complexes composition clearly influence their target genes transcriptional activity and some studies using proteomic analysis have determined several co-repressors and co-activators which associate with SOX2. Moreover, some authors had determined that SOX2 uses distinct and multiples domains to associate with its multiple partner proteins. These findings were tested by mapping the SOX2 domains that were necessary to co-immunoprecipitate SOX2-associated proteins. Together, these studies had provided novel insights about SOX2 ability to associate with a wide range of nuclear proteins that control the transcription of genes (218, 219).

Chapter 2 | Hypothesis and Aims

1.1 | Hypothesis

We hypothesize that cancer stem cells are controlled by a limited set of key transcription factors whose expression can impose CSCs properties in cancer cells.

1.2 | Aims

The aims of this project were to isolate and characterize intestinal CSCs and to identify the TFs and major signaling networks associated with a CSC phenotype and with SOX2, in CRC.

Chapter 3 | Materials and Methods

1| Transcription factors selection

Besides performing a bibliographic search to identify TFs related to colorectal CSCs, available gene expression datasets generated from colorectal primary tumors were analyzed by our group. Tumor samples were separated based on SOX2 expression, into SOX2-high and SOX2-low, and gene expression of TFs were compared between these samples, in order to find SOX2-related TFs (table III).

Table III | Overview of the TFs identified as related to colorectal CSCs (bibliographic search) **or as SOX2-related in CRC** (by data mining analysis) **or as a SOX2 co-factor** (Biogrid https://thebiogrid.org/112540/table/homo-sapiens/sox2.html)

Gene	Colon CSCs related	SOX2 related in CRC	SOX2 cofactor (Biogrid)
SOX4	J	J	
SOX8			J
SOX9	J		
KLF4	J	J	
STAT3	J		
c-MYC	J	J	
GLI1	J		
SALL4	J		
beta catenin	J	J	J
TCF4	J		J
HMGA1	J	J	
HMGA2	J		
FOXC2		J	
OCT1		J	J
HOXA10		J	
MEIS2		J	J

2| SORE6-GFP cellular model

In the present study, SW480 human colon cancer cell line (ATCC) that has been previously transduced using lentiviral supernants with the SORE6-GFP reporter system by our group, were used (66, 116). The lentiviral reporter construct - SORE6-GFP - has a SOX2/OCT4 response element from the proximal human NANOG promoter repeated six times (SORE6) and joined to a minimal cytomegalovirus (CMV) promoter to drive the expression of a fluorescent protein reporter (GFP).

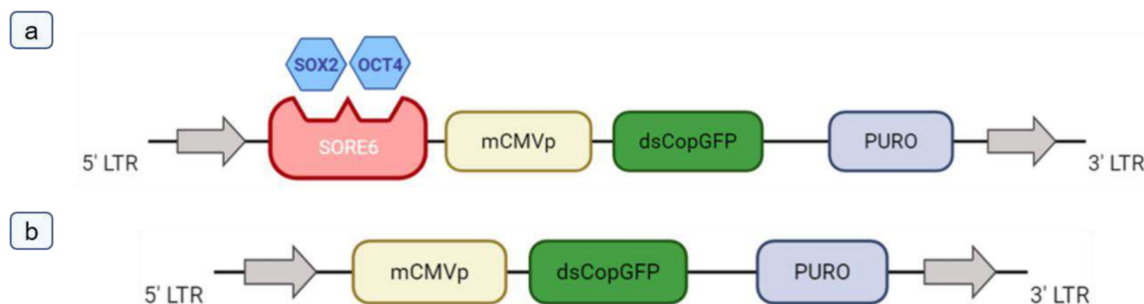


Figure 6 | a| SW480 cells were transduced with the SORE6-GFP reporter (SORE6 - six repeats of a composite SOX2/OCT4 response element) b| SW480 cells were transduced with an empty vector (without SORE6).

2.1| Cell culture

Cells were grown and maintained as a monolayer in Dulbecco's minimal essential media (DMEM) without phenol red (Gibco, Life Technologies) supplemented with 10% (v/v) inactivated fetal bovine serum (FBS) (Biowest), 1% Sodium Pyruvate (Gibco, Life Technologies) and 1% (v/v) antibiotics - 100 U/ml penicillin and 100 µg/ml streptomycin (Thermo Fisher Scientific, Massachusetts, USA). They were maintained at 37°C in a humidified 5% CO_2 incubator, under standard conditions.

When monolayers achieved approximately 80% confluence, trypsinization was done. In order to detach adherent cells from a culture flask, culture medium was collected and discarded, cells were washed with phosphate buffered saline (PBS) 1x (Grisp Research solutions, Oporto, Portugal), 0.05% trypsin-EDTA (ThermoFisher Scientific) was added and cells were incubated at 37°C in a 5% CO_2 humidified atmosphere until detachment. Trypsin-EDTA was neutralized by the addiction of culture medium and a cell pellet was obtained by centrifugation at 1200 rpm (revolutions per minute) for 5 min (minutes) at RT (room temperature). Afterwards, culture medium was collected and discarded, the cell pellet was re-suspended in complete media, transferred into a sterile tissue culture flask and incubated at 37°C in a 5% CO_2 humidified atmosphere.

For cryopreservation, cell pellets were re-suspended in FBS containing 10% (v/v) DMSO (dimethyl sulfoxide) (AppliChem, Barcelona, Spain), transferred into a cryovial and

incubated at -80°C. To defrost cryopreserved cells, cryovials were gathered from -80°C and cells were immediately re-suspended in pre-warmed (37ºC) complete media and centrifuged at 1200 rpm for 5 min at RT. Supernatant was collected and discarded, cell pellets were re-suspended in complete media with 20% FBS, transferred into a sterile tissue culture flasks and incubated at 37°C in a 5% CO_2 humidified atmosphere.

To count cells, a Neubauer chamber was used (Marienfeld Superior™ Counting Chamber; Marienfeld, Lauda-Königshofen, Germany). Trypan blue at 0.4% (ThermoFisher Scientific) was used to distinguish viable from non-viable cells. This solution penetrates in cells once the membrane is damaged and when observed under microscope, viable cells exclude dye and dead cells internalize trypan blue. Cell suspension was diluted 1:1 in trypan blue solution and 10 µl of cell mixture were placed in a Neubauer chamber. Cells were counted under a Leica DMi1 inverted phase contrast microscope (Leica Microsystems, Wetzlar, Germany) at 50x magnification.

2.2| SORE6-GFP Fluorescence-activated cell-sorting

By trypsinization, SW480 SORE6-GFP cells were collected, washed with PBS and resuspended in PBS containing 10% FBS. By using a FACS Aria II Cell Sorter flow cytometer (BD Bioscience, Franklin Lakes, NJ, USA) SW480 SORE6-GFP cells were sorted into cells with and without GFP expression - SORE6+ and SORE6- cells, respectively. They were seeded in six-well plates at a density of 7.5×10^4 cells/ml for SORE6- cells and 6.5×10^4 cells/ml for SORE6+ cells. Negative control cells transduced with an empty reporter (Ø) viral supernatant were used for gating purposes. Data was analyzed by FlowJo software v10.0.7 (Oregon, USA).

2.3 | Knockdown and overexpression of SOX2

In order to silence SOX2 expression, $1.5x10^5$ of SW480 SORE6+ and SORE6- cells were seeded in each well of a 24-well plate and transiently transfected the next day by using Lipofectamine 2000 1:1 (Thermo Fisher Scientific, Massachusetts, USA) with either a mix of three small interfering RNAs (siRNAs) targeting SOX2 or a non-targeting scrambled siRNA used as a negative control (Thermo Fisher Scientific, Massachusetts, USA). The siRNAs and the Lipofectamine were diluted in Opti-MEM medium (Thermo Fisher Scientific, Massachusetts, USA) prior to the utilization.

For SOX2 overexpression, $3x10^5$ SW480 SORE6+ and SORE6- cells were seeded in each well of 6-well plates and transfected with 1 µg of a human SOX2 expression vector containing two FLAG tags at the N-terminus or the corresponding empty vector

(pcDNA3.1) in a ratio of 1:1.5 to Lipofectamine 2000 reagent, both previously diluted in Opti-MEM medium, as recommended by the manufacturer (Thermo Fisher Scientific, Massachusetts, USA).

Cells were incubated for 48h after which the efficiency of the SOX2 silencing/overexpression was evaluated by Real-Time PCR and Western Blot (WB).

3| Expression studies

3.1| Protein extraction, quantification and Western Blot

SW480 SORE6+ and SORE6-cells were washed with cold PBS 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, USA) supplemented with Complete protease inhibitor cocktail (Roche Applied Science, Indianapolis, USA), 1 mM Na_3VO_4 and 1 mM PMSF. Cells were scrapped and the mixture was centrifuged at 4°C for 15min at 14000 rpm. The soluble proteins concentration was evaluated using Pierce BCA protein Assay Kit (Thermo Fisher Scientific, Massachusetts, USA) according to the instructions of the manufacturer. 30 µg of total protein extract were used for WB analysis.

Proteins were separated in a 10% sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane (Amersham, GE Healthcare, UK) which was labelled with *Ponceau Red* (Sigma) to monitor the efficiency of the transfer. Then, the membrane was incubated with the blocking solution for 1h and then blotted overnight at 4°C with a primary antibody (conditions in table IV) with gentle agitation. The following day, the membrane was washed 3 times with TBS-1% Tween-20 to remove the unbound primary antibody. Afterwards, the membrane was incubated for 1h at RT with gentle agitation with a secondary antibody: anti-mouse IgG HRP-linked antibody (A4416, 1:4000; ThermoFisher Scientific) diluted in 5% non-fat milk in TBS-1% Tween-20 or anti-rabbit IgG antibody (1:10000, Santa Cruz Biotechnology) diluted in TBS-T and washed 3 times with TBS-1% Tween-20.

The loading controls used were GAPDH (glyceraldehyde-3-phosphate dehydrogenase) and β-actin (Actin, cytoplasmic 1). Membrane was first incubated overnight at 4°C with gentle agitation with the primary antibody for GADPH or β-actin, then, after 3 consecutive washes, it was incubated for 1h with the anti-mouse IgG HRP-linked secondary antibody (A4416, 1:4000; ThermoFisher Scientific) diluted in 5% non-fat milk in TBS-1% Tween-20.

Table IV | Primary antibodies and conditions used in immunoblot procedure.

Protein	Animal	Origin	Molecular weight (kDa)	Dilution		Incubation conditions
ALDH1	Rabbit	Cell Signaling Technology	55	1:1000	Low fat milk 5%	4°C overnight with gentle shaking
CD44V6	Mouse	Life technologies	100			
CD49F	Rabbit	Sigma-Aldrich	125			
GAPDH	Mouse	Santa Cruz Biotechnology	37			
-actin	Mouse	Santa Cruz Biotechnology	42-44	1:4000		
SOX2	Rabbit	Cell marque	34	1:500		
C-MYC	Rabbit	Cell Signaling Technology	57 - 65	1:700	BSA 5%	

After incubation the membrane was washed three times with TBS-T and protein bands were visualized using GE Healthcare Amersham™ ECL™ Western Blotting Detection Reagents (GE Healthcare, Illinois, USA) for the detection of chemiluminescence, according to the instructions of the manufacturer. Detection was achieved in ChemiDocTM XRS+ system (Bio-Rad Laboratories) and protein bands were evaluated using Image Lab Software (Bio-Rad Laboratories).

3.2| RNA extraction and Real-time PCR

Cells from each well were lysed in 300µL of Lysis Buffer, containing 1:100 β-mercaptoethanol, for approximately 30 min 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 instructions. This protocol includes a PureLink DNase Treatment, in which DNA is removed from RNA, which remains bound to a Spin Cartridge. After extraction, the purified RNA was stored at -80°C. RNA concentration was measured using a Nano-Drop 1000 instrument (Thermo Fisher Scientific, Massachusetts, USA). RNA (1µg) was reverse transcribed to cDNA using the SuperScript IV (Thermo Fisher Scientific, Massachusetts, USA), according to the instructions of the manufacturer. The genes were amplified with SYBR Green (Thermo Fisher Scientific, Massachusetts, USA) using the primers detailed in table V. Real-time PCR was performed in a fluorescence reader ABI Prism 7500. 4µL of cDNA diluted 1:10 in DEPC-treated water were used as template for each real-time PCR with 10µL *SYBR Green* and 0.3µM of specific

forward (F) and reverse (R) primers, in a final volume of 20µL. The following Real time-PCR protocol has been applied, consisting in a denaturation step at 95 ºC for 10 min, a 40-cycle stage, including denaturation at 95 ºC for 15s and annealing at 60ºC for 1 min, and a melting curve program (60-95ºC) with continuous fluorescence measurement. The amount of mRNA was determined using the threshold cycle (Ct) values and ΔΔCt method (220). The levels of 18S and HPRT1 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.

Table V | Primers used for real-time PCR

Gene		Sequence
18S	F	5' CGCGCGCTAGAGGTGAAATTC 3'
	R	5' CATTCTTGGCAAATGCTTTCG 3'
HPRT1	F	5' GCAGACTTTGCTTTCCTTGGTCAG 3'
	R	5' GTCTGGCTTATATCCAACACTTCGTG 3'
OCT4	F	5' CTTCTCGCCCCCTCCAGGT 3'
	R	5' GCTCACCCTGGGGTTCTATTT 3'
SOX1	F	5' GGAAAATACTGGAGACGAACGC 3'
	R	5' GCTTTCCGCTTCCTCCGTA 3'
SOX2	F	5' AACGGCTCGCCCACCTACAGC 3'
	R	5' AGTGGGAGGAAGAGGTAACC 3'
SOX3	F	5' GGACTGTTGCCTTGTACCGA 3'
	R	5' CAGCGATTCCCAGCCTACAA 3'
SOX4	F	5' GACTTCGAGTTTGCTCCCCT 3'
	R	5' TAACTCGCCTTCTTGCTGGG 3'
SOX6	F	5' GCGCTTTGAGAATTTGGGGC 3'
	R	5' GCCATTCATTGCTTTACTTCCCT 3'
SOX7	F	5' GGCCAAGGACGAGAGGAAAC 3'
	R	5' CCTTCCACGACTTTCCCAGC 3'
SOX8	F	5' AAGCTGTGGCGCTTGCT 3'
	R	5' GCTGGTACTTGTAGTCGGGG 3'
SOX9	F	5' CGGAGGAAGTCGGTGAAG 3'
	R	5' CTGGGATTGCCCCGAGTGCT 3'
SOX10	F	5' CTCTGGAGGCTGCTGAACGA 3'
	R	5' GGCCTTCCCGTTCTTCCG 3'
SOX12	F	5' TACCTGTATCTCACCGGCGT 3'
	R	5' CTGGGACCGTTTGTTCCTGA 3'
SOX13	F	5' TTCCCTGCAAACCAGTGGAG 3'
	R	5' GGCTTGGCTGTGAGGTTCAG 3'
SOX17	F	5' CCAAGGGCGAGTCCCGTAT 3'
	R	5' CCACGACTTGCCCAGCATC 3'
SOX18	F	5' CAAGATGCTGGGCAAAGCG 3'
	R	5' GCGGCCGGTACTTGTAGTT 3'
SOX21	F	5' GAATGTATAGGTGCCAGGTAGAG 3'
	R	5' AACCGCCTGCTTTCGAGTT 3'
KLF4	F	5' CAGAGGAGCCCAAGCCAAAG 3'
	R	5' TTTCTCACCTGTGTGGGTTCG 3'
STAT3	F	5' GAGGACTGAGCATCGAGCA 3'
	R	5' CATGTGATCTGACACCCTGAA 3'
C-MYC	F	5' GCGACTCTGAGGAGGAACAAG 3'
	R	5' TGGGCTGTGAGGAGGTTTG 3'
GLI1	F	5' GAAGTCATACTCACGCCTCGAA 3'
	R	5' CAGCCAGGGAGCTTACATACAT 3'
β-catenin	F	5' CAATGGCTTGGAATGAGACTGC 3'
	R	5' GGATCATCCTGGTGATATCCAAG 3'

TCF4	F	5' TTTGGAAGAAGCGGCCAAGAGG 3'
	R	5' TTGGGGAGGTAGGGGCTCGT 3'
FOXC2	F	5' CTACAGCTACATCGCGCTCATCA 3'
	R	5' ACTGGTAGATGCCGTTCAAGGTG 3'
MEIS2	F	5' TCCAGCATCTCACACATCCG 3'
	R	5' GAAAACCTGCTCGATTTGACTGG 3'
TEAD2	F	5' TTTGGGGTGTGCCCAGATG 3'
	R	5' TCCTCACTGCCTTCCTCACT 3'
OCT1	F	5' CAAAATGGCGGACGGAGGA 3'
	R	5' GTTCATTCTTGAGTCTGCTGCTG 3'
CNOT3	F	5' GGCTCACGAATACCATCGACA 3'
	R	5' GCTTATCCTTGTCGCCCTTCT 3'
SALL4	F	5' CAGCACATCAACTCGGAGGA 3'
	R	5' TTCACTGGAGCACCCAGC 3'
HOXA10	F	5' CCCTTCCGAGAGCAGCAAA 3'
	R	5' TCTTCCGACCACTCTTTGCC 3'
HMGA1	F	5' GTGCCAACACCTAAGAGACCT 3'
	R	5' TCTGCTGGTTTTCCGGCTC 3'
HMGA2	F	5' CCCAAAGGCAGCAAAAACAA 3'
	R	5' GCCTCTTGGCCGTTTTTCTC 3'

3.3| Statistical analysis

Results were expressed as mean ± SD. Statistical significance to assess significant differences in expression levels between the SORE6+ and SORE6- samples was determined by unpaired two-tailed t test using GraphPad Prism 8.0 software (GraphPad Software, La Jolla, CA, USA) and Microsoft Excel (Microsoft Corporation, 2018). Differences were considered significant when $P \leq 0.05$.

4 | Cell Proliferation assay

SW480 SORE6 GFP+ and GFP- cells were seeded in 24-well plates at a density of 1.5×10^5 cells/ml in DMEM (1X) without phenol red, supplemented with 10% FBS and 1% Sodium Pyruvate. After 48h, cells were treated with BrdU (Roche) diluted 1:1000 in the same medium, for 1 h at 37 °C. Cells were collected by trypsinization, pelleted, washed with PBS and fixed in 1 ml of ice-cold methanol for 30 min. After washing with PBS, cells were permeabilized with 1 ml of HCl 4M for 20 min at room temperature and washed with PBS. After a blocking step in PBS containing 0.5% Tween-20 and 0.05% BSA for 10 min, cells were incubated with a monoclonal mouse primary antibody anti-BrdU (Cell Signaling) diluted 1:20 in the blocking solution, for 1h at RT. After washing again with PBS, cells were incubated with the polyclonal rabbit anti-mouse secondary antibody labelled with FITC

(Dako) diluted 1:150 during 30 min at room temperature and protected from light. After a final washing step, the pellet was resuspended in 500 µl of PBS and the percentage of BrdU labeled positive cells was evaluated using the BD Accuri C6 flow cytometer (BD Biosciences). The data was analyzed using FlowJo software (Version 7.6.1).

5 | Colonosphere forming assay

Coating of 6-wells plates with a hydrophobic surface (non-adherent conditions) was performed by adding to each well 2ml of a sterile PolyHEMA solution, done by dissolving 12g of Poly (2-hydroxyethyl methacrylate) in 1 liter of 95% ethanol in a hot plate (< 50ºC). Afterwards the 6-well plates were left overnight at 54ºC in an incubator without CO_2.

The colonosphere medium was prepared by using 50ml of phenol red-free DMEM medium was supplemented with 1ml of B27 supplement, 0.5ml of N2 supplement, 10µl of hEGF (100ng/ul), 5 µl of bFGF (100ng/ul) and 500 µl of Penicillin/Streptomycin, afterwards the medium was filtered through a 0,45 µm filter.

SW480 SORE6 GFP+ and GFP- cells were resuspended in 2ml of cold PBS and passed through a 25 gauge needle using a syringe 3 times in order to separate cells into a single cell suspension. Cell suspensions were plated at a density of 5000 cells/well containing the colonosphere media and incubated in a 37˚C humidified atmosphere and 5% CO_2 for 5 days without replenishing the media. The colonosphere forming efficiency (CFE) was assessed after 14 days by counting the number of colonospheres that appear with ≥ 100µm diameter.

6 | 5-FU resistance assay

24h after transfection done to down-regulate SOX2, SW480 SORE6+ and SORE6- cells were treated with a 5-FU dose - 56.9 µg/mL - corresponding to the previously determined IC_{50} for SW480. After 48h, cells were washed once with PBS 1x and then 50µl of PrestoBlue Viability Reagent 1x (Invitrogen, California, USA), 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. The results are expressed as means ± SD of representative triplicates.

7 | Co-immunoprecipitation and mass-spectrometry analysis

After being resuspended by vortexing, 50 µL of Dynabeads magnetic beads (Invitrogen) were transferred to different tubes. The tubes were placed on the magnet and supernatants were removed in order to separate the beads. The tubes were removed from the magnet and 10 µg of the Flag (Abnova, PAB0900) antibody (Ab) diluted in 200 µL PBS with Tween-20 were incubated with the beads for 10 min with rotation at RT. Then, the tubes were placed on the magnet, the supernatants were discarded, the tubes were removed from the magnet and the magnetic beads-Ab complex was resuspended in 200 µL PBS with Tween-20 and washed by gentle pipetting. The tubes were placed on the magnet, the supernatants were removed and the samples containing respectively 200µg of protein extracted from SORE6+ or SORE6- cells previously transfected with SOX2 expression vector (as in 2.3) were added to the magnetic beads-Ab complex and gently pipetted. Then, an incubation was performed overnight with rotation and at 4°C. In the next day, tubes were placed on the magnet, the magnetic beads-Ab-antigen (Ag) complex was washed 3 times by gentle pipetting with 200 µL of PBS and by removing the supernatant. The magnetic beads-Ab-Ag complex was resuspended in 100 µL PBS and transferred to clean tubes to avoid co-elution of proteins possibly bound to the tube walls. These samples were kept at -80°C until mass spectrometry analysis.

The two samples, concerning SORE6+ and SORE6-, were reduced, alkylated, and later digested with trypsin. The mass spectra were obtained by LC-MS. Protein identification was performed by Proteome Discoverer (Thermo Scientific) software and Uniprot protein sequence database was used for Homo sapiens taxonomic selection as well as a mass spectra database. It was also considered a database of contaminant proteins. Protein quantification was performed using the LFQ - Label Free Quantification. This protocol was performed by the i3S Proteomics platform.

Chapter 4 | Results

1| Isolation and characterization of SW480 SORE6 subpopulations

When SW480 cells transduced with SORE6-GFP were sorted using FACS (fluorescent activated cell sorting) it was possible to distinguish two subpopulations: SORE6+ and SORE6- cells. The SORE6+ cells subpopulation, corresponding to a total of 3.84% of the cells, is capable of activating GFP expression while the SORE6- cells subpopulation, without GFP expression, corresponds to a higher percentage of cells (76.7%) (figure 7).

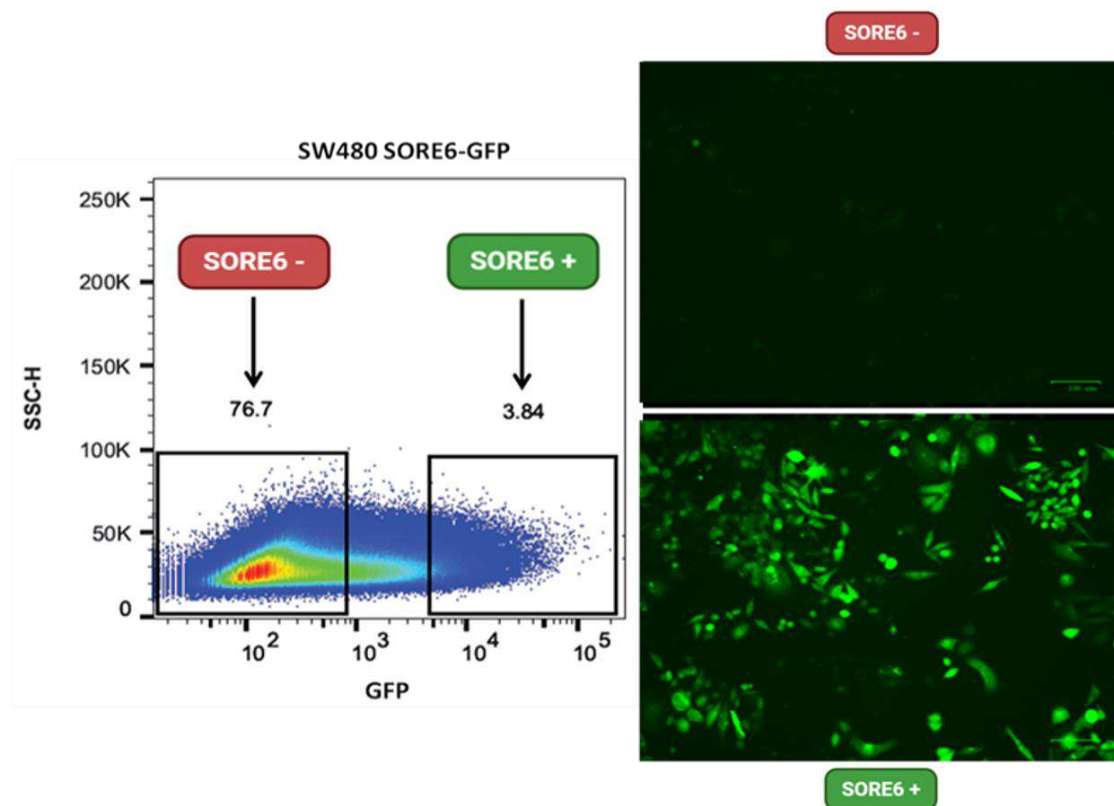


Figure 7 | FACS analysis with gate selection and cell sorting of two cell subpopulations, SORE6+ and SORE6- cells in SW480 SORE6-GFP cell line with further fluorescence images of the two subpopulations sorted (ZOE Fluorescent Cell Imager, Bio-Rad). Scale bar = 1

The two cell subpopulations, SORE6+ and SORE6-, were characterized for different parameters such as SOX2 and OCT4 expression, cell proliferation capacity, ability to form colonospheres, drug resistance after 5-FU treatment and expression of CSC markers.

1.1 | SOX2 and OCT4 expression

The expression of SOX2 and OCT4 was tested and evaluated in the SORE6+ and SORE6- subpopulations. A very low expression of OCT4 was detected by real-time PCR, corresponding to CTs (cycle threshold) ≥ 33.5.

In what concerns SOX2, and contrarily to what was expected, no difference was found in its expression between SORE6+ and SORE6- cells, when evaluated by both WB and real-time PCR in cells collected at different timepoints (14, 21, 25 days) after sorting. The housekeeping control gene, 18S was used for target gene abundance normalization and data was analyzed by the comparative method ΔΔCT (figure 8).

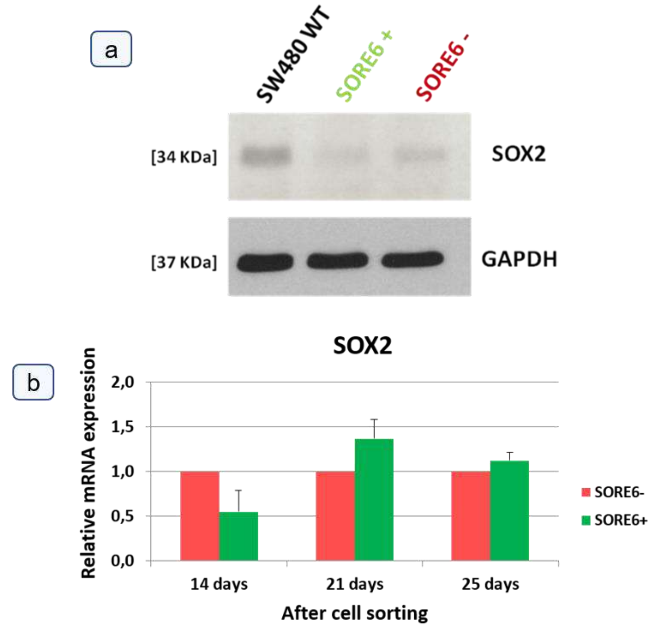


Figure 8 | a| SOX2 expression by WB analysis in FACS-sorted SORE6+ and SORE6- cells and (WT) wild-type SW480 cell line. The internal control used was GAPDH. **b| SOX2 relative mRNA expression by real-time PCR analysis** at different timepoints after sorting.

1.2 | Cell proliferation

Despite not observing differences in the expression of these TFs, between SORE6+ and SORE6- cells, contrary to what was expected, cells were further characterized. The proliferation of both subpopulations was evaluated through BrdU incorporation, used to identify the actively dividing cells. The proliferation rate (% of Brdu positive cells) was similar for SORE6+ (47.9 ± 3.9%), SORE6- (49.4 ± 5.7%) and the corresponding WT SW480 (48.4 ± 3.0%) cell line (figure 9).

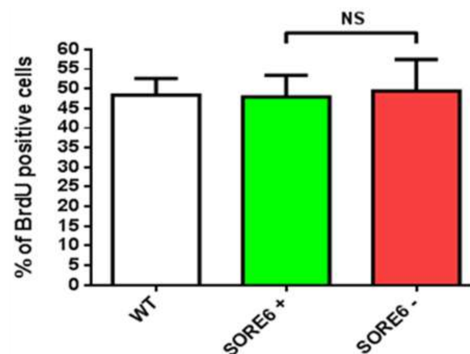


Figure 9 | Proliferation rate obtained by BrdU incorporation in SW480 WT, SORE6+ and SORE6- cells. NS: not significant

1.3| Colonosphere formation

Another important characteristic of normal stem cells and CSCs is the self-renewal ability to proliferate and to form and develop sphere-like structures. Cells were grown in serum-free medium under non-adherent conditions as only CSCs are capable of surviving and proliferating in this environment. In these conditions, the SORE6+ cell subpopulation was significantly enriched in cells that were capable of generating a higher number of colonospheres (2.04±0.36 times more), when compared to the SORE6- cell subpopulation (figure 10).

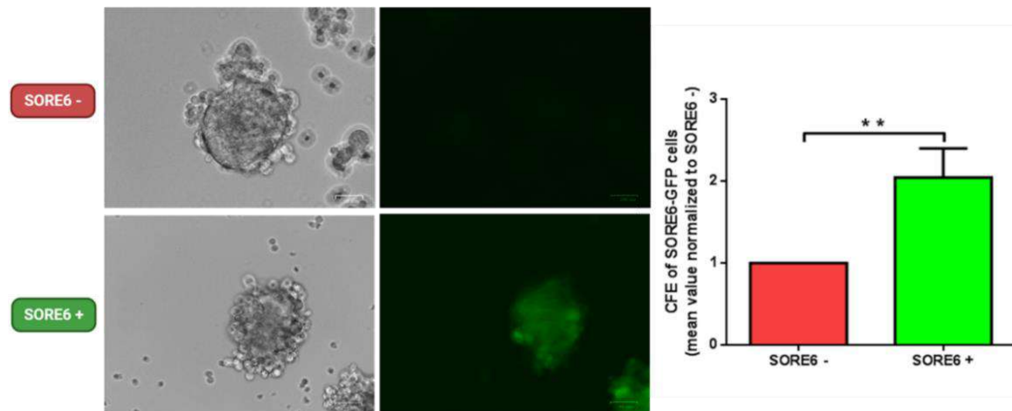


Figure 10 | Colonosphere-forming ability of SW480 SORE6- and SORE6+ cells. Spheres with only >100 µm diameter were considered as colonospheres. Scale bar SORE 6- cells = 100µm and SORE6+ cells = 64 µm. The results are mean ± SD of three independent experiments. Significant differences (**$P \leq 0.01$).

1.4| 5-FU resistance

CSCs are known to be more resistant to chemotherapeutic drugs leading to significant clinical implications as tumor recurrence and 5-FU is still considered the gold standard of CRC first-line treatment (221). SORE6-GFP cell subpopulations were treated with 5-FU and the cell viability was assessed. SORE6+ cells have shown a significantly higher viability and, therefore, more resistance to 5-FU treatment, when compared to SORE6- cells and the respective WT SW480 cell line (figure 11).

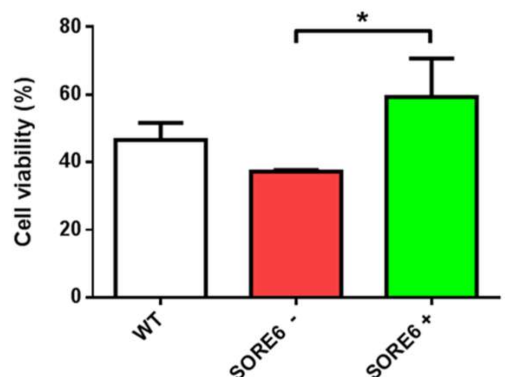


Figure 11 | 5-FU treatment and subsequent cell viability assessment using Presto blue in SW480 WT, SORE6- and SORE6+ cells. The results are mean ± SD of three independent experiments. Significant differences (*$P \leq 0.05$).

SOX2 knockdown was performed by transient transfection with SOX2 siRNA *versus* a non-targeting scrambled siRNA as a negative control. After SOX2 knockdown, cells were incubated with 5-FU and there was a decrease in the SORE6+ (0.65 ± 0.12) and SORE6- (0.68 ± 0.17) cells viability ratio (considering the viability of the same cells transfected with scrambled siRNA as 1), therefore suggesting that SOX2 has a role in therapy resistance, although being expressed in the two subpopulations (figure 12).

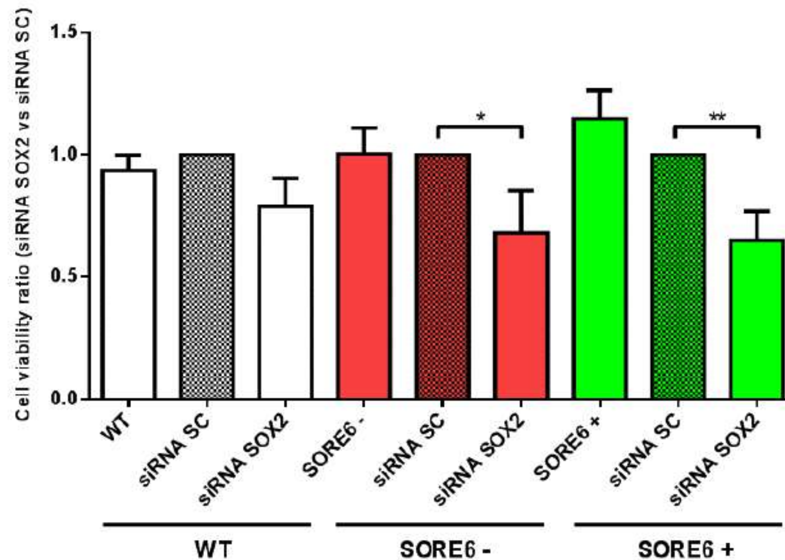


Figure 12 | 5-FU treatment and subsequent cell viability assessment using Presto blue, after SOX2 knockdown in SW480 WT, SORE6- and SORE6+ cells. The results are mean ± SD of three independent experiments. Significant differences (*$P \leq 0.05$; **$P \leq 0.01$).

1.5 | Cancer stem cell markers expression

WB technique was performed to evaluate the expression of certain CSC surface markers such as CD44v6 and CD49f, used as targets in CSCs identification and isolation and proven in the literature to be expressed in SW480 cell line (222, 223). Both CD44v6 and CD49f CSC surface markers were expressed in the two subpopulations and no difference between the expression in SORE6+ and SORE6- cells was found (figure 13).

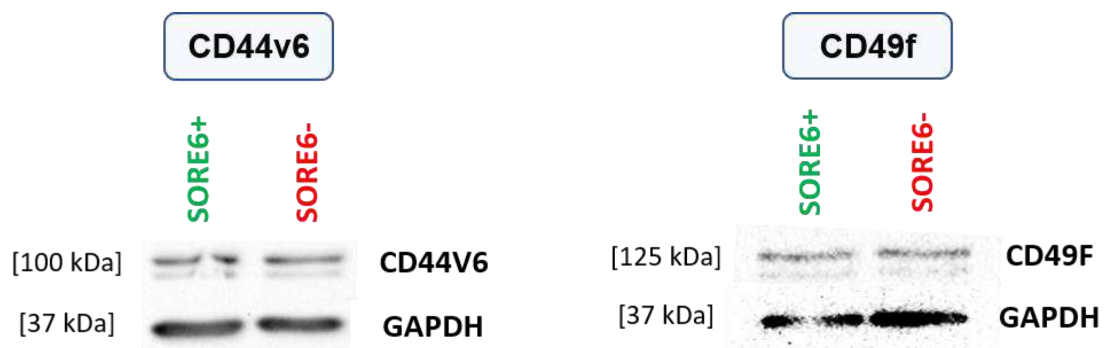


Figure 13 | WB analysis of CSC surface markers CD44v6 and CD49f on sorted SW480 cell subpopulations. The internal control used was GAPDH.

Other CSC markers such as ALDH1 have been already reported with a low expression and c-MYC was described to be expressed in SW480 cell line (224, 225). WB was performed and no expression of ALDH1 was found in both subpopulations, but it was expressed in the chosen positive control (A549 lung carcinoma cell line) concluding that ALDH1 is not expressed in our SW480 cell subpopulations. In what concerns to c-MYC, it was expressed in SORE6+ and SORE6- cells but no difference was found between its expression in the two subpopulations (figure 14).

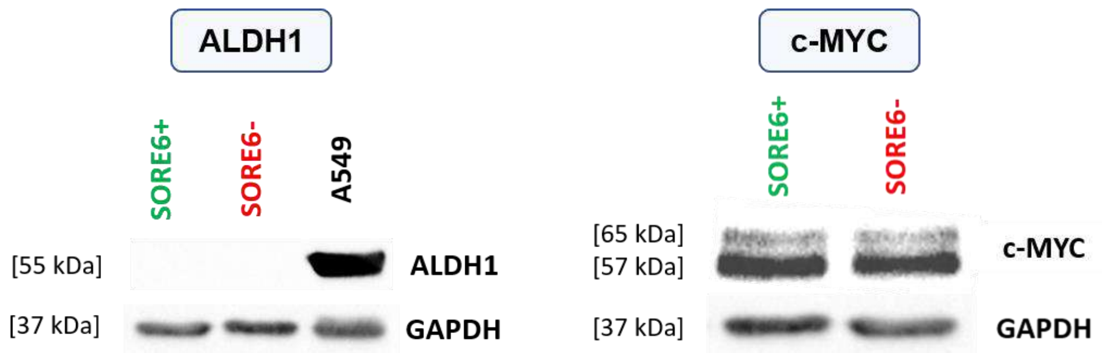


Figure 14 | WB analysis of CSC markers ALDH1 and c-MYC on sorted SW480 cell subpopulations. The internal control used was GAPDH.

In summary, it was possible to conclude that none of these CSC markers were enriched in the SORE6+ subpopulation from SW480 colon cancer cell line.

2| Transcription factors expression analysis

Since both SORE6+ and SORE6- have similar levels of SOX2 expression but SORE6+ reveals activation of GFP and characteristics of a CSC phenotype, we hypothesized that other transcription factors could be expressed in SORE6+, that can bind to the SOX2 binding site in SORE6 or directly to SOX2 protein and activate GFP expression. This assumption could help to explain the differences previously observed in the SORE6+ *versus* the SORE6- cells, such as increased ability in colonospheres formation and 5-FU resistance.

2.1| SOX family

The expression of TFs of the SOX family such as SOX1, SOX3, SOX4, SOX6, SOX7, SOX8, SOX9, SOX10, SOX12, SOX13, SOX17 and SOX18 was evaluated by real-time PCR in SORE6+ and SORE6- cells collected at different timepoints (14, 21, 25 days)

after the sorting. The housekeeping control gene, 18S was used for target gene abundance normalization and data was analyzed by the comparative method ΔΔCT.

SOX1, SOX3, SOX6, SOX7, SOX10, SOX17 and SOX21 were not expressed in SORE6+ and SORE6- cells (amplification occurred at later cycles - CTs ≥ 30 - and the melting curves were not satisfactory). No significant differences were found in the expression of SOX4, SOX8, SOX9, SOX12, and SOX13 between the SORE6+ and SORE6- cells. However, SOX18 relative mRNA expression was significantly higher in SORE6+ cells when compared to SORE6- cells but only for cells collected 25 days after sorting (1.180 ± 0.006; *P*=0.015) (figure 15).

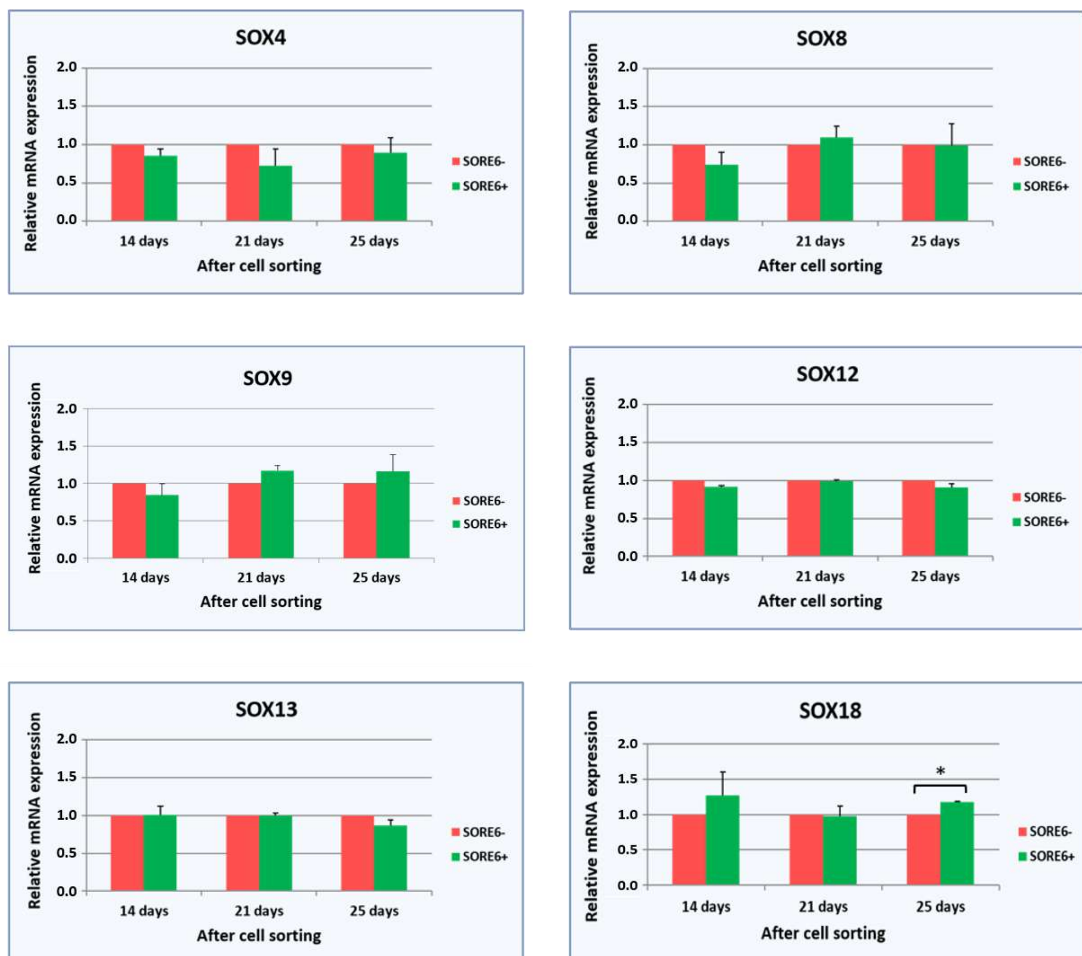


Figure 15 | SOX4, SOX8, SOX9, SOX12, SOX13 and SOX18 relative mRNA expression of SORE6+ and SORE6- cells by real-time PCR analysis at different timepoints after cells sorting. Significant differences (*P ≤ 0.05).

2.2| Colorectal CSCs-related

The expression of some TFs described to be associated with CSCs such as the KLF4, STAT3, c-MYC, GLI1, SALL4, β-catenin, TCF4, HMGA1 and HMGA2 was evaluated by real time PCR in SORE6+ and SORE6- cells collected at different timepoints (14, 21, 25

days) after the sorting. The housekeeping control genes, 18S and HPRT1 were used for target gene abundance normalization and data was analyzed by the comparative method ΔΔCT.

SALL4 was tested but it was not expressed in SORE6+ and SORE6- cells (CTs ≥ 35.0). Between the SORE6+ and SORE6- cells no significant differences were found in the expression of KLF4, c-MYC, GLI1, β-catenin, TCF4, and HMGA2. However, STAT3 and HMGA1 mRNA expression was significantly higher in SORE6+ cells when compared to SORE6- cells, only for cells collected 25 days (1.177 ± 0.004; *P*=0.010) and 14 days after sorting (1.307 ± 0.036; *P*=0.053), respectively (figure 16).

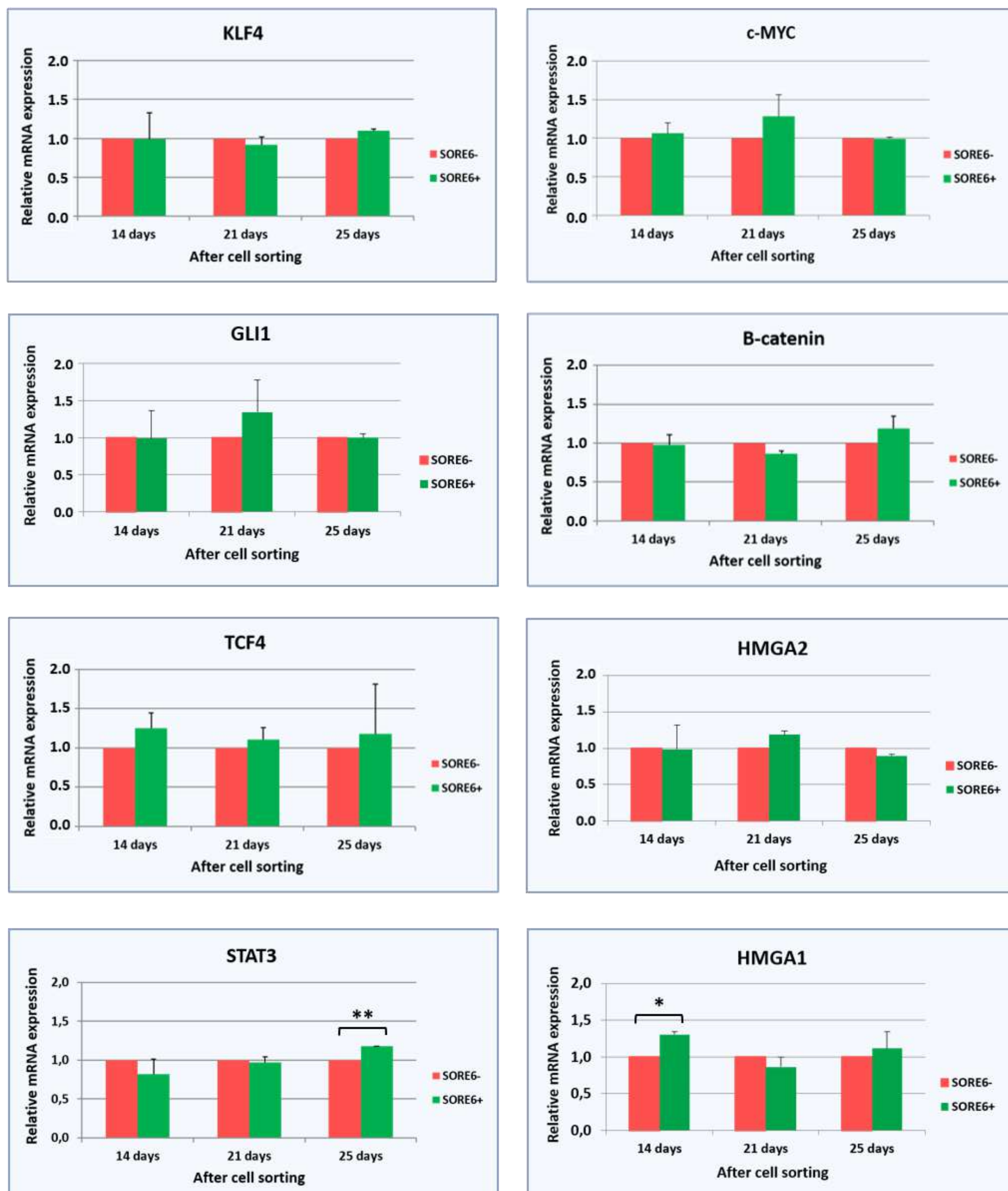


***Figure 16* | KLF4, c-MYC, GLI1, β-catenin, TCF4, HMGA2, STAT3 and HMGA1 relative mRNA expression** of SORE6+ and SORE6- cells by real-time PCR analysis at different timepoints after cells sorting. Significant differences (**P* ≤ 0.05; ** *P* ≤ 0.01).

2.3| Colorectal cancer - related

Numerous transcription factors have been described to be expressed in CRC namely FOXC2, TEAD2, OCT1, CNOT3, HOXA10 and MEIS2. The expression of these TFs was also evaluated by real-time PCR in SORE6+ and SORE6- cells collected at different timepoints (14, 21, 25 days) after the sorting. The housekeeping control genes, 18S and HPRT1 were used for target gene abundance normalization and data was analyzed by the comparative method ΔΔCT.

Between the SORE6+ and SORE6- cells no significant differences were found in the expression of FOXC2, TEAD2, CNOT3, HOXA10 and MEIS2. However, OCT1 relative mRNA expression was significantly higher in SORE6+ cells when compared to SORE6- cells only for cells collected 25 days after sorting (1.182 ± 0.016; P=0.041) (figure17).

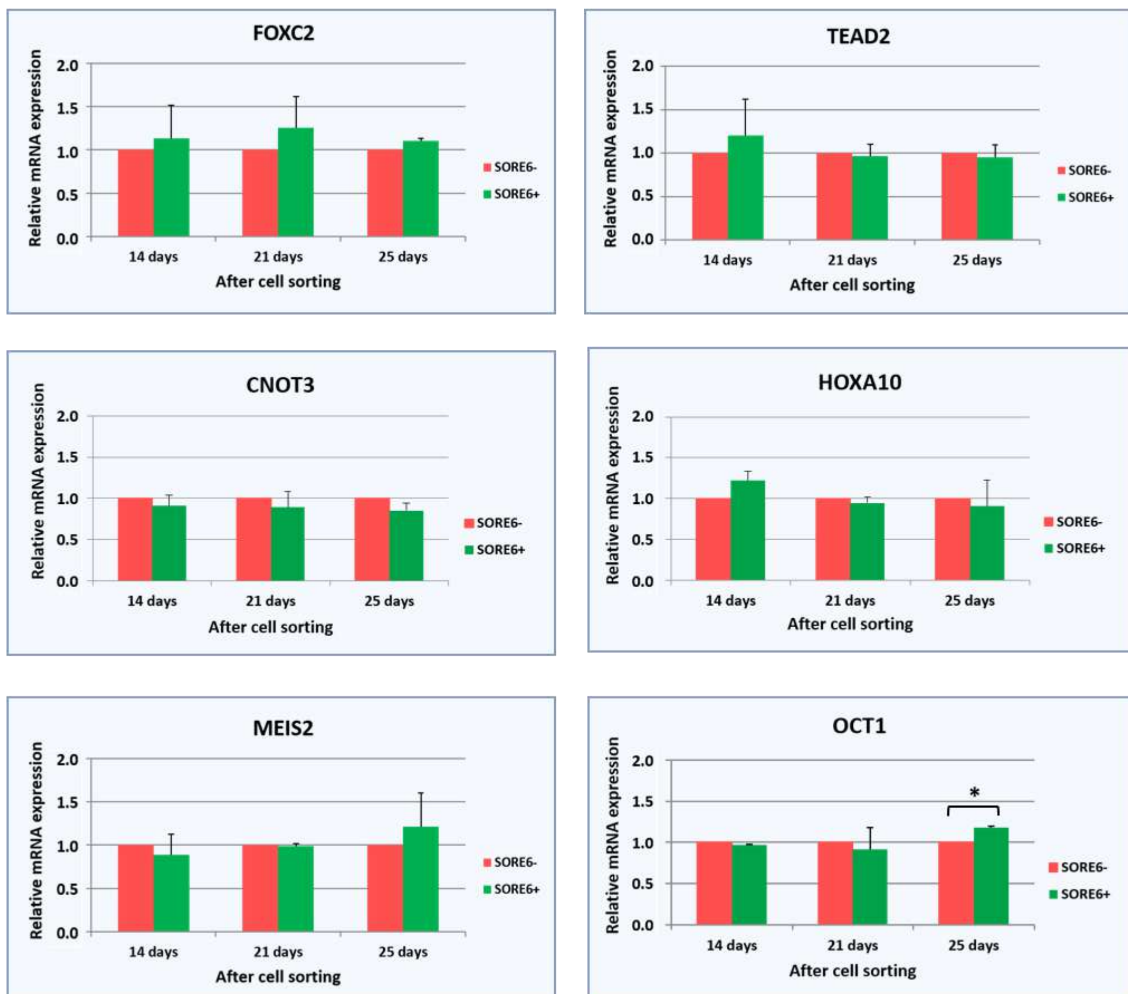


Figure 17 | FOXC2, TEAD2, CNOT3, HOXA10, MEIS2 and OCT1 relative mRNA expression of SORE6+ and SORE6- cells by real-time PCR analysis in different timepoints after cells sorting. Significant differences (*$P \leq 0.05$).

3| SOX2 co-immunoprecipitation and mass-spectrometry

Co-immunoprecipitation was done in protein samples from SORE6+ and SORE6- cells previously transfected with a SOX2 expression vector (figure 18) and by using a FLAG antibody. WB was performed to ensure the up-regulation of SOX2 in the two cell subpopulations. Its expression was found in both SORE6+ and SORE6- subpopulations transfected with the SOX2 vector and not in those transfected with the corresponding empty vector (pcDNA3.1). Also, SOX2 expression was found in the positive control (AGS gastric cancer cell line).

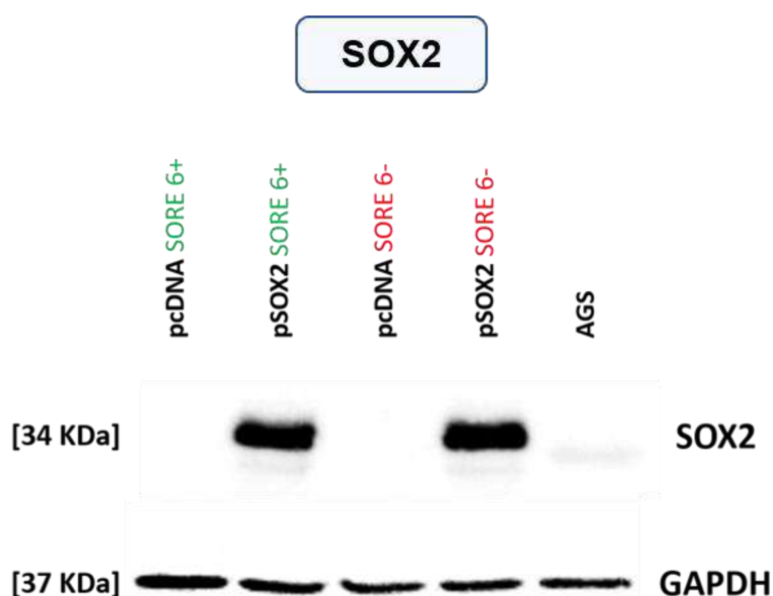


Figure 18 | WB analysis of transfection with an expression vector for SOX2 (pSOX2) and with an empty vector (pcDNA) on sorted SORE6+ and SORE6- cell subpopulations. The internal control used was GAPDH.

The subsequent mass spectrometry analysis was performed by the i3S Proteomics platform. From all the possible SOX2 interacting proteins obtained, we discarded the ones appearing with unique peptides ≤ 2. Proteins were considered as significantly more present in SORE6+ cells when their abundance in SORE6+ cells in relation to SORE6- cells was ≥ 2 such as KRT1, KRT2, KRT9, KRT10, KRT14, HRNR, RPL10A, LDHA and DCD - (table VI) - and in the other hand when it was ≤ 0.5 such as IMPDH1, IMPDH2, PARP1 and S100A9 - (table VII) - both comparisons with an adjusted P-value < 0.05. The only exception considered was HMGA1 protein for which the observed P-value was 0.106.

Table VI | SOX2-associated proteins with an abundance ≥ 2 in SORE6+ in relation to SORE6- cells.

Gene	Protein	SORE6+ to SORE6- abundance	SORE6+ to SORE6- abundance adjusted *P*	Unique Peptides	Coverage (%)	Sum PEP Score
KRT1	Keratin, type II cytoskeletal 1	9.541	< 0.001	33	65	407.046
KRT9	Keratin, type I cytoskeletal 9	14.220	< 0.001	27	58	299.522
KRT10	Keratin, type I cytoskeletal 10	10.641	< 0.001	28	58	293.729
KRT2	Keratin, type II cytoskeletal 2 epidermal	8.875	< 0.001	25	66	284.520
KRT14	Keratin, type I cytoskeletal 14	7.355	< 0.001	5	42	109.932
HRNR	Hornerin	2.824	0.002	4	6	43.792
RPL10A	60S ribosomal protein L10a	2.332	0.021	4	26	28.925
LDHA	L-lactate dehydrogenase A chain	2.593	0.005	3	16	27.450
DCD	Dermcidin	2.175	0.045	3	20	16.698
HMGA1	High mobility group A1	1.987	0.106	5	37	45.781

Table VII | SOX2-associated proteins with an abundance ≤ 0.5 in SORE6+ in relation to SORE6- cells.

Gene	Protein	SORE6+ to SORE6- abundance	SORE6 +/- abundance adjusted *P*	Unique Peptides	Coverage (%)	Sum PEP Score
IMPDH2	Inosine-5'-monophosphate dehydrogenase 2	0.465	0.012	19	53	170.499
IMPDH1	Inosine-5'-monophosphate dehydrogenase 1	0.524	0.046	8	32	37.914
PARP1	Poly (ADP-ribose) polymerase 1	0.519	0.042	9	11	34.760
S100A9	Protein S100-A9	0.206	0.001	3	31	21.375

3.1| HMGA1 as a CSC marker

Our previous results have shown that HMGA1 mRNA expression was significantly higher in SORE6+ cells when compared to SORE6- cells, but only for cells collected 14 days after sorting. By co-immunoprecipitation, HMGA1 presented a relative abundance of 1.987 in SORE6+ in relation to SORE6- cells showing that this protein was approximately two times more expressed in SORE6+ cells than in SORE6- cells, although this difference was not statistically significant ($P = 0.106$).

We performed a WB to see if HMGA1 was already more expressed in SORE6+, when compared to SORE6- cells and in fact it was shown to be only expressed in the SORE+ cells subpopulation (figure 19).

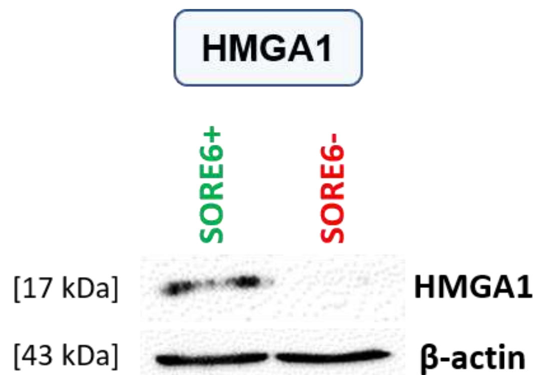


Figure 19 | WB analysis of HMGA1 on sorted SW480 SORE6 cell subpopulations. In this WB the internal control used was β-actin.

Chapter 5 | Discussion and conclusion

CRC is the third most frequent and the second major cause of death-related to cancer in the world (1, 2). CRC cancer incidence is directly related not only to environmental but also to hereditary factors with up nearly 65% of CRC cases occurring sporadically. Predictive biomarkers are remarkably needed in the decision-making process for therapy approaches and optimization of CRC patients management and follow up, with mass spectrometry representing an effective tool in biomarkers identification (43, 48, 59). Despite all the therapeutic approaches that are presently available, their efficacy remains limited as a consequence of drug resistance.

The CSC model hypothesizes the existence of a subpopulation of tumor cells with stem-like characteristics and self-renewal ability. These cells are capable of initiating and sustaining tumorigenesis, metastasis, cancer relapse and resistance to both chemotherapy and radiotherapy (62-66). Moreover, TFs together with the development of reporter systems that can actively monitor intracellular markers, represent an effective tool in CSCs identification and study in a more efficient way than the use of cell-surface markers (66, 85, 97-100).

In our study we used a SORE6-GFP lentiviral-based reporter system to transduce SW480 colon cancer cell line. This reporter system have allowed breast and gastric CSCs direct visualization, depending on the expression of SOX2 and OCT4 (66). By using FACS we were able to isolate and then characterize two subpopulations - SORE6+, which have the SORE6 activated and GFP expression, and SORE6-, without GFP expression. A low expression of OCT4 was detected in both subpopulations and, interestingly, no significant differences were found between them in what concerns the expression of SOX2. Moreover, when cell proliferation ability was evaluated no significant differences were found between SORE6+ and SORE6- cells.

Some CSCs-related-features were then evaluated, such as the ability to proliferate and generate large sphere-like structures - colonospheres - and the resistance to chemotherapeutic drugs as 5-FU. SORE6+ cells revealed more capable of forming a significantly higher number of colonospheres and showed to be more resistant to 5-FU treatment. With these results, similar to those obtained by Tang et al. and by our group in gastric cancer cell lines, it is possible to hypothesize that SORE6+ cells behave as cancer stem cells, as compared to SORE6- cells (66, 116). When SOX2 knockdown was performed before 5-FU treatment, in both SORE6+ and SORE6-, the viability in the two subpopulations has decreased, supporting a crucial role for SOX2 in therapy resistance.

We had also evaluated the expression of CSCs surface markers. Both CD44v6 and CD49f markers were expressed in the two subpopulations (as previously described for

SW480), with no differences found between SORE6+ and SORE6- cells (222, 223). No ALDH1 expression was detected in our cell subpopulations and these results are consistent with those found in the literature, showing low expression of ALDH1 in SW480 (224). In what concerns c-MYC, no differences were seen when comparing SORE6+ and SORE6- cells, although being expressed, as already described for SW480 cell line (225).

Despite the interesting observations, SOX2 (or OCT4) differential expression could not be used to explain the phenotypical differences between SORE6+ and SORE6- cells, therefore our subsequent approach consisted in verifying if other TFs that can also bind to SOX2 binding site or directly to SOX2 protein could be expressed differently in our subpopulations. In what concerns to SOX family of TFs we found a significantly higher expression of SOX18 in SORE6+ cells when compared to SORE6- cells but only for cells collected 25 days after sorting. Indeed, SOX18 upregulation has already been found in CRC tissues and cells lines and it was related with prognosis, but as far as we know there is no report of its relation with CSCs (169). Also, HMGA1, STAT3 and OCT1 expression was significantly higher in SORE6+ cells when compared to SORE6- cells (only at one timepoint), supporting the results found in the literature which described HMGA1 and STAT3 as being highly expressed in CSCs from SW480 and other CRC cell lines (175, 226) and OCT1 as being associated with bad prognosis in CRC (213).

By real-time PCR we obtained significant differences in the expression of four TFs – SOX18, and HMGA1, STAT3 and OCT1– between the two subpopulations and these have to be further studied. Meanwhile, we decided to develop another strategy - to immunoprecipitate and analyse by mass spectrometry SOX2-interacting proteins in both SORE6+ and SORE6- cells. As a first step, we had to express SOX2 in both cell lines, since the endogenous levels of expression were not sufficient to allow this analysis and we have used a flag antibody to co-immunoprecipitate proteins bound to SOX2.

Some proteins - KRT1, KRT2, KRT9, KRT10, KRT14, HRNR, RPL10A, LDHA and DCD - appeared significantly increased in SORE6+, when compared to SORE6- cells:

The role of cytoskeletal keratins - as KRT1, KRT2, KRT9, KRT10 and KRT14 - in colon carcinogenesis and in colon CSCs is not yet established, however keratin activity has been associated with cancer cell growth, metastasis and invasion. Moreover, some studies have shown that the expression of keratins is important in the maintenance of CSCs characteristics (227-230). Furthermore, KRT14, used as a CSC marker, was found in the literature to be expressed in SOX2-positive CSCs in human bladder cancer (231).

HRNR (hornerin) is a member of the S100 protein family, that is involved in pathological processes and whose expression has been found in some tumors, however its

role in CRC and in CSCs is not yet described. In hepatocellular carcinoma HRNR was associated with tumor progression and with a poor prognosis and in breast cancer it was related to tumor development and malignant transformation (232, 233).

RPL10A encodes a ribosomal protein from 60S subunit and has been described to be involved in cell proliferation, tumorigenesis and, therefore, in the development of cancer (234). RPL10A has been described to be expressed in colorectal adenomas and CRC and, in glioblastoma it was shown to interact with SOX2, together with other ribonucleoproteins, which suggests an important function for SOX2 in post-transcriptional regulation (235, 236).

LDHA (Lactate dehydrogenase A) is an essential regulator of glycolysis and it has been described to be highly expressed in SW480 colon cancer cell line and involved in CRC cell proliferation. Moreover, by generating extracellular lactate, LDHA provides propitious conditions for colorectal CSCs growth and invasion. Zhang et al. had found that LDHA expression was significantly associated with OCT4 and that LDHA knockdown was capable of decreasing OCT4 expression and, consequently, its tumorigenicity (237, 238).

DCD (Dermcidin) has been described to have a role in carcinogenesis in some types of tumors as gastric cancer and to be significantly associated with lymph node metastasis, tumor differentiation and poor prognosis.

Besides, the expression of other proteins - IMPDH1, IMPDH2, PARP1 and S100A9 - was significantly higher in SORE6- cells:

Inosine-5′-monophosphate dehydrogenase (IMPDH) is an enzyme with two distinct isoforms, IMPDH1 and IMPDH2, that share 84% similarity in their sequence of amino acids (239). It catalyzes the oxidation of IMP to xanthosine monophosphate and determines the guanine nucleotide pool and consequently DNA replication and cell proliferation. It binds to chromatin and represses transcription expression of histone genes and a transcription factor important for DNA replication (240). Whereas IMPDH1 is expressed in normal lymphocytes and leukocytes and in small cell lung cancer, IMPDH2 has been related to chemoresistance, differentiation and has been described to be upregulated in proliferating cells and in tumor tissues including CRC (239, 241-243). Other studies had reported IMPDH2 upregulation in some CRC cell lines such as SW480, being capable of promoting invasion, proliferation and tumorigenicity of CRC cells. It is associated with EMT transition, more aggressive features and poor prognosis (242). Nevertheless, IMPDH2 have already been described as a SOX2-interactor in brain tumors, and this information is available in BIOGRID biomedical interaction repository (244).

PARP1 (poly(ADP-ribose) polymerase 1), a nucleosome binding protein that is responsible for a DNA damage response, recruits some repair factors and leads to the activation of effector proteins. Importantly, it has been described as a SOX2 interactor in human glioblastoma and in mouse ESCs, stabilizes and facilitates SOX2 binding to nucleosomal sites in the chromatin suggesting that PARP1 functions as a cofactor in SOX-mediated activity (107, 219, 236). This information is also available in BIOGRID biomedical interaction repository.

S100A9, an element of S100 calcium binding proteins, has been described to be overexpressed in some types of cancers and as being involved in chemotherapy resistance (245, 246). Also, it has been described in colon CSCs to regulate mTORC1, a kinase that is activated in some cancers (247). Other study in glioma cells concluded that a higher expression of S100A9 was associated with tumorigenesis and CSCs (248).

HMGA1 (high mobility group AT-hook 1) protein, appeared around two times more co-immunoprecipitated with SOX2 in SORE6+ than in SORE6- cells, but the difference was not statistically significant (P=0.106). This TF appeared already associated with SOX2 expression when we did the data mining analysis and its mRNA expression was significantly higher in SORE6+ cells when compared to SORE6- cells. To investigate if the HMGA1 protein had higher levels in SORE6+ cells, we performed an WB and, in fact, it has shown to be expressed only in the SORE6+ cells subpopulation. Puca *et al.* have shown a key role for HMGA1 in colon CSCs self-renewal sphere-forming capacity (226). D'Angelo et al. reported that HMGA1 protein led to 5-FU chemoresistance in colon and thyroid cancer (249). The high expression of HMGA1 was associated to CRC progression, mainly through the Wnt signaling pathway (199, 202, 250). HMGA1 enhanced the reprogramming of somatic cells in order to induce stem cells, when added to reprogramming TFs such as SOX2, OCT4, c-MYC and KLF4 (196, 226, 251). Also, HMGA1 expression was related with the expression of OCT4, SOX2, KLF4 and NANOG, and induced SOX2 expression (196, 201). These studies are in line with our results and suggest the HMGA1 crucial role in the maintenance of a CSC phenotype, proposing HMGA1 as a powerful tumor marker and a propitious target for CSCs eradication therapies and for the management and treatment of numerous types of cancers, including CRC.

Overall, we conclude that SORE6-GFP reporter is capable of detecting cells with a CSC phenotype in SW480 colon cancer cell line, since SORE6+ cells have more resistance to 5-FU chemotherapeutic drug and are capable of forming a larger number of colonospheres. Co-immunoprecipitation and mass spectrometry analysis, revealed different proteins co-precipitating with SOX2 in SORE6+ and SORE6- and these

observations can justify the different phenotypes observed, despite similar expression of SOX2. However, further studies are necessary to elucidate the role of HMGA1 in colorectal CSCs reprogramming and SOX2 interaction and regulation, in this model. The same is true for some other interesting proteins that were shown to differentially co-precipitate with SOX2 and are undoubtedly CSCs-related.

Chapter 6 | Future perspectives

This study has revealed some promising factors, but important validations and additional work should be performed to support and clarify the data obtained. Our future perspectives consist in validating all the data obtained in both real-time PCR and co-immunoprecipitation, by Western Blot, as we did for HMGA1. Moreover, and because we are truly convinced of the role of HMGA1 in CSCs, we intend to overexpress/down-regulate HMGA1 in SORE6-/SORE6+, respectively and evaluate differences in CSC phenotype, namely drug resistance and formation of colonospheres. We would very much like to investigate the role of HMGA1 as a SOX2-cofactor and in SOX2 regulation, namely by using luciferase reporter assays. The same analysis should be performed in other CRC cell lines that present a higher expression of SOX2, in order to better understand its role in CSC reprogramming.

Chapter 7 | References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017;66(4):683-91.

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.

3. Cancer IAfRo. Globocan 2018: Cancer Fact Sheets - Colorectal Cancer. 2018.

4. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019;16(12):713-32.

5. Wilsnack RW, Wilsnack SC, Kristjanson AF, Vogeltanz-Holm ND, Gmel G. Gender and alcohol consumption: patterns from the multinational GENACIS project. Addiction. 2009;104(9):1487-500.

6. Karastergiou K, Smith SR, Greenberg AS, Fried SK. Sex differences in human adipose tissues - the biology of pear shape. Biol Sex Differ. 2012;3(1):13.

7. Huxley RR, Ansary-Moghaddam A, Clifton P, Czernichow S, Parr CL, Woodward M. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. Int J Cancer. 2009;125(1):171-80.

8. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. The Lancet. 2019;394(10207):1467-80.

9. Buccafusca G, Proserpio I, Tralongo AC, Rametta Giuliano S, Tralongo P. Early colorectal cancer: diagnosis, treatment and survivorship care. Crit Rev Oncol Hematol. 2019;136:20-30.

10. Patel SG, Ahnen DJ. Colorectal Cancer in the Young. Curr Gastroenterol Rep. 2018;20(4):15.

11. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin. 2018;68(4):250-81.

12. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal Cancer Screening: Recommendations for Physicians and Patients From the U.S. Multi-Society Task Force on Colorectal Cancer. Gastroenterology. 2017;153(1):307-23.

13. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. Gastroenterology. 2010;138(6):2044-58.

14. Pearlman R, Frankel WL, Swanson B, Zhao W, Yilmaz A, Miller K, et al. Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. JAMA Oncol. 2017;3(4):464-71.

15. Hemminki K, Chen B. Familial risk for colorectal cancers are mainly due to heritable causes. Cancer Epidemiol Biomarkers Prev. 2004;13(7):1253-6.

16. Quintero E, Carrillo M, Leoz ML, Cubiella J, Gargallo C, Lanas A, et al. Risk of Advanced Neoplasia in First-Degree Relatives with Colorectal Cancer: A Large Multicenter Cross-Sectional Study. PLoS Med. 2016;13(5):e1002008.

17. Syngal S, Brand RE, Church JM, Giardiello FM, Hampel HL, Burt RW, et al. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. Am J Gastroenterol. 2015;110(2):223-62; quiz 63.

18. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759-67.

19. Dow LE, O'Rourke KP, Simon J, Tschaharganeh DF, van Es JH, Clevers H, et al. Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. Cell. 2015;161(7):1539-52.

20. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. Gastrointest Cancer Res. 2012;5(1):19-27.

21. Itzkowitz SH, Yio X. Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. Am J Physiol Gastrointest Liver Physiol. 2004;287(1):G7-17.

22. Kedrin D, Gala MK. Genetics of the serrated pathway to colorectal cancer. Clin Transl Gastroenterol. 2015;6(4):e84.

23. Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. Gastroenterology. 2010;138(6):2088-100.

24. Mundade R, Imperiale TF, Prabhu L, Loehrer PJ, Lu T. Genetic pathways, prevention, and treatment of sporadic colorectal cancer. Oncoscience. 2014;1(6):400-6.

25. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology. 2008;135(4):1079-99.

26. Ogino S, Goel A. Molecular classification and correlates in colorectal cancer. J Mol Diagn. 2008;10(1):13-27.

27. Li D. Recent advances in colorectal cancer screening. Chronic Dis Transl Med. 2018;4(3):139-47.

28. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med. 1988;319(9):525-32.

29. Pawlik TM, Raut CP, Rodriguez-Bigas MA. Colorectal carcinogenesis: MSI-H versus MSI-L. Dis Markers. 2004;20(4-5):199-206.

30. Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. N Engl J Med. 2009;361(25):2449-60.

31. Nazemalhosseini Mojarad E, Kuppen PJ, Aghdaei HA, Zali MR. The CpG island methylator phenotype (CIMP) in colorectal cancer. Gastroenterol Hepatol Bed Bench. 2013;6(3):120-8.

32. Funkhouser WK, Jr., Lubin IM, Monzon FA, Zehnbauer BA, Evans JP, Ogino S, et al. Relevance, pathogenesis, and testing algorithm for mismatch repair-defective colorectal carcinomas: a report of the association for molecular pathology. J Mol Diagn. 2012;14(2):91-103.

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

34. Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. Gastroenterology. 2015;149(5):1177-90.e3.

35. Arvelo F, Sojo F, Cotte C. Biology of colorectal cancer. Ecancermedicalscience. 2015;9:520.

36. Rutka M, Bor R, Bálint A, Fábián A, Milassin Á, Nagy F, et al. Diagnostic Accuracy of Five Different Fecal Markers for the Detection of Precancerous and Cancerous Lesions of the Colorectum. Mediators of inflammation. 2016;2016:2492081-.

37. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, et al. Cancer treatment and survivorship statistics, 2014. CA Cancer J Clin. 2014;64(4):252-71.

38. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. Ann Intern Med. 2011;154(1):22-30.

39. Al Bandar MH, Kim NK. Current status and future perspectives on treatment of liver metastasis in colorectal cancer (Review). Oncology reports. 2017;37(5):2553-64.

40. Feo L, Polcino M, Nash GM. Resection of the Primary Tumor in Stage IV Colorectal Cancer: When Is It Necessary? Surg Clin North Am. 2017;97(3):657-69.

41. Greene FL. Current TNM staging of colorectal cancer. Lancet Oncol. 2007;8(7):572-3.

42. Ang CS, Baker MS, Nice EC. Mass Spectrometry-Based Analysis for the Discovery and Validation of Potential Colorectal Cancer Stool Biomarkers. Methods Enzymol. 2017;586:247-74.

43. Coghlin C, Murray GI. Biomarkers of colorectal cancer: recent advances and future challenges. Proteomics Clin Appl. 2015;9(1-2):64-71.

44. Alvarez-Chaver P, Otero-Estévez O, Páez de la Cadena M, Rodríguez-Berrocal FJ, Martínez-Zorzano VS. Proteomics for discovery of candidate colorectal cancer biomarkers. World journal of gastroenterology. 2014;20(14):3804-24.

45. de Wit M, Fijneman RJ, Verheul HM, Meijer GA, Jimenez CR. Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. Clin Biochem. 2013;46(6):466-79.

46. Campos-da-Paz M, Dórea JG, Galdino AS, Lacava ZGM, de Fatima Menezes Almeida Santos M. Carcinoembryonic Antigen (CEA) and Hepatic Metastasis in Colorectal Cancer: Update on Biomarker for Clinical and Biotechnological Approaches. Recent Pat Biotechnol. 2018;12(4):269-79.

47. Gonzalez-Pons M, Cruz-Correa M. Colorectal Cancer Biomarkers: Where Are We Now? Biomed Res Int. 2015;2015:149014.

48. Alves Martins BA, de Bulhões GF, Cavalcanti IN, Martins MM, de Oliveira PG, Martins AMA. Biomarkers in Colorectal Cancer: The Role of Translational Proteomics Research. Front Oncol. 2019;9.

49. Alnabulsi A, Murray GI. Proteomics for early detection of colorectal cancer: recent updates. Expert Rev Proteomics. 2018;15(1):55-63.

50. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. Nat Rev Dis Primers. 2015;1:15065.

51. Sánchez-Gundín J, Fernández-Carballido AM, Martínez-Valdivieso L, Barreda-Hernández D, Torres-Suárez AI. New Trends in the Therapeutic Approach to Metastatic Colorectal Cancer. Int J Med Sci. 2018;15(7):659-65.

52. Brown KGM, Solomon MJ, Mahon K, O'Shannassy S. Management of colorectal cancer. Bmj. 2019;366:l4561.

53. Messersmith WA. NCCN Guidelines Updates: Management of Metastatic Colorectal Cancer. J Natl Compr Canc Netw. 2019;17(5.5):599-601.

54. Van Cutsem E, Cervantes A, Nordlinger B, Arnold D. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2014;25 Suppl 3:iii1-9.

55. Vera R, Alonso V, Gállego J, González E, Guillén-Ponce C, Pericay C, et al. Current controversies in the management of metastatic colorectal cancer. Cancer Chemother Pharmacol. 2015;76(4):659-77.

56. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. Signal Transduct Target Ther. 2020;5(1):22.

57. André T, de Gramont A, Vernerey D, Chibaudel B, Bonnetain F, Tijeras-Raballand A, et al. 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. J Clin Oncol. 2015;33(35):4176-87.

58. Duineveld LA, van Asselt KM, Bemelman WA, Smits AB, Tanis PJ, van Weert HC, et al. Symptomatic and Asymptomatic Colon Cancer Recurrence: A Multicenter Cohort Study. Ann Fam Med. 2016;14(3):215-20.

59. You W, Yan L, Cai Z, Xie L, Sheng N, Wang G, et al. Clinical Significances of Positive Postoperative Serum CEA and Post-preoperative CEA Increment in Stage II and III Colorectal Cancer: A Multicenter Retrospective Study. Front Oncol. 2020;10:671-.

60. Valent P, Bonnet D, De Maria R, Lapidot T, Copland M, Melo JV, et al. Cancer stem cell definitions and terminology: the devil is in the details. Nat Rev Cancer. 2012;12(11):767-75.

61. Marjanovic ND, Weinberg RA, Chaffer CL. Cell plasticity and heterogeneity in cancer. Clin Chem. 2013;59(1):168-79.

62. Clevers H. The cancer stem cell: premises, promises and challenges. Nat Med. 2011;17(3):313-9.

63. Iseghohi SO. Cancer stem cells may contribute to the difficulty in treating cancer. Genes Dis. 2016;3(1):7-10.

64. Magee JA, Piskounova E, Morrison SJ. Cancer stem cells: impact, heterogeneity, and uncertainty. Cancer Cell. 2012;21(3):283-96.

65. Friedmann-Morvinski D, Verma IM. Dedifferentiation and reprogramming: origins of cancer stem cells. EMBO Rep. 2014;15(3):244-53.

66. Tang B, Raviv A, Esposito D, Flanders Kathleen C, Daniel C, Nghiem Bao T, et al. A Flexible Reporter System for Direct Observation and Isolation of Cancer Stem Cells. Stem cell reports. 2015;4(1):155-69.

67. Brunner TB, Kunz-Schughart LA, Grosse-Gehling P, Baumann M. Cancer stem cells as a predictive factor in radiotherapy. Semin Radiat Oncol. 2012;22(2):151-74.

68. Dalerba P, Cho RW, Clarke MF. Cancer stem cells: models and concepts. Annu Rev Med. 2007;58:267-84.

69. Baccelli I, Trumpp A. The evolving concept of cancer and metastasis stem cells. The Journal of cell biology. 2012;198(3):281-93.

70. Basu S, Haase G, Ben-Ze'ev A. Wnt signaling in cancer stem cells and colon cancer metastasis. F1000Res. 2016;5.

71. Phi LTH, Sari IN, Yang YG, Lee SH, Jun N, Kim KS, et al. Cancer Stem Cells (CSCs) in Drug Resistance and their Therapeutic Implications in Cancer Treatment. Stem Cells Int. 2018;2018:5416923.

72. Peitzsch C, Kurth I, Kunz-Schughart L, Baumann M, Dubrovska A. Discovery of the cancer stem cell related determinants of radioresistance. Radiother Oncol. 2013;108(3):378-87.

73. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100(7):3983-8.

74. Ignatova TN, Kukekov VG, Laywell ED, Suslov ON, Vrionis FD, Steindler DA. Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. Glia. 2002;39(3):193-206.

75. O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature. 2007;445(7123):106-10.

76. Zhang S, Balch C, Chan MW, Lai HC, Matei D, Schilder JM, et al. Identification and characterization of ovarian cancer-initiating cells from primary human tumors. Cancer Res. 2008;68(11):4311-20.

77. Eramo A, Lotti F, Sette G, Pilozzi E, Biffoni M, Di Virgilio A, et al. Identification and expansion of the tumorigenic lung cancer stem cell population. Cell Death Differ. 2008;15(3):504-14.

78. Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J, et al. Identification of a cancer stem cell in human brain tumors. Cancer Res. 2003;63(18):5821-8.

79. Schatton T, Murphy GF, Frank NY, Yamaura K, Waaga-Gasser AM, Gasser M, et al. Identification of cells initiating human melanomas. Nature. 2008;451(7176):345-9.

80. Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, et al. Identification of pancreatic cancer stem cells. Cancer Res. 2007;67(3):1030-7.

81. Yang ZF, Ho DW, Ng MN, Lau CK, Yu WC, Ngai P, et al. Significance of CD90+ cancer stem cells in human liver cancer. Cancer Cell. 2008;13(2):153-66.

82. Chen K, Huang YH, Chen JL. Understanding and targeting cancer stem cells: therapeutic implications and challenges. Acta Pharmacol Sin. 2013;34(6):732-40.

83. Dalerba P, Dylla SJ, Park IK, Liu R, Wang X, Cho RW, et al. Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A. 2007;104(24):10158-63.

84. Ginestier C, Hur MH, Charafe-Jauffret E, Monville F, Dutcher J, Brown M, et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell stem cell. 2007;1(5):555-67.

85. Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nat Rev Cancer. 2008;8(10):755-68.

86. Deng Y, Zhou J, Fang L, Cai Y, Ke J, Xie X, et al. ALDH1 is an independent prognostic factor for patients with stages II-III rectal cancer after receiving radiochemotherapy. British journal of cancer. 2014;110(2):430-4.

87. Rassouli FB, Matin MM, Saeinasab M. Cancer stem cells in human digestive tract malignancies. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(1):7-21.

88. Wang Z, Tang Y, Xie L, Huang A, Xue C, Gu Z, et al. The Prognostic and Clinical Value of CD44 in Colorectal Cancer: A Meta-Analysis. Front Oncol. 2019;9:309.

89. Huang R, Wang G, Song Y, Tang Q, You Q, Liu Z, et al. Colorectal cancer stem cell and chemoresistant colorectal cancer cell phenotypes and increased sensitivity to Notch pathway inhibitor. Molecular medicine reports. 2015;12(2):2417-24.

90. Haraguchi N, Ishii H, Mimori K, Ohta K, Uemura M, Nishimura J, et al. CD49f-positive cell population efficiently enriches colon cancer-initiating cells. Int J Oncol. 2013;43(2):425-30.

91. Tachezy M, Zander H, Gebauer F, Marx A, Kaifi JT, Izbicki JR, et al. Activated leukocyte cell adhesion molecule (CD166)--its prognostic power for colorectal cancer patients. J Surg Res. 2012;177(1):e15-20.

92. Zhang SS, Huang ZW, Li LX, Fu JJ, Xiao B. Identification of CD200+ colorectal cancer stem cells and their gene expression profile. Oncology reports. 2016;36(4):2252-60.

93. Hsu HC, Liu YS, Tseng KC, Hsu CL, Liang Y, Yang TS, et al. Overexpression of Lgr5 correlates with resistance to 5-FU-based chemotherapy in colorectal cancer. International journal of colorectal disease. 2013;28(11):1535-46.

94. Wu W, Cao J, Ji Z, Wang J, Jiang T, Ding H. Co-expression of Lgr5 and CXCR4 characterizes cancer stem-like cells of colorectal cancer. Oncotarget. 2016;7(49):81144-55.

95. Kemper K, Prasetyanti PR, De Lau W, Rodermond H, Clevers H, Medema JP. Monoclonal antibodies against Lgr5 identify human colorectal cancer stem cells. Stem Cells. 2012;30(11):2378-86.

96. Leng Z, Xia Q, Chen J, Li Y, Xu J, Zhao E, et al. Lgr5+CD44+EpCAM+ Strictly Defines Cancer Stem Cells in Human Colorectal Cancer. Cell Physiol Biochem. 2018;46(2):860-72.

97. Hadjimichael C, Chanoumidou K, Papadopoulou N, Arampatzi P, Papamatheakis J, Kretsovali A. Common stemness regulators of embryonic and cancer stem cells. World journal of stem cells. 2015;7(9):1150-84.

98. Zhou Y, Xia L, Wang H, Oyang L, Su M, Liu Q, et al. Cancer stem cells in progression of colorectal cancer. Oncotarget. 2017;9(70):33403-15.

99. Pádua D, Figueira P, Ribeiro I, Almeida R, Mesquita P. The Relevance of Transcription Factors in Gastric and Colorectal Cancer Stem Cells Identification and Eradication. Front Cell Dev Biol. 2020;8:442-.

100. Saygin C, Samour M, Chumakova A, Jarrar A, Lathia JD, Reizes O. Reporter Systems to Study Cancer Stem Cells. Methods Mol Biol. 2016;1516:319-33.

101. Young RA. Control of the embryonic stem cell state. Cell. 2011;144(6):940-54.

102. Schmidt R, Plath K. The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. Genome Biol. 2012;13(10):251.

103. Pevny LH, Lovell-Badge R. Sox genes find their feet. Curr Opin Genet Dev. 1997;7(3):338-44.

104. Wegner M. From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res. 1999;27(6):1409-20.

105. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 2003;17(1):126-40.

106. Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol. 2007;9(6):625-35.

107. Julian LM, McDonald AC, Stanford WL. Direct reprogramming with SOX factors: masters of cell fate. Curr Opin Genet Dev. 2017;46:24-36.

108. Kamachi Y, Uchikawa M, Kondoh H. Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet. 2000;16(4):182-7.

109. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-76.

110. Gangemi RM, Griffero F, Marubbi D, Perera M, Capra MC, Malatesta P, et al. SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity. Stem Cells. 2009;27(1):40-8.

111. Arnold K, Sarkar A, Yram MA, Polo JM, Bronson R, Sengupta S, et al. Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. Cell stem cell. 2011;9(4):317-29.

112. Li XL, Eishi Y, Bai YQ, Sakai H, Akiyama Y, Tani M, et al. Expression of the SRY-related HMG box protein SOX2 in human gastric carcinoma. Int J Oncol. 2004;24(2):257-63.

113. Han X, Fang X, Lou X, Hua D, Ding W, Foltz G, et al. Silencing SOX2 induced mesenchymal-epithelial transition and its expression predicts liver and lymph node metastasis of CRC patients. PloS one. 2012;7(8):e41335.

114. Amini S, Fathi F, Mobalegi J, Sofimajidpour H, Ghadimi T. The expressions of stem cell markers: Oct4, Nanog, Sox2, nucleostemin, Bmi, Zfx, Tcl1, Tbx3, Dppa4, and Esrrb in bladder, colon, and prostate cancer, and certain cancer cell lines. Anat Cell Biol. 2014;47(1):1-11.

115. Takeda K, Mizushima T, Yokoyama Y, Hirose H, Wu X, Qian Y, et al. Sox2 is associated with cancer stem-like properties in colorectal cancer. Sci Rep. 2018;8(1):17639.

116. Pádua D, Barros R, Amaral AL, Mesquita P, Freire AF, Sousa M, et al. A SOX2 Reporter System Identifies Gastric Cancer Stem-Like Cells Sensitive to Monensin. Cancers (Basel). 2020;12(2).

117. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet. 2009;41(11):1238-42.

118. Boumahdi S, Driessens G, Lapouge G, Rorive S, Nassar D, Le Mercier M, et al. SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. Nature. 2014;511(7508):246-50.

119. Vanner RJ, Remke M, Gallo M, Selvadurai HJ, Coutinho F, Lee L, et al. Quiescent sox2(+) cells drive hierarchical growth and relapse in sonic hedgehog subgroup medulloblastoma. Cancer Cell. 2014;26(1):33-47.

120. Wuebben EL, Rizzino A. The dark side of SOX2: cancer - a comprehensive overview. Oncotarget. 2017;8(27):44917-43.

121. Lundberg IV, Löfgren Burström A, Edin S, Eklöf V, Öberg Å, Stenling R, et al. SOX2 expression is regulated by BRAF and contributes to poor patient prognosis in colorectal cancer. PloS one. 2014;9(7):e101957-e.

122. Lundberg IV, Edin S, Eklöf V, Öberg Å, Palmqvist R, Wikberg ML. SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer. BMC cancer. 2016;16:471-.

123. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet. 2006;38(4):431-40.

124. Han S-M, Han S-H, Coh Y-R, Jang G, Chan Ra J, Kang S-K, et al. Enhanced proliferation and differentiation of Oct4- and Sox2-overexpressing human adipose tissue mesenchymal stem cells. Experimental & molecular medicine. 2014;46(6):e101-e.

125. Hochedlinger K, Yamada Y, Beard C, Jaenisch R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell. 2005;121(3):465-77.

126. Atlasi Y, Mowla SJ, Ziaee SA, Bahrami AR. OCT-4, an embryonic stem cell marker, is highly expressed in bladder cancer. Int J Cancer. 2007;120(7):1598-602.

127. Saigusa S, Tanaka K, Toiyama Y, Yokoe T, Okugawa Y, Ioue Y, et al. Correlation of CD133, OCT4, and SOX2 in rectal cancer and their association with distant recurrence after chemoradiotherapy. Ann Surg Oncol. 2009;16(12):3488-98.

128. Chen Z, Xu WR, Qian H, Zhu W, Bu XF, Wang S, et al. Oct4, a novel marker for human gastric cancer. J Surg Oncol. 2009;99(7):414-9.

129. Peng S, Maihle NJ, Huang Y. Pluripotency factors Lin28 and Oct4 identify a sub-population of stem cell-like cells in ovarian cancer. Oncogene. 2010;29(14):2153-9.

130. Wen J, Park JY, Park KH, Chung HW, Bang S, Park SW, et al. Oct4 and Nanog expression is associated with early stages of pancreatic carcinogenesis. Pancreas. 2010;39(5):622-6.

131. Guo Y, Liu S, Wang P, Zhao S, Wang F, Bing L, et al. Expression profile of embryonic stem cell-associated genes Oct4, Sox2 and Nanog in human gliomas. Histopathology. 2011;59(4):763-75.

132. He W, Li K, Wang F, Qin YR, Fan QX. Expression of OCT4 in human esophageal squamous cell carcinoma is significantly associated with poorer prognosis. World J Gastroenterol. 2012;18(7):712-9.

133. Yin X, Li YW, Zhang BH, Ren ZG, Qiu SJ, Yi Y, et al. Coexpression of stemness factors Oct4 and Nanog predict liver resection. Ann Surg Oncol. 2012;19(9):2877-87.

134. Matsuoka J, Yashiro M, Sakurai K, Kubo N, Tanaka H, Muguruma K, et al. Role of the stemness factors sox2, oct3/4, and nanog in gastric carcinoma. J Surg Res. 2012;174(1):130-5.

135. Dai X, Ge J, Wang X, Qian X, Zhang C, Li X. OCT4 regulates epithelial-mesenchymal transition and its knockdown inhibits colorectal cancer cell migration and invasion. Oncology reports. 2013;29(1):155-60.

136. Basati G, Mohammadpour H, Emami Razavi A. Association of High Expression Levels of SOX2, NANOG, and OCT4 in Gastric Cancer Tumor Tissues with Progression and Poor Prognosis. J Gastrointest Cancer. 2020;51(1):41-7.

137. Miyoshi N, Fujino S, Ohue M, Yasui M, Takahashi Y, Sugimura K, et al. The POU5F1 gene expression in colorectal cancer: a novel prognostic marker. Surg Today. 2018;48(7):709-15.

138. Mertin S, McDowall SG, Harley VR. The DNA-binding specificity of SOX9 and other SOX proteins. Nucleic acids research. 1999;27(5):1359-64.

139. Schepers GE, Teasdale RD, Koopman P. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. Dev Cell. 2002;3(2):167-70.

140. Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell stem cell. 2013;12(1):15-30.

141. Schmitz M, Temme A, Senner V, Ebner R, Schwind S, Stevanovic S, et al. Identification of SOX2 as a novel glioma-associated antigen and potential target for T cell-based immunotherapy. Br J Cancer. 2007;96(8):1293-301.

142. Ye YW, Wu JH, Wang CM, Zhou Y, Du CY, Zheng BQ, et al. Sox17 regulates proliferation and cell cycle during gastric cancer progression. Cancer letters. 2011;307(2):124-31.

143. Lee Y-H, Saint-Jeannet J-P. Sox9 function in craniofacial development and disease. Genesis (New York, NY : 2000). 2011;49(4):200-8.

144. Mao DD, Gujar AD, Mahlokozera T, Chen I, Pan Y, Luo J, et al. A CDC20-APC/SOX2 Signaling Axis Regulates Human Glioblastoma Stem-like Cells. Cell Rep. 2015;11(11):1809-21.

145. Zhao T, Yang H, Tian Y, Xie Q, Lu Y, Wang Y, et al. SOX7 is associated with the suppression of human glioma by HMG-box dependent regulation of Wnt/β-catenin signaling. Cancer letters. 2016;375(1):100-7.

146. Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. Genes & development. 2001;15(10):1272-86.

147. Reményi A, Lins K, Nissen LJ, Reinbold R, Schöler HR, Wilmanns M. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes & development. 2003;17(16):2048-59.

148. Kondoh H, Kamachi Y. SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. Int J Biochem Cell Biol. 2010;42(3):391-9.

149. Aksoy I, Jauch R, Eras V, Chng WB, Chen J, Divakar U, et al. Sox transcription factors require selective interactions with Oct4 and specific transactivation functions to mediate reprogramming. Stem Cells. 2013;31(12):2632-46.

150. Merino F, Ng CKL, Veerapandian V, Schöler HR, Jauch R, Cojocaru V. Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. Structure. 2014;22(9):1274-86.

151. Tapia N, MacCarthy C, Esch D, Gabriele Marthaler A, Tiemann U, Araúzo-Bravo MJ, et al. Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. Sci Rep. 2015;5:13533.

152. Huang Y-H, Jankowski A, Cheah KSE, Prabhakar S, Jauch R. SOXE transcription factors form selective dimers on non-compact DNA motifs through multifaceted interactions between dimerization and high-mobility group domains. Scientific reports. 2015;5:10398-.

153. Narasimhan K, Pillay S, Huang YH, Jayabal S, Udayasuryan B, Veerapandian V, et al. DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors. Nucleic Acids Res. 2015;43(3):1513-28.

154. Xu YR, Yang WX. SOX-mediated molecular crosstalk during the progression of tumorigenesis. Semin Cell Dev Biol. 2017;63:23-34.

155. Penzo-Méndez AI. Critical roles for SoxC transcription factors in development and cancer. Int J Biochem Cell Biol. 2010;42(3):425-8.

156. Jafarnejad SM, Ardekani GS, Ghaffari M, Li G. Pleiotropic function of SRY-related HMG box transcription factor 4 in regulation of tumorigenesis. Cell Mol Life Sci. 2013;70(15):2677-96.

157. Wang L, Zhang J, Yang X, Chang YW, Qi M, Zhou Z, et al. SOX4 is associated with poor prognosis in prostate cancer and promotes epithelial-mesenchymal transition in vitro. Prostate Cancer Prostatic Dis. 2013;16(4):301-7.

158. Wang B, Li Y, Tan F, Xiao Z. Increased expression of SOX4 is associated with colorectal cancer progression. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(7):9131-7.

159. Wang Y, Yang W, Liu T, Bai G, Liu M, Wang W. Over-expression of SOX8 predicts poor prognosis in colorectal cancer: A retrospective study. Medicine (Baltimore). 2019;98(27):e16237.

160. Bastide P, Darido C, Pannequin J, Kist R, Robine S, Marty-Double C, et al. Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J Cell Biol. 2007;178(4):635-48.

161. Huch M, Clevers H. Sox9 marks adult organ progenitors. Nat Genet. 2011;43(1):9-10.

162. Schwitalla S, Fingerle AA, Cammareri P, Nebelsiek T, Göktuna SI, Ziegler PK, et al. Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. Cell. 2013;152(1-2):25-38.

163. Lü B, Fang Y, Xu J, Wang L, Xu F, Xu E, et al. Analysis of SOX9 expression in colorectal cancer. Am J Clin Pathol. 2008;130(6):897-904.

164. Espersen ML, Olsen J, Linnemann D, Høgdall E, Troelsen JT. Clinical implications of intestinal stem cell markers in colorectal cancer. Clin Colorectal Cancer. 2015;14(2):63-71.

165. Javier BM, Yaeger R, Wang L, Sanchez-Vega F, Zehir A, Middha S, et al. Recurrent, truncating SOX9 mutations are associated with SOX9 overexpression, KRAS mutation, and TP53 wild type status in colorectal carcinoma. Oncotarget. 2016;7(32):50875-82.

166. Huang W, Chen Z, Shang X, Tian D, Wang D, Wu K, et al. Sox12, a direct target of FoxQ1, promotes hepatocellular carcinoma metastasis through up-regulating Twist1 and FGFBP1. Hepatology. 2015;61(6):1920-33.

167. Du F, Chen J, Liu H, Cai Y, Cao T, Han W, et al. SOX12 promotes colorectal cancer cell proliferation and metastasis by regulating asparagine synthesis. Cell death & disease. 2019;10(3):239.

168. Du F, Li X, Feng W, Qiao C, Chen J, Jiang M, et al. SOX13 promotes colorectal cancer metastasis by transactivating SNAI2 and c-MET. Oncogene. 2020;39(17):3522-40.

169. Miao Z, Deng X, Shuai P, Zeng J. Upregulation of SOX18 in colorectal cancer cells promotes proliferation and correlates with colorectal cancer risk. Onco Targets Ther. 2018;11:8481-90.

170. Leng Z, Tao K, Xia Q, Tan J, Yue Z, Chen J, et al. Kruppel-like factor 4 acts as an oncogene in colon cancer stem cell-enriched spheroid cells. PloS one. 2013;8(2):e56082.

171. Shie JL, Chen ZY, O'Brien MJ, Pestell RG, Lee ME, Tseng CC. Role of gut-enriched Kruppel-like factor in colonic cell growth and differentiation. Am J Physiol Gastrointest Liver Physiol. 2000;279(4):G806-14.

172. Wei D, Kanai M, Huang S, Xie K. Emerging role of KLF4 in human gastrointestinal cancer. Carcinogenesis. 2006;27(1):23-31.

173. Lin Q, Lai R, Chirieac LR, Li C, Thomazy VA, Grammatikakis I, et al. Constitutive activation of JAK3/STAT3 in colon carcinoma tumors and cell lines: inhibition of JAK3/STAT3 signaling induces apoptosis and cell cycle arrest of colon carcinoma cells. Am J Pathol. 2005;167(4):969-80.

174. Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. Nat Rev Cancer. 2009;9(11):798-809.

175. Lin L, Liu A, Peng Z, Lin HJ, Li PK, Li C, et al. STAT3 is necessary for proliferation and survival in colon cancer-initiating cells. Cancer Res. 2011;71(23):7226-37.

176. Munro MJ, Wickremesekera SK, Peng L, Tan ST, Itinteang T. Cancer stem cells in colorectal cancer: a review. J Clin Pathol. 2018;71(2):110-6.

177. Chung SS, Adekoya D, Enenmoh I, Clarke O, Wang P, Sarkyssian M, et al. Salinomycin Abolished STAT3 and STAT1 Interactions and Reduced Telomerase Activity in Colorectal Cancer Cells. Anticancer research. 2017;37(2):445-53.

178. Dang CV. MYC, metabolism, cell growth, and tumorigenesis. Cold Spring Harb Perspect Med. 2013;3(8).

179. Yang L, Shi P, Zhao G, Xu J, Peng W, Zhang J, et al. Targeting cancer stem cell pathways for cancer therapy. Signal Transduct Target Ther. 2020;5(1):8.

180. Wang W, Deng J, Wang Q, Yao Q, Chen W, Tan Y, et al. Synergistic role of Cul1 and c-Myc: Prognostic and predictive biomarkers in colorectal cancer. Oncology reports. 2017;38(1):245-52.

181. Zhang HL, Wang P, Lu MZ, Zhang SD, Zheng L. c-Myc maintains the self-renewal and chemoresistance properties of colon cancer stem cells. Oncol Lett. 2019;17(5):4487-93.

182. Yang Z, Zhang C, Qi W, Cui Y, Xuan Y. GLI1 promotes cancer stemness through intracellular signaling pathway PI3K/Akt/NFkappaB in colorectal adenocarcinoma. Exp Cell Res. 2018;373(1-2):145-54.

183. Varnat F, Duquet A, Malerba M, Zbinden M, Mas C, Gervaz P, et al. Human colon cancer epithelial cells harbour active HEDGEHOG-GLI signalling that is essential for tumour growth, recurrence, metastasis and stem cell survival and expansion. EMBO molecular medicine. 2009;1(6-7):338-51.

184. Forghanifard MM, Moghbeli M, Raeisossadati R, Tavassoli A, Mallak AJ, Boroumand-Noughabi S, et al. Role of SALL4 in the progression and metastasis of colorectal cancer. J Biomed Sci. 2013;20(1):6-.

185. Zhang X, Yuan X, Zhu W, Qian H, Xu W. SALL4: an emerging cancer biomarker and target. Cancer letters. 2015;357(1):55-62.

186. Behrens J, von Kries JP, Kuhl M, Bruhn L, Wedlich D, Grosschedl R, et al. Functional interaction of beta-catenin with the transcription factor LEF-1. Nature. 1996;382(6592):638-42.

187. Kolligs FT, Bommer G, Goke B. Wnt/beta-catenin/tcf signaling: a critical pathway in gastrointestinal tumorigenesis. Digestion. 2002;66(3):131-44.

188. Kanwar SS, Yu Y, Nautiyal J, Patel BB, Majumdar AP. The Wnt/beta-catenin pathway regulates growth and maintenance of colonospheres. Mol Cancer. 2010;9:212.

189. Chen Z, He X, Jia M, Liu Y, Qu D, Wu D, et al. β-catenin overexpression in the nucleus predicts progress disease and unfavourable survival in colorectal cancer: a meta-analysis. PloS one. 2013;8(5):e63854-e.

190. Segditsas S, Tomlinson I. Colorectal cancer and genetic alterations in the Wnt pathway. Oncogene. 2006;25(57):7531-7.

191. Shin SH, Lim DY, Reddy K, Malakhova M, Liu F, Wang T, et al. A Small Molecule Inhibitor of the β-Catenin-TCF4 Interaction Suppresses Colorectal Cancer Growth In Vitro and In Vivo. EBioMedicine. 2017;25:22-31.

192. Sun S, Yang X, Qin X, Zhao Y. TCF4 promotes colorectal cancer drug resistance and stemness via regulating ZEB1/ZEB2 expression. Protoplasma. 2020;257(3):921-30.

193. Kriegl L, Horst D, Reiche JA, Engel J, Kirchner T, Jung A. LEF-1 and TCF4 expression correlate inversely with survival in colorectal cancer. Journal of translational medicine. 2010;8:123.

194. Chiappetta G, Avantaggiato V, Visconti R, Fedele M, Battista S, Trapasso F, et al. High level expression of the HMGI (Y) gene during embryonic development. Oncogene. 1996;13(11):2439-46.

195. Resar LM. The high mobility group A1 gene: transforming inflammatory signals into cancer? Cancer Res. 2010;70(2):436-9.

196. Shah SN, Kerr C, Cope L, Zambidis E, Liu C, Hillion J, et al. HMGA1 reprograms somatic cells into pluripotent stem cells by inducing stem cell transcriptional networks. PloS one. 2012;7(11):e48533.

197. Fusco A, Fedele M. Roles of HMGA proteins in cancer. Nat Rev Cancer. 2007;7(12):899-910.

198. Grade M, Hörmann P, Becker S, Hummon AB, Wangsa D, Varma S, et al. Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas. Cancer research. 2007;67(1):41-56.

199. E. MP, Liu T, Zhang X, Yang H, Wang J, Huang R, et al. High-mobility group A1 (HMGA1) gene expressions in various colorectal cancer cell lines and correlation with prognosis. 2020. 2020;9(2):763-73.

200. Belton A, Gabrovsky A, Bae YK, Reeves R, Iacobuzio-Donahue C, Huso DL, et al. HMGA1 induces intestinal polyposis in transgenic mice and drives tumor progression and stem cell properties in colon cancer cells. PloS one. 2012;7(1):e30034-e.

201. Shah SN, Resar LM. High mobility group A1 and cancer: potential biomarker and therapeutic target. Histol Histopathol. 2012;27(5):567-79.

202. Xian L, Georgess D, Huso T, Cope L, Belton A, Chang YT, et al. HMGA1 amplifies Wnt signalling and expands the intestinal stem cell compartment and Paneth cell niche. Nature communications. 2017;8:15008.

203. Zhang S, Mo Q, Wang X. Oncological role of HMGA2 (Review). Int J Oncol. 2019;55(4):775-88.

204. Wang X, Liu X, Li AY-J, Chen L, Lai L, Lin HH, et al. Overexpression of HMGA2 promotes metastasis and impacts survival of colorectal cancers. Clinical cancer research : an official journal of the American Association for Cancer Research. 2011;17(8):2570-80.

205. Xu X, Wang Y, Deng H, Liu C, Wu J, Lai M. HMGA2 enhances 5-fluorouracil chemoresistance in colorectal cancer via the Dvl2/Wnt pathway. Oncotarget. 2018;9(11):9963-74.

206. Sano H, Leboeuf JP, Novitskiy SV, Seo S, Zaja-Milatovic S, Dikov MM, et al. The Foxc2 transcription factor regulates tumor angiogenesis. Biochemical and biophysical research communications. 2010;392(2):201-6.

207. Watanabe T, Kobunai T, Yamamoto Y, Matsuda K, Ishihara S, Nozawa K, et al. Gene expression of mesenchyme forkhead 1 (FOXC2) significantly correlates with the degree of lymph node metastasis in colorectal cancer. Int Surg. 2011;96(3):207-16.

208. Yang C, Cui X, Dai X, Liao W. Downregulation of Foxc2 enhances apoptosis induced by 5-fluorouracil through activation of MAPK and AKT pathways in colorectal cancer. Oncol Lett. 2016;11(2):1549-54.

209. Merlos-Suárez A, Barriga FM, Jung P, Iglesias M, Céspedes MV, Rossell D, et al. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. Cell stem cell. 2011;8(5):511-24.

210. Esch D, Vahokoski J, Groves MR, Pogenberg V, Cojocaru V, Vom Bruch H, et al. A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat Cell Biol. 2013;15(3):295-301.

211. Maddox J, Shakya A, South S, Shelton D, Andersen JN, Chidester S, et al. Transcription factor Oct1 is a somatic and cancer stem cell determinant. PLoS genetics. 2012;8(11):e1003048.

212. Li Y, Dong M, Kong F, Zhou J. Octamer transcription factor 1 mediates epithelial-mesenchymal transition in colorectal cancer. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2015;36(12):9941-6.

213. Wang Y-P, Song G-H, Chen J, Xiao C, Li C, Zhong L, et al. Elevated OCT1 participates in colon tumorigenesis and independently predicts poor prognoses of colorectal cancer patients. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(3):3247-55.

214. Cejas P, Cavazza A, Yandava CN, Moreno V, Horst D, Moreno-Rubio J, et al. Transcriptional Regulator CNOT3 Defines an Aggressive Colorectal Cancer Subtype. Cancer research. 2017;77(3):766-79.

215. Yuan Y, Sun S, Jiao N, Shu Y, Zhang Y. Upregulation of HOXA10 Protein Expression Predicts Poor Prognosis for Colorectal Cancer. Genet Test Mol Biomarkers. 2018;22(6):390-7.

216. Wan Z, Chai R, Yuan H, Chen B, Dong Q, Zheng B, et al. MEIS2 promotes cell migration and invasion in colorectal cancer. Oncology reports. 2019;42(1):213-23.

217. Wang X, Ghareeb WM, Zhang Y, Yu Q, Lu X, Huang Y, et al. Hypermethylated and downregulated MEIS2 are involved in stemness properties and oxaliplatin-based chemotherapy resistance of colorectal cancer. J Cell Physiol. 2019;234(10):18180-91.

218. Cox JL, Mallanna SK, Luo X, Rizzino A. Sox2 uses multiple domains to associate with proteins present in Sox2-protein complexes. PloS one. 2010;5(11):e15486-e.

219. Mallanna SK, Ormsbee BD, Iacovino M, Gilmore JM, Cox JL, Kyba M, et al. Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. Stem cells (Dayton, Ohio). 2010;28(10):1715-27.

220. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif). 2001;25(4):402-8.

221. Francipane MG, Bulanin D, Lagasse E. Establishment and Characterization of 5-Fluorouracil-Resistant Human Colorectal Cancer Stem-Like Cells: Tumor Dynamics under Selection Pressure. Int J Mol Sci. 2019;20(8):1817.

222. Davis TL, Rabinovitz I, Futscher BW, Schnölzer M, Burger F, Liu Y, et al. Identification of a novel structural variant of the alpha 6 integrin. J Biol Chem. 2001;276(28):26099-106.

223. Lv L, Liu HG, Dong SY, Yang F, Wang QX, Guo GL, et al. Upregulation of CD44v6 contributes to acquired chemoresistance via the modulation of autophagy in colon cancer SW480 cells. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(7):8811-24.

224. Alowaidi F, Hashimi SM, Alqurashi N, Alhulais R, Ivanovski S, Bellette B, et al. Assessing stemness and proliferation properties of the newly established colon cancer 'stem' cell line, CSC480 and novel approaches to identify dormant cancer cells. Oncology reports. 2018;39(6):2881-91.

225. Ren L, Zhou T, Wang Y, Wu Y, Xu H, Liu J, et al. RNF8 induces β-catenin-mediated c-Myc expression and promotes colon cancer proliferation. Int J Biol Sci. 2020;16(12):2051-62.

226. Puca F, Colamaio M, Federico A, Gemei M, Tosti N, Bastos AU, et al. HMGA1 silencing restores normal stem cell characteristics in colon cancer stem cells by increasing p53 levels. Oncotarget. 2014;5(10):3234-45.

227. Zhang N, Zhang R, Zou K, Yu W, Guo W, Gao Y, et al. Keratin 23 promotes telomerase reverse transcriptase expression and human colorectal cancer growth. Cell death & disease. 2017;8(7):e2961.

228. Chin IK-N, Cheung GH-Y, Kwok HY-Y, Ma VW-S, Lau EY-T, Cho WC-S. Abstract 3067: The novel function of keratin 14 (KRT14) in regulating cancer stem cell properties in lung adenocarcinoma. Cancer Research. 2018;78(13 Supplement):3067.

229. Yang B, Zhang W, Zhang M, Wang X, Peng S, Zhang R. KRT6A Promotes EMT and Cancer Stem Cell Transformation in Lung Adenocarcinoma. Technology in cancer research & treatment. 2020;19:1533033820921248.

230. Saha SK, Kim K, Yang GM, Choi HY, Cho SG. Cytokeratin 19 (KRT19) has a Role in the Reprogramming of Cancer Stem Cell-Like Cells to Less Aggressive and More Drug-Sensitive Cells. Int J Mol Sci. 2018;19(5).

231. Zhu F, Qian W, Zhang H, Liang Y, Wu M, Zhang Y, et al. SOX2 Is a Marker for Stem-like Tumor Cells in Bladder Cancer. Stem cell reports. 2017;9(2):429-37.

232. Choi J, Kim DI, Kim J, Kim BH, Kim A. Hornerin Is Involved in Breast Cancer Progression. Journal of breast cancer. 2016;19(2):142-7.

233. Fu SJ, Shen SL, Li SQ, Hua YP, Hu WJ, Guo B, et al. Hornerin promotes tumor progression and is associated with poor prognosis in hepatocellular carcinoma. BMC cancer. 2018;18(1):815.

234. Fang E, Zhang X. Identification of breast cancer hub genes and analysis of prognostic values using integrated bioinformatics analysis. Cancer biomarkers : section A of Disease markers. 2017;21(1):373-81.

235. Lü B, Xu J, Zhu Y, Zhang H, Lai M. Systemic analysis of the differential gene expression profile in a colonic adenoma-normal SSH library. Clinica chimica acta; international journal of clinical chemistry. 2007;378(1-2):42-7.

236. Fang X, Yoon JG, Li L, Tsai YS, Zheng S, Hood L, et al. Landscape of the SOX2 protein-protein interactome. Proteomics. 2011;11(5):921-34.

237. Zhang Y, Zhang X, Wang X, Gan L, Yu G, Chen Y, et al. Inhibition of LDH-A by lentivirus-mediated small interfering RNA suppresses intestinal-type gastric cancer tumorigenicity through the downregulation of Oct4. Cancer letters. 2012;321(1):45-54.

238. Ji Y, Yang C, Tang Z, Yang Y, Tian Y, Yao H, et al. Adenylate kinase hCINAP determines self-renewal of colorectal cancer stem cells by facilitating LDHA phosphorylation. Nature communications. 2017;8:15308.

239. Nagai M, Natsumeda Y, Konno Y, Hoffman R, Irino S, Weber G. Selective up-regulation of type II inosine 5'-monophosphate dehydrogenase messenger RNA expression in human leukemias. Cancer Res. 1991;51(15):3886-90.

240. Kozhevnikova EN, van der Knaap JA, Pindyurin AV, Ozgur Z, van Ijcken WF, Moshkin YM, et al. Metabolic enzyme IMPDH is also a transcription factor regulated by cellular state. Molecular cell. 2012;47(1):133-9.

241. Jackson RC, Weber G, Morris HP. IMP dehydrogenase, an enzyme linked with proliferation and malignancy. Nature. 1975;256(5515):331-3.

242. Duan S, Huang W, Liu X, Liu X, Chen N, Xu Q, et al. IMPDH2 promotes colorectal cancer progression through activation of the PI3K/AKT/mTOR and PI3K/AKT/FOXO1 signaling pathways. Journal of experimental & clinical cancer research : CR. 2018;37(1):304.

243. Huang F, Ni M, Chalishazar MD, Huffman KE, Kim J, Cai L, et al. Inosine Monophosphate Dehydrogenase Dependence in a Subset of Small Cell Lung Cancers. Cell metabolism. 2018;28(3):369-82.e5.

244. Cox JL, Wilder PJ, Gilmore JM, Wuebben EL, Washburn MP, Rizzino A. The SOX2-interactome in brain cancer cells identifies the requirement of MSI2 and USP9X for the growth of brain tumor cells. PloS one. 2013;8(5):e62857.

245. Markowitz J, Carson WE, 3rd. Review of S100A9 biology and its role in cancer. Biochim Biophys Acta. 2013;1835(1):100-9.

246. Yang WS, Moon HG, Kim HS, Choi EJ, Yu MH, Noh DY, et al. Proteomic approach reveals FKBP4 and S100A9 as potential prediction markers of therapeutic response to neoadjuvant chemotherapy in patients with breast cancer. Journal of proteome research. 2012;11(2):1078-88.

247. Wang Y, Yin K, Tian J, Xia X, Ma J, Tang X, et al. Granulocytic Myeloid-Derived Suppressor Cells Promote the Stemness of Colorectal Cancer Cells through Exosomal S100A9. Advanced science (Weinheim, Baden-Wurttemberg, Germany). 2019;6(18):1901278.

248. Chen S, Zhao H, Deng J, Liao P, Xu Z, Cheng Y. Comparative proteomics of glioma stem cells and differentiated tumor cells identifies S100A9 as a potential therapeutic target. Journal of cellular biochemistry. 2013;114(12):2795-808.

249. D'Angelo D, Mussnich P, Rosa R, Bianco R, Tortora G, Fusco A. High mobility group A1 protein expression reduces the sensitivity of colon and thyroid cancer cells to antineoplastic drugs. BMC cancer. 2014;14:851.

250. Resar L, Chia L, Xian L. Lessons from the Crypt: HMGA1-Amping up Wnt for Stem Cells and Tumor Progression. Cancer Res. 2018;78(8):1890-7.

251. Kim DK, Seo EJ, Choi EJ, Lee SI, Kwon YW, Jang IH, et al. Crucial role of HMGA1 in the self-renewal and drug resistance of ovarian cancer stem cells. Experimental & molecular medicine. 2016;48(8):e255.