UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas

Development of a mathematical model for early prediction of relapses and assessment of chemotherapy resistance in colorectal cancer

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Marta Sánchez Díez
Bachelor's in Biotechnology
Master's in Biomolecules and Cell Dynamics

Madrid, 2025

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Agronómica,
Alimentaria y de Biosistemas

Doctoral Degree in Complex Systems

Development of a mathematical model for early prediction of relapses and assessment of chemotherapy resistance in colorectal cancer

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Marta Sánchez Díez
Bachelor's in Biotechnology
Master's in Biomolecules and Cell Dynamics

Under the supervision of:

Dr. Juan Manuel Pastor Ruiz
Dra. María del Carmen Ramírez Castillejo

Madrid, 2025

Title: Development of a mathematical model for early prediction of relapses and assessment of chemotherapy resistance in colorectal cancer

Author: Marta Sánchez Díez

Doctoral Programme: Complex Systems

Thesis Supervision:

Dr. Juan Manuel Pastor Ruiz, Associate Professor (Universidad Politécnica de Madrid)

Dra. María del Carmen Ramírez Castillejo, Associate Professor (Universidad Politécnica de Madrid)

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been partially supported by “Programa Propio UPM” both through the doctoral studies grant and the international placement grant.

A quienes me acompañan día a día en cada aventura, y a quienes, aunque no estén de cuerpo presente, permanecen en mi corazón.

Agradecimientos

Hacer una tesis es como correr una maratón sin conocer la distancia exacta: a veces la meta parece cercana, otras inalcanzable. Una maratón no se consigue si no es con el acompañamiento y motivación de todo el mundo que rodea esta carrera. Y así ha sido en mi caso, este camino lo he recorrido gracias a muchas personas, a las que les debo cada meta alcanzada. Por ello, quiero agradecer a quienes han sido mi aliento cuando flaqueaba, mi avituallamiento en los momentos de mayor desgaste y mi brújula cuando me perdía. Por supuesto, también a aquellas personas que han estado ahí para celebrar cada hito alcanzado, en esos momentos de subidón por haber conseguido cumplir los objetivos y obtener buenos resultados.

Me gustaría agradecer a todas aquellas personas que han contribuido en mi formación desde el inicio, a todos los profesores del Grado en Biotecnología de la Universidad Politécnica de Madrid. En especial a Juan Orellana, pues no solo fue un pilar en mi formación sino también a nivel personal. Para mi fuiste una guía, un ejemplo a seguir y tengo muy buenos recuerdos. También a Soledad Sacristán, pues mi primera experiencia en un laboratorio fue en el tuyo. Gracias por el apoyo y por todo lo que aprendí de ti en esos meses. Por supuesto a Patricia Giraldo, gracias por esa gran oportunidad que me ofreciste, la de ser mi tutora académica, de un gran trabajo que llevé a cabo en el laboratorio del Dr. Lluís Montoliu. Gracias de corazón a los dos, porque de ahí me llevo gran parte del conocimiento y experiencia que tengo ahora.

Cómo no, a Carmen Ramírez, que fue mi profesora en varias asignaturas de la carrera y posteriormente me dio la oportunidad de formar parte de su laboratorio, donde he llevado a cabo esta tesis doctoral. Comenzamos montando el laboratorio allí en la escuela, haciendo hueco para poder comenzar con este gran proyecto ¿te acuerdas? Muchas gracias por todo tu apoyo, por luchar por nosotros en todo momento, por apagar fuegos cuando más lo necesitábamos, por buscar ante todo un buen futuro para nosotros. Gracias por la oportunidad que me diste, porque he aprendido muchísimo en estos años.

A mi otro director de tesis, Juan Manuel Pastor. En las primeras reuniones en tu despacho, ¿quién nos iba a decir que nos íbamos a entender tan bien al final de la tesis? Muchísimas gracias, Juanma, por enseñarme tantísimas cosas en un mundo tan desconocido para mí. Sin lugar a duda, sin ti no hubiera sido posible llevar a cabo esta tesis doctoral. Quiero agradecerte tu disponibilidad, tu apoyo en todo momento y tu interés por entender la biología detrás de los números.

Es imposible no mencionar los inicios de este proyecto, en el laboratorio de la escuela elucubrando con Javier Guerrero, Javier Pareja y Beatriz Cózar, cómo podríamos implementar un sistema diagnóstico en el Sistema Nacional de Salud. Entre risas, dejábamos volar la imaginación, intentando cumplir un sueño, que hoy, gracias a esta tesis doctoral, creo que está más cerca de hacerse realidad. Muchas gracias por todos esos momentos.

No tengo palabras para agradecer a todo el equipo de laboratorio, por vuestro apoyo durante el proyecto, por esas charlas reparadoras, y esas quedadas, que, aunque no del todo frecuentes, fueron chutes de energía para seguir adelante. Gracias a Josefa, por esa compañía desde el inicio y sobre todo en los momentos finales de estrés para presentar la tesis. Nico, por todas esas risas y el cariño que me has dado este tiempo. Es un auténtico placer haber conocido a esta pareja tan maravillosa. Raquel, muchísimas gracias por tu apoyo y esas conversaciones tan necesarias para despejar la mente de vez en cuando. Clara, porque además de amiga has sido un pilar fundamental en el laboratorio. Javi, gracias por desprender tanta felicidad en todo momento. Ha sido un placer compartir este camino con todos vosotros.

Quiero agradecer también a todos los estudiantes que han pasado por el laboratorio, en especial Marta López y Ana García, por su apoyo incondicional en el proyecto, por su disposición y ganas de aprender. Ha sido un placer trabajar con vosotras.

I extend my deepest gratitude to Dr. Frédéric Hollande for the invaluable opportunity to be part of his research group for three months. During this time, I have learned a lot both from you and your team. Your guidance has been instrumental in advancing this project. Culturing cancer stem cells is a challenging task, but thanks to your expertise, we achieved significant progress. I truly appreciate the way you looked after me during my stay, ensuring my placement was possible despite the tight timeframe. Special thanks to Robin for the warm welcome, for integrating me into your project, and for generously sharing your knowledge. To both Robin and Maruša, thank you for the great moments and laughter. We built an amazing team, I enjoyed being part of it a lot. Finally, thanks to Morgan, Sarah, Antonin, Jessica, Dijina and Oliver, for your kindness and making me feel so welcome. It was a pleasure to share those three months with such a dedicated and welcoming team.

A todos mis amigos de Tudela, en especial al Kun, las Elenas, Andrea y Yaiza. Gracias por estar siempre ahí, por hacerme reír, por tantos buenos momentos juntos. Por ser un apoyo siempre que lo he necesitado, por preguntar qué tal va mi tesis a pesar de la lejanía en el campo. Gracias de corazón.

A mis más fieles compañeras y amigas desde primero de carrera. A Coral, Clara y Paula, no sabéis lo feliz que me hace compartir este momento con vosotras. Desde que nos conocimos hemos sido un grupo ejemplar, porque por más que haya distancia de por medio, nuestra relación no cambia. Porque sois lo más grande, y os quiero un montón. Porque sois un gran apoyo, siempre, da igual el momento, siempre estáis ahí para lo que necesite. Coral, por esos momentos de risas juntas, porque somos imparables, nos da igual dónde y con quién, la diversión está garantizada. Porque eres un amor y me encanta estar contigo. Clara, porque eres tan dulce, tan buena amiga, tan original, tan … todo. Por esas vueltas a casa desde hace muchos años, por esa compañía a todos lados, por ese apoyo durante toda mi tesis, por esas charlas a velocidad de la luz. Paula, ¿te acuerdas de esas primeras prácticas juntas? Creo que ya no la lío tanto, al menos he aprendido a usar un frasco lavador. Por esos momentos divertidos durante la carrera, porque, aunque estemos lejos, cuando hablamos, siento la misma energía que cuando estábamos juntas. Porque me alegro de todos tus éxitos y porque siempre estás ahí. No me olvido de vosotros, Javi y Dani. Muchas gracias por esos buenos momentos en la universidad y por mantener nuestra amistad, aunque los caminos se separen.

Estoy muy orgullosa de la familia que tengo, de lo unidos que estamos todos. Eso es gracias a mi abuela Luisa, que hacía lo imposible por juntarnos a todos cada domingo. Gracias a mis tíos y primos, y en especial a mis abuelos. Ellos fueron un pilar muy importante desde que era pequeña, y gracias a ellos soy quien soy. Sé que estaríais muy orgullosos de verme cumplir esta gran meta. Os quiero.

Gracias a las Ericas, Francisco y Alberto, habéis sido un gran apoyo durante estos años. Gracias por hacerme sentir una más de la familia, por ese cariño que me habéis dado, por compartir todos los momentos juntos. Gracias a Lucía, porque da igual en qué momento esté, esa carita siempre me saca una sonrisa. Por esos: "¡¡¡Tía Marta!!!" que me derriten.

A mis padres. Mamá, desde que éramos pequeños has sido nuestro gran pilar, el ejemplo a seguir. Un ejemplo de superación, gracias a ti he podido llegar hasta aquí. Sé lo muy orgullosa que estás de mí, y eso me hace muy feliz. Gracias por la oportunidad que me has dado siempre, por apoyarme y cuidarme. Papá, ojalá pudieras estar aquí para verme cumplir esta gran meta. Me acuerdo cada día de ti, cada uno de mis logros siempre irán dedicados con mucho cariño para ti. Sé que estarías muy orgulloso de mi. Siempre te llevaré en el corazón. Os quiero.

A mis hermanos, ¿qué seríamos los unos sin los otros? Si algo hemos aprendido a lo largo de los años es a apoyarnos cuando el otro lo necesita. Compartimos cada alegría y cada éxito juntos. Es muy divertido ser cuatro, siempre hay alguien en casa con quien pasarlo bien. Laura, porque eres mi bebo, mi pequeña a la que nadie puede tocar. Gracias por todas las cosas nuevas que me enseñas. A pesar de que seas la pequeña, podemos ser transparentes contigo y eres madura como para ser uno más. Gracias por entendernos y ayudarnos. Quiero que sepas que no cambiaría por nada del mundo todos tus besos y abrazos, eres mi pequeño gran amor. Dani, por tus palabras tranquilas y alentadoras. Da igual lo que pase que un abrazo tuyo hace sentir que todo está bien. Porque nuestras llamadas de "5 minutos" se convierten en llamadas de dos horas, nos da igual de qué hablar y qué contarnos. Eres un gran apoyo para mí y siempre nos has cuidado, desde bien pequeñas. Te debemos mucho, gracias a ti también salimos adelante. Patri, ¿no serás mi fiel compañera de vida? Todo lo hacemos juntas desde pequeñas, somos inseparables. Porque el odio de las peleas de pequeñas se convierte en un amor incondicional de mayor. Eres la persona a la que primero llamo, porque juntas somos más. Eres mi refugio en los días grises y mi luz en los días felices. Gracias por todos esos paseos cuando más lo necesitaba, por cuidarme siempre. Os quiero.

Por último, a Diego, mi amor. Quiero que sepas que a tu lado soy inmensamente feliz, que cada momento contigo es un regalo y que no cambiaría nada de lo que somos. Gracias por ser mi punto de apoyo durante toda la tesis, en especial durante estos últimos meses. Juntos hemos superado todos los obstáculos que han aparecido por el camino. Gracias a esta tesis ahora eres un poquito más científico. Muchas gracias por poner de tu parte para entender este mundo tan complejo. Gracias por estar ahí en todo momento, en los días buenos de emoción y celebración y en aquellos días no tan buenos donde lo que necesitaba era un empujón. Gracias por ser la persona que tengo a mi lado cuando me despierto y cuando me acuesto. Gracias por estar siempre ahí. Te quiero.

Abstract

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths and the third most common cancer worldwide. While primary tumors are effectively managed in most cases, approximately 20–40% of patients experience recurrence within the first two to three years after treatment. Cancer stem cells (CSCs) are most responsible for relapse due to their unique characteristics, including their low division rate, their self-renewal capacity, and their ability to evade the primary tumor, travel through the bloodstream (due to its epithelial-mesenchymal transition capacity), and colonize other locations, both proximally and distally (metastasis). Moreover, traditional chemotherapy treatments target rapidly dividing cells, leaving CSCs unaffected due to their slow division rate. This also makes them key contributors to chemotherapy resistance. Although several biomarkers have been described in the literature for the detection of CSCs, none are exclusive to this cell population. This limitation highlights the need for a combination of biomarkers to improve early relapse prediction.

In this PhD thesis, we propose the use of a panel of biomarkers for CSCs to enhance the early prediction of relapses in CRC patients. Additionally, we focus on studying resistance to treatment, a significant challenge as many patients develop chemotherapy-induced toxicity and relapse due to ineffective treatments. This underscores the need for new strategies to accurately measure chemotherapy efficacy and to identify novel biomarkers for resistance detection. Precision medicine, especially through the use of patient-derived tumors, plays a critical role in studying resistance and selecting the most effective treatment for each patient.

The main objectives of this PhD thesis were to develop a mathematical model for early relapse prediction and to study resistance in CRC. To achieve these goals, we first selected biomarkers reported in the literature and measured their expression levels in CRC cell lines and in cells isolated from the blood of patients diagnosed with CRC. The expression levels together with the clinicopathological characteristics of the patients were the inputs for the developed mathematical model for early prediction of relapses. To achieve this, we used Machine Learning (ML) models, with Logistic Regression yielding the best results. This model achieved a specificity of 82.4% and a sensitivity of 82.6%, offering reliable predictive capabilities. We also analyzed the contribution of each biomarker to the model to better understand their roles in relapse prediction.

After predicting relapses, we shifted our focus to resistance to chemotherapy. Our analysis revealed that 40 and 60% of the patients in the study experienced treatment resistance or toxicity, respectively. To address the limitations of the conventional IC_{50} method to measure the efficacy of different treatments, particularly its time-dependent nature, we developed a novel growth rate-based method. This method allowed us to compare the resistance levels of various CRC cell lines and to assess the expression of different biomarkers to identify candidates associated with resistance.

Through this work, we identified trophoblast cell surface antigen 2 (TROP2), a transmembrane glycoprotein and calcium signal transducer, as a potential biomarker for resistance. We observed that the expression of TROP2, both at the gene and protein levels, correlated strongly with resistance. Further investigations revealed that chronic exposure to chemotherapy led to increased expression of TROP2. Additionally, we explored the effects of silencing the TROP2 gene and observed its influence on resistance. Since resistance was assessed through a cell viability assay, the impact of TROP2 on proliferation complicated the experiment. Despite these challenges and the need for further studies, the findings support TROP2 as a promising biomarker for resistance detection.

In conclusion, this PhD thesis resulted in the development of a robust mathematical model for the early prediction of relapse in CRC patients. We also introduced a novel growth rate-based method for studying chemotherapy resistance and identified TROP2 as a promising biomarker candidate for resistance detection. These findings contribute to advancing precision medicine and improving the quality of life for colorectal cancer patients by enabling more targeted and effective interventions, maximizing response rates, minimizing unnecessary treatments, and reducing side effects.

Resumen

El cáncer colorrectal (CRC) es la segunda causa principal de muertes relacionadas con cáncer y el tercer tipo de cáncer más común en todo el mundo. Aunque los tumores primarios suelen superarse en la mayoría de los casos, aproximadamente el 20-40% de los pacientes experimentan recaídas en los primeros dos o tres años tras el tratamiento. Las células madre cancerígenas (CSCs) son las principales responsables de las recaídas debido a sus características, que incluyen una baja tasa de división, capacidad de autorrenovación y la habilidad de abandonar el tumor primario, viajar a través del torrente sanguíneo (por su capacidad de transición epitelio-mesénquima) y colonizar otros órganos, tanto proximales como distales (metástasis). Además, los tratamientos de quimioterapia tradicionales están diseñados para atacar células de rápida división, por lo que las CSCs permanecen intactas debido a su lenta tasa de división. Esto las convierte en un factor clave en la resistencia a la quimioterapia. Aunque en la literatura se han descrito varios biomarcadores para la detección de CSCs, ninguno es exclusivo de esta población. Esta limitación resalta la necesidad de combinar biomarcadores para mejorar la predicción de recaídas.

En esta tesis doctoral, proponemos el uso de un panel de biomarcadores de CSCs para mejorar la predicción de recaídas en pacientes con CRC. Además, nos enfocamos en el estudio de la resistencia al tratamiento, un desafío importante, ya que muchos pacientes desarrollan toxicidad tras la administración de quimioterapia y recaen debido a la ineficiencia de los tratamientos. Esto subraya la necesidad de desarrollar nuevas estrategias para medir con precisión la efectividad de los tratamientos quimioterapéuticos e identificar biomarcadores novedosos para la detección de resistencia. La medicina de precisión, particularmente a través del desarrollo de tumores derivados de pacientes, desempeña un papel fundamental en el estudio de la resistencia y en la selección del tratamiento más efectivo para cada paciente.

Los principales objetivos de esta tesis doctoral fueron desarrollar un modelo matemático para la predicción temprana de recaídas y estudiar la resistencia en el CRC. Para alcanzar estos objetivos, primero seleccionamos biomarcadores descritos en la literatura y medimos sus niveles de expresión en líneas celulares de CRC y en células aisladas de la sangre de pacientes diagnosticados con CRC.

Los niveles de expresión, junto con la historia clínica de los pacientes, se utilizaron como variables de entrada en el modelo matemático que desarrollamos para la predicción de recaídas. Para ello, empleamos modelos de aprendizaje automático (Machine Learning, ML), siendo la regresión logística el que mostró los mejores resultados. Este modelo obtuvo una especificidad del 82.4% y una sensibilidad del 82.6%, demostrando una buena predicción de las recaídas. También analizamos la contribución de cada biomarcador al modelo para entender mejor su papel en la predicción de recaídas.

Tras desarrollar un modelo de predicción de recaídas, nos centramos en estudiar la resistencia a la quimioterapia. El 40 y el 60% de los pacientes incluidos en el estudio experimentaron resistencia al tratamiento o toxicidad, respectivamente. Para abordar las limitaciones del método convencional IC_{50} que mide la eficacia de distintos tratamientos, particularmente su dependencia del tiempo, desarrollamos un método novedoso basado en la tasa de crecimiento. Este método nos permitió comparar los niveles de resistencia de varias líneas celulares de CRC y evaluar la expresión de diferentes biomarcadores para identificar posibles candidatos asociados con la resistencia.

A través de este trabajo, identificamos a TROP2 como un potencial biomarcador para la resistencia. Observamos que la expresión de TROP2, tanto a nivel génico como proteico, mostró una fuerte correlación con la resistencia. Además, la exposición crónica a la quimioterapia llevó a un aumento en la expresión de TROP2. Además, exploramos los efectos del silenciamiento del gen TROP2 y evaluamos su influencia en la resistencia. Dado que la resistencia se evaluó mediante un ensayo de viabilidad celular, el impacto de TROP2 en la proliferación complicó el experimento. A pesar de los desafíos encontrados y la necesidad de estudios adicionales, los resultados obtenidos posicionan a TROP2 como un biomarcador candidato para la detección de resistencia.

En conclusión, esta tesis doctoral ha permitido desarrollar un modelo matemático robusto para la predicción temprana de recaídas en pacientes con CRC. También introdujimos un nuevo método basado en la tasa de crecimiento para estudiar la resistencia a la quimioterapia e identificamos a TROP2 como un biomarcador candidato prometedor para la detección de resistencia. Estos hallazgos impulsan el avance de la medicina de precisión y mejoran la calidad de vida de los pacientes con cáncer colorrectal, al permitir intervenciones más específicas y efectivas, optimizar las tasas de respuesta, reducir tratamientos innecesarios y minimizar los efectos secundarios.

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

1L/2L/3L	First, second and third-line of treatment, respectively
5-FU	5-Fluorouracil
ABCG2	ATP-Binding Cassette Subfamily G Member 2
AEMPS	National Agency for Medicines and Medical Devices
AI	Artificial Intelligence
AKT	Protein Kinase B
ALCAM	Activated Leukocyte Cell Adhesion Molecule
ANOVA	Analysis of Variance
APC	Allophycocyanin
ATCC	American Type Culture Collection
ATP	Adenosine Triphosphate
AUC	Area Under the Curve
β-ACTIN	Beta-actin
B.C.	Before Christ
BCRP1	Breast Cancer Resistance Protein 1
BMI	Body Mass Index
BOM	Basal Organoid Medium
BSA	Bovine Serum Albumin
CA	Classification accuracy
CA19.9	Cancer Antigen 19.9
CAM	Chorioallantoic membrane
CAR	Chimeric Antigen Receptor
CD	Cluster of Differentiation
cDNA	Complementary DNA
CDX	CTC Derived Xenograft
CEA	Carcinoembryonic Antigen

CO_2	Carbon Dioxide
COM	Complete Organoid Medium
CpG	Cytosine-phosphate-Guanine
CRC	Colorectal Cancer
CRH	Ciudad Real Hospital
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSC	Cancer Stem Cell
CT	Computed Tomography
C_T	Cycle threshold
CTC	Circulating Tumor Cells
ctDNA	Circulating Tumor DNA
CTLA	Cytotoxic T-Lymphocyte Antigen
DCLK1	Doublecortin Like Kinase 1
DHC	Daraca Health Center
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
dNTPs	Deoxyribonucleotide Triphosphates
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial-Mesenchymal Transition
EpCAM	Epithelial Cell Adhesion Molecule
Eq	Equation
ERK	Extracellular Signal-Regulated Kinase
ESMO	European Society for Medical Oncology
F1	F1 Score
FAK	Focal Adhesion Kinase
FBS	Fetal Bovine Serum
FDA	Food and Drug Administration
FIT	Fecal Immunochemical Test
FITC	Fluorescein Isothiocyanate
FOLFOX	Fluorouracil and Oxaliplatin

FSC	Forward Scatter
FW	Forward Primer
H_2O	Water
HER2	Human Epidermal Growth Factor Receptor 2
IC_{50}	Half-maximal Inhibitory Concentration
IC_{r0}	Inhibition Concentration in which growth rate is 0
IC_{rmed}	Inhibition Concentration in which growth rate is half the control
ID	Identification
IGF	Insulin-like Growth Factor
JDH	Jimenez Díaz Hospital
kNN	k-Nearest Neighbors
KRAS	Kirsten Rat Sarcoma Viral Oncogene
LD	Localized disease
LGR5	Leucine-Rich Repeat-Containing G-Protein Coupled Receptor 5
MACS	Magnetic-Activated Cell Sorting
MAPK	Mitogen-Activated Protein Kinase
ML	Machine Learning
MRI	Magnetic Resonance Imaging
mRNA	Messenger RNA
MTT	Methylthiazolyldiphenyl-tetrazolium Bromide
NPV	Negative Predictive Value
NSG	NOD-SCID Gamma (Mouse model)
PBMCs	Peripheral Blood Mononuclear Cells
PBS	Phosphate-Buffered Saline
PDOs	Patient-Derived Organoids
PDX	Patient-Derived Xenograft
PE	Phycoerythrin
PET	Positron Emission Tomography
PI3K	Phosphoinositide 3-Kinase
PPV	Positive Predictive Value
Prec	Precision
PROM1	Prominin-1

RT-qPCR	Reverse Transcription Quantitative Polymerase Chain Reaction
RAGE	Receptor for Advanced Glycation Endproducts
RCH	Ramon y Cajal Hospital
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
RT	Room Temperature
RV	Reverse Primer (in PCR)
SE	Standard Error
siRNA	Small Interfering RNA
SP	Side Population
Spec	Specificity
SSC	Side Scatter
SVM	Support Vector Machine
Sx	Surgery
TACE	Tumor Necrosis Factor Alpha Converting Enzyme
TACSTD2	Tumor-Associated Calcium Signal Transducer 2
Tm	Melting Temperature
TME	Tumor Microenvironment
TNM	Tumor-Node-Metastasis
TROP2	Tumor-Associated Calcium Signal Transducer 2
TUH	Toledo University Hospital
VEGF	Vascular Endothelial Growth Factor
WHO	World Health Organization

1. Introduction

Cancer is a complex disease characterized by the uncontrolled growth and spread of malignant cells. Affecting millions of people worldwide, cancer remains one of the leading causes of mortality despite significant advancements in diagnosis and treatment. While progress in early detection and effective management of primary tumors has considerably improved remission rates, the challenge often emerges in subsequent years due to the risk of relapse. Apart from the commonly known tumoral cells, characterized by their high division rates, a less-known population called cancer stem cells (CSCs) are also present in many tumors, such as glioblastoma, breast and colon cancer. They are characterized by a low division rate, self-renewal and resistance, and are thought to be responsible for relapses. A more detailed explanation of CSCs can be found in section 1.6.

Many patients experience tumor recurrence, either locally or as distant metastasis, with the latter typically being more aggressive, resistant to conventional therapies and associated with poorer prognoses. This highlights the need to deepen our understanding of the mechanisms underlying resistance and to develop a method for relapse prediction based on biomarkers' expression, which are the focus of this PhD thesis. We centered our study on colorectal cancer (CRC) due to its high recurrence and mortality rates, which not only underscored its clinical relevance but also ensured the project's feasibility by enabling the collection of a substantial number of samples within the limited timeframe of a doctoral thesis.

This section aims to explain both the biological and medical aspects, as well as the mathematical approach, in sufficient detail to ensure clear understanding for professionals from different fields, as it forms part of a multidisciplinary PhD thesis. In this section we will first provide a general introduction to the disease, covering its origins, classification, epidemiology, risk factors, prevention, diagnosis and treatment. Subsequently, we will focus on CSCs as key contributors to relapses, highlighting several biomarkers reported in the literature for identifying this population. Lastly, we will explore the use of Machine Learning (ML) techniques, which were used to develop a mathematical model for predicting relapse.

1.1. Origin and main discoveries in cancer disease

A representation of the main discoveries and advances of cancer is shown in Figure I1. The first evidence of cancer in humans' dates to the ancient Egyptian and Greek civilizations, although traces of bone tumors have been found in prehistorical animals before [1]. Cancer was described in the Edwin Smith [2] and Ebers papyrus [3] but was not called as such; it was in 400 B.C. when Hippocrates first coined this term, derived from the Greek word karkínos.

Cancer was previously not considered to have physical or biological causes. Hippocrates was the first to describe it in a scientific manner. He considered this disease an imbalance of the four main body humors (blood, phlegm, yellow and black bile) [4]. From the Egyptian period until the end of the nineteenth century, the treatment was predominantly surgery. In relation with that, Galen, a Roman and Greek surgeon, was the one who indicated the first surgical strategies for the treatment of tumors in 130-200 B.C [5] and Seishu Hanoaka, a Japanese surgeon, performed a mastectomy as the first surgical procedure under general anaesthesia in 1804 [6]. Finally, William Halsted published the "Radical Mastectomy" procedure in 1894, leading to its practice worldwide [7].

Regarding scientific advances in the disease, the early development of microscopes in 1600-1620 and the later publication of "Micrographia" by Robert Hooke in 1665 in which he first recorded observations of cells [8], allowed Johannes Müller to demonstrate in 1838 that cancer is made up of cells, although he believed that they come from abnormal cells (cells that already have irregular characteristics). It was in 1858 when Rudolf Virchow identified that cancer cells come from normal cells that undergo transformation and become cancerous due to mutations or external factors. Finally, another notable historical event was in 1860 when Karl Thiersch showed that cancer metastasizes through the spread of malignant cells [9].

The first great advance in cancer treatment was the development of radiotherapy thanks to the discovery of X-rays by Wilhelm Conrad Röntgen in 1895 and radium by Marie Curie and Pierre Curie in 1898 [10,11]. These latter two suggested using X-rays to treat tumors, which have been applied since 1920. Moreover, the development of computerized tomography (CT) scanning by Godfrey Hounsfield in 1971 improved imaging for both surgery and radiation treatment [12].

The second great advance in cancer treatment was the development of chemotherapy. Starting with the accidental discovery of the first DNA alkylating agent, a nitrogen mustard derived from iprite. First used for military purposes during the Second World War, the pharmacological effect of nitrogen mustards on organisms affected by certain tumors was later described in 1946 by Louis Goodman and Alfred Gilman [13]. This led to the development of Mechlorethamine, the first chemotherapy treatment that was approved by the U.S Food and Drug Administration (FDA) in 1949. Mechlorethamine directly modifies DNA, leading to cell cycle arrest and programmed cell death (apoptosis). From this point on, several chemotherapy treatments were developed [14]. Chemotherapy is nowadays a very widespread technique worldwide and is usually administered as a combination of multiple drugs, including mitotic inhibitors, topoisomerase inhibitors and antimetabolites. Vincent DeVita first described the efficiency of combination treatments in 1965 [15].

At the beginning of the ‘80s, new discoveries in the field of immunology, cell and molecular biology allowed researchers to investigate the molecular mechanisms responsible for the neoplastic transformation of cells, which triggers uncontrolled growth. These advancements facilitated the identification of new molecular targets [16–19], leading to the development of targeted therapy, including immunotherapy and monoclonal antibodies (laboratory-produced antibodies targeting a distinct region of an antigen, called epitope).

The first monoclonal antibody, Rituximab (which specifically recognizes a non-glycosylated transmembrane phosphoprotein antigen CD20), was approved by the FDA in 1997. Nevertheless, Trastuzumab, which selectively binds to the human epidermal growth factor receptor type 2 (HER2), was the first antibody tested in clinical trial (FDA approval in 1998). The introduction of both antibodies represented landmark events in the revolution of anti-tumor treatments [20,21]. New monoclonal antibodies directed toward tumor antigens or T-lymphocyte receptors that downregulate the immune response have also been developed. The first approved immune checkpoint inhibitor was Ipilinumab in 2011, which specially binds to CTLA-4 protein [22]. In 2014, Nivolumab and Pembrolizumab, which detect the PD-1 molecule, were introduced to the market with previous studies demonstrating their efficacy [23]. These two checkpoint-targeted antibodies have been a major immunotherapeutic breakthrough against tumor progression, as recognized by the synchronous Nobel Prize in Medicine and Physiology awarded in 2018 to James P. Allison (anti-CTLA-4) and Tasuku Honjo (anti-PD1) for their discoveries of cancer therapy through regulation of the immune system.

In the 2020s, many clinical trials are currently underway to develop new approaches for cancer treatment. In particular, the novel technique CAR-T cell therapy (Chimeric-Antigen Receptor) involves the integration of chimeric receptors (laboratory-designed molecules that combine elements from different proteins to form an artificial receptor with specific functions) into T cells, enabling them to target antigens that are specifically expressed on tumor cells. This technique implies the removal of T-cells from the patient and the *in vitro* genetic modification for the addition of the chimeric receptor. These cells are then reinfused into the patients where they selectively interact with cancer cells [24,25].

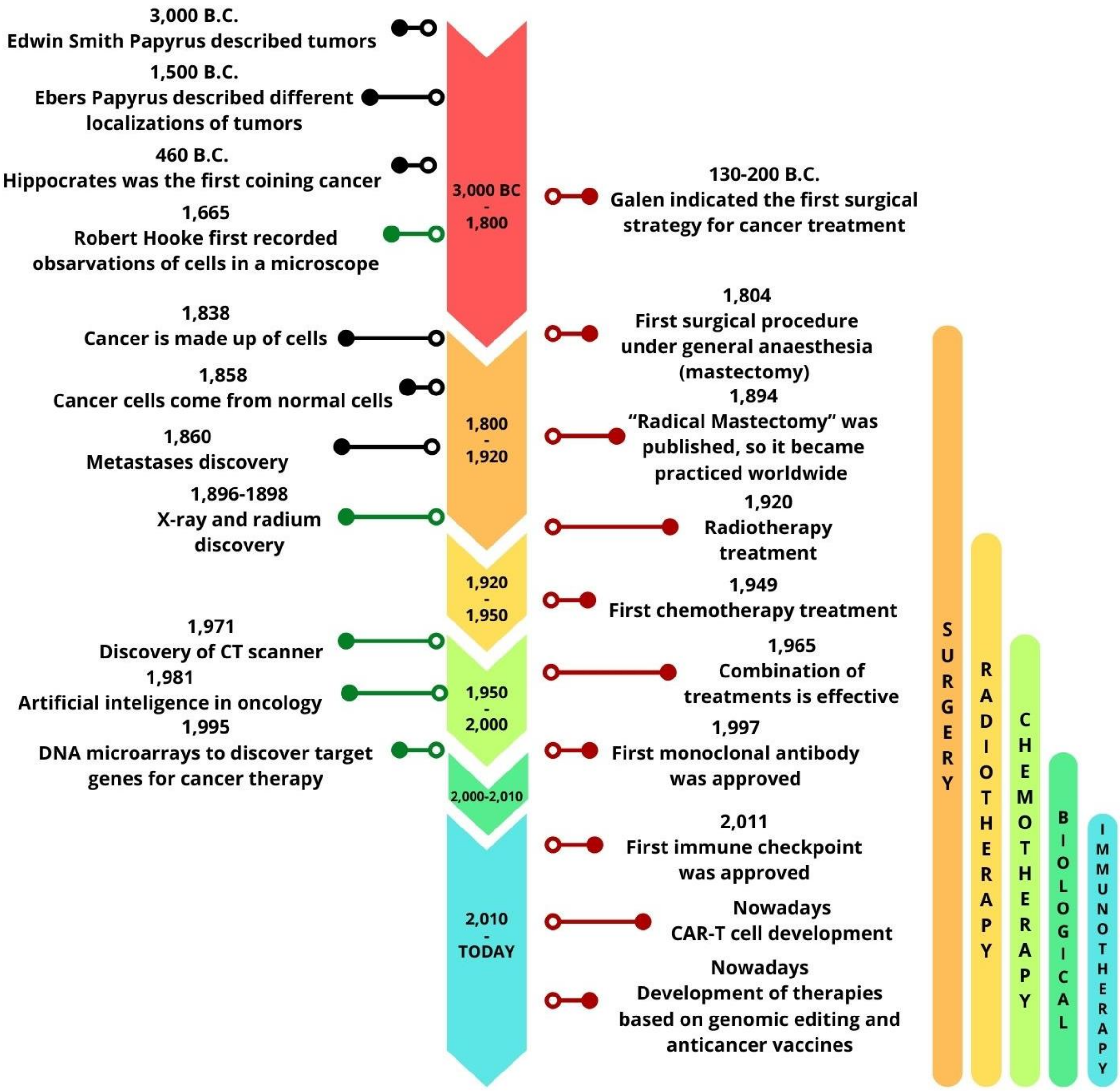

Figure I1. Schematic timeline that includes the most important milestones of oncology (3000 BC – today). The main scientific developments and discoveries in cancer are shown on the left (novel techniques are labelled in green). The advances in cancer therapy are shown on the right side (labelled in red). Original figure created by the author with data obtained from Falzone et al [26].

Finally, many researchers are trying to develop new therapies based on genomic editing using CRISPR/Cas9 technology to correct genetic aberrations, although these types of treatments are still in clinical trials with no approved therapies to date [27]. Some researchers are also studying the possibility of developing anticancer vaccines designed according to the individual characteristics of tumors. Although there are some approved vaccines, production difficulties make this approach particularly expensive and not suitable for all patients [28].

Advances in bioinformatics have also been important for scientific research and treatment of cancer. Although first mentioned in 1965, artificial intelligence (AI) was not applied to oncology until the 1980s, with the arrival of ONCOCIN. Developed by researchers at the University of Stanford, this AI-based system aimed to help oncologists make decisions about patients' cancer treatment [29]. Since then, AI has been used for clinical decision making, imaging analysis for early detection, predicting adverse drug reactions and measuring efficacy of treatments among patients [30–32].

1.2. Classification of cancer

Classification of cancer is crucial for pathologists, as it determines the severity and progression of a patient's cancer, influences their prognosis, and guides the selection of the most effective treatment options. Different classification systems have been established (Figure I2) based on varied features, leading from broader to more detailed separation. Some commonly used classifications are:

- By organ or tissue of origin.

  Cancer is broadly classified according to the location in the body where the cancer first developed, such as breast cancer, prostate cancer, colon cancer, lung cancer, etc. From a histological point of view, there are multiple different cancers that are grouped into six major categories: carcinoma (originating from epithelial cells), sarcomas (from supportive or connective tissue), myeloma (from plasma cells), leukemia (from leukocytes), lymphoma (from lymphoid tissue) and there is a group with mixed histological features in origin, called mixed type. This classification is described in the International Classification of Diseases for Oncology, Third Edition (ICD-O-3) [33].

- By grade.

  The World Health Organization (WHO) grades tumors depending on the cytological features (mostly focused on the differentiation of the cells) and morphological-structural observations (mitotic count and necrosis). There are four numerical grades, from grade 1 indicating a high level of differentiated and non-malignant cells (low grade and high survival rate) to a grade 4 in which almost all cells are undifferentiated and abnormal (high grade and low survival rate).

- By stage.

  Cancers are also classified according to their stage [34]. This usually refers to the size and spread of the tumor although there are several types of staging methods. The most widely used method divides cancer in terms of tumor size (T), the degree of regional spread or node involvement (N), and distant metastasis (M). This is called TNM staging. Each capital letter contains a letter or a number behind it, indicating the grade of this term. The following explains the meaning of the corresponding letters and numbers:

  - Primary tumor (T)
    - Tx: main tumor cannot be measured
    - T0: main tumor cannot be found
    - T1-T4: refers to the size of the main tumor, the higher the number, the larger the tumor
  - Regional lymph nodes (N)
    - Nx: the impact of the tumor on nearby lymph nodes is unknown
    - N0: no cancer in nearby lymph nodes
    - N1-3: the higher the number, the more lymph nodes that contain cancer
  - Distal metastasis (M)
    - Mx: metastasis cannot be confirmed or ruled out
    - M0: cancer has not spread to other parts of the body
    - M1: cancer has spread to other parts of the body

  The TNM method provides a detailed classification of tumors. However, cancer can also be categorized into five broader stages, with higher numbers indicating more advanced disease and a poorer prognosis. The following explains these five stages in detail:

  - Stage 0 indicates cancer *in situ* or limited to surface cells.
  - Stage I indicates cancer limited to the tissue of origin.

- Stage II indicates limited local spread.
- Stage III indicates extensive local and regional spread.
- Stage IV corresponds to advanced cancers with distant spread and metastasis.

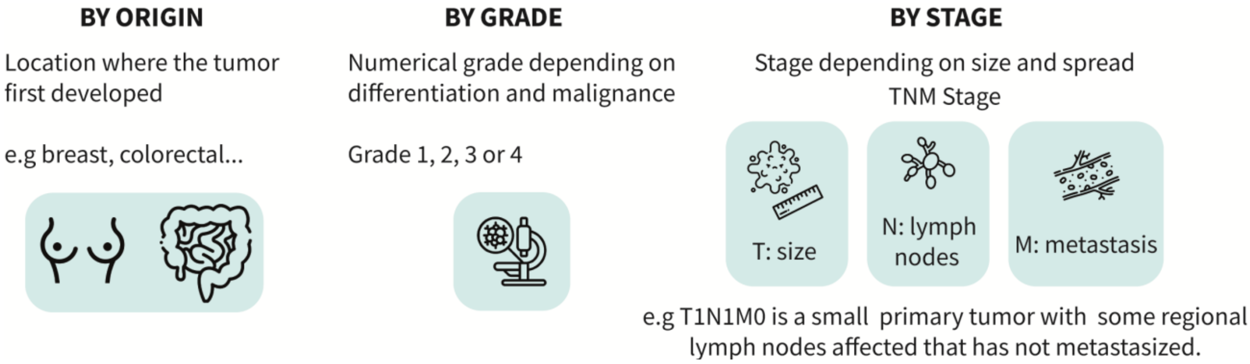

Figure I2. Classification of cancer. Cancer is broadly classified by origin (depending on the location of the primary tumor), grade (depending on the differentiation and malignance, there are four possible grades) or stage (depending on the size and spread, the TNM staging method is the most common one). Original figure created by the author using symbols designed by Freepick (https://www.freepik.es/).

1.3. Epidemiology

Cancer is a global health concern, with an incidence of nearly 20 million cases worldwide and almost 300,000 cases reported in Spain in 2022. Most cancer patients are aged over 50. Globally, lung cancer has the highest incidence, followed by breast cancer and CRC cancer. However, in Spain, this pattern differs slightly, with CRC having the highest incidence, followed by breast and prostate cancer. Consequently, CRC ranks third globally but holds the top position in Spain in terms of incidence. It is also the second leading cause of cancer-related deaths, preceded only by lung cancer [35]

CRC accounts for 9.6% and 14.1% of all cancer cases worldwide and in Spain, respectively. Although it affects both men and women almost equally, there is a slight variation, with around 55-60% of cases occurring in men and 40-45% in women and boasts a 5-year survival rate of 65%. The prognosis for patients depends on the stage of the disease, with a 92% 5-year survival rate for localized disease, 70% for cases where it has spread to regional lymph nodes, and 12% for cases where it has metastasized distantly. These survival prognoses are reflected in current prevalence data, with approximately 5.25 million people affected worldwide and 150,000 in Spain [35]. A summary of these metrics is shown in Figure I3.

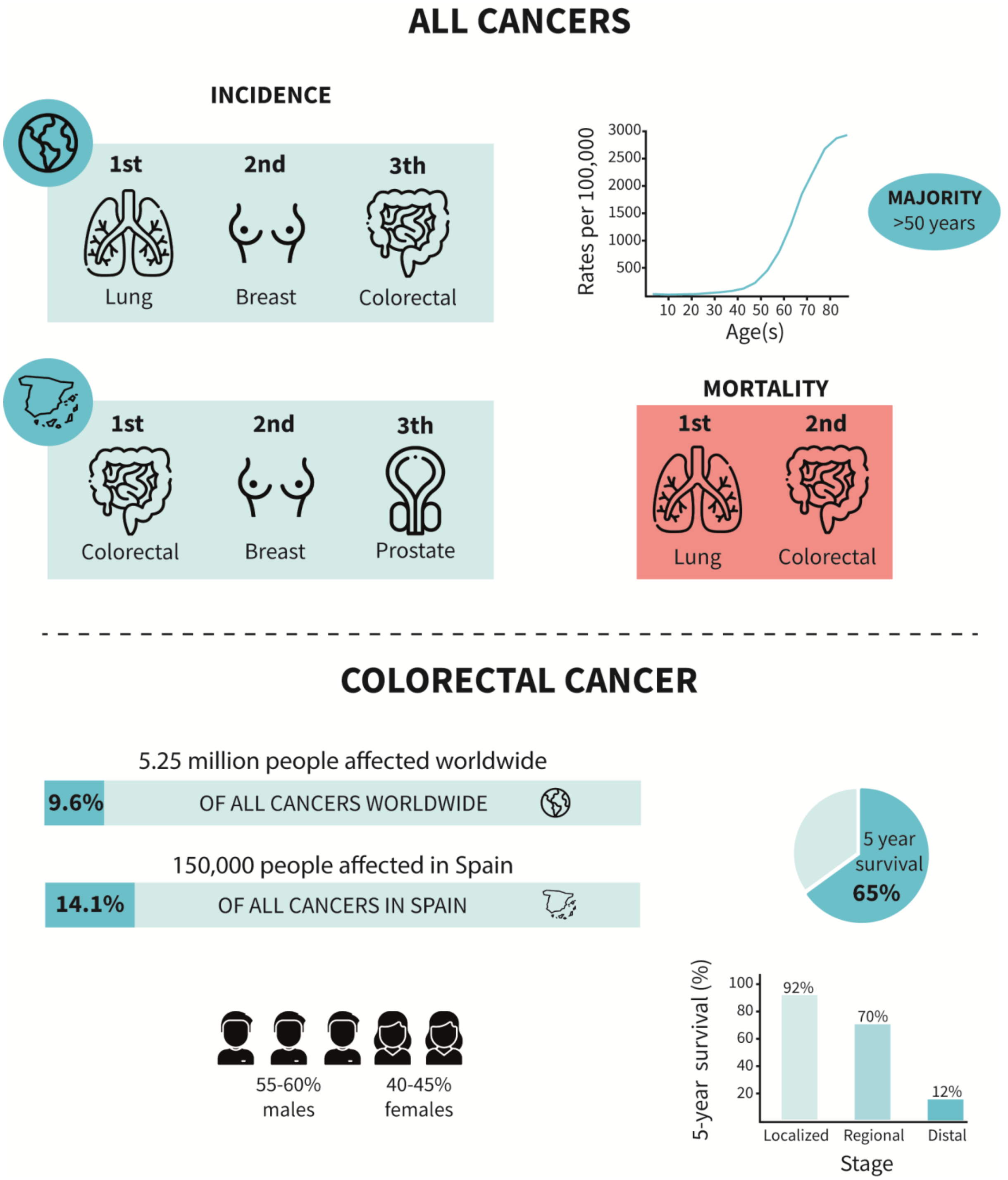

Figure I3. Epidemiology in cancer disease focused on CRC. The incidence, mortality and age of cancer patients is shown above. CRC incidence, sex and overall survival is shown below. Original figure created by the author with data taken from the Global Cancer Observatory [36] and using symbols designed by Freepick (https://www.freepik.es/).

1.4. Risk factors and prevention

Cancer occurs due to mutations of a patients' normal cells, leading to the appearance of tumor cells, with patients' genetics or the exposure to external agents being the most important causative agents. In this regard, genetic causes of cancer may refer to sporadic mutations in the DNA that happen as cells multiply or to mutations inherited from one or both parents. These mutations typically cause chromosomal alterations and translocations in genes that can lead to microsatellite instability (MSI), CpG island methylator phenotype (CIMP), and chromosomal instability (CIN). Mutations often cause dysregulation of various pathways that are responsible for cancer progression, including the PI3K/Akt, Wnt, TP53, and MAPK pathways, among others [37–39].

For their part, external agents can be classified in three main groups: physical carcinogens (such as ultraviolet and ionizing radiation), chemical carcinogens (such as asbestos, components of tobacco smoke, alcohol, food or drink contaminants, etc.) or biological carcinogens (such as infections from certain viruses, bacteria or parasites). In most cases, external agents induce mutations or affect the expression of oncogenes and tumor suppressor genes [40]. Viruses may also have other mechanisms to cause cancer such as encoding some viral proteins that are oncogenic or cause chronic inflammation [41,42]. In the case of chemical carcinogens, they can directly or indirectly react with DNA. Most of them do not react directly with DNA but undergo metabolic activation of reactive intermediates that react with DNA, leading to mutations that increase tumor risk [43].

Regarding CRC, numerous risk factors are known to predispose to it, including people with a family history of colon or rectum cancer, familiar polyposis syndrome or hereditary nonpolyposis, and chronic inflammatory diseases (ulcerative colitis and Crohn´s disease). Some environmental factors are also relevant, such as smoking, excess weight, diabetes, excessive fat and alcohol consumption, lack of regular physical activity and inappropriate diet [44–47]. While none of these factors alone have been directly linked to CRC, a healthier lifestyle overall has been associated with a lower incidence of CRC. Moreover, the WHO has determined that 23% of cancer cases attributable to excess body mass and 21.1% to alcohol consumption corresponds to CRC patients. Thus, CRC is one of the cancers that is most affected by diet [48].

The overall accumulation of risks along with the reduced effectiveness of DNA repair systems as a person gets older, causes the incidence of cancer to increase drastically with age [49]. Patients' genetics, environment and lifestyle are important factors in cancer development and risk can be largely reduced by controlling factors such as not smoking, maintaining a healthy body weight and a healthy diet, getting vaccinated, avoiding or reducing consumption of alcohol and controlling exposure to radiation and air pollution, etc. [50].

However, the development of preventive strategies is also essential. The fecal immunochemical test (FIT) is a voluntary biennial procedure in patients between the ages of 50 to 69. It is not a diagnosis technique but allows screening of the population depending on cancer risk. It should be noted that positive results can also be obtained due to some benign diseases such as hemorrhoids, necessitating colonoscopy assessment of patients whose FIT result was higher than a particular threshold (determined by each country) to diagnose CRC and remove possible polyps present in the colon. Although this technique is widely established, some investigations are being carried out to improve the test's specificity, as colonoscopy, according to some studies, could have been prevented in 38% of cases [51,52]. A summary of the main risk factors and prevention methods in cancer is shown in Figure I4.

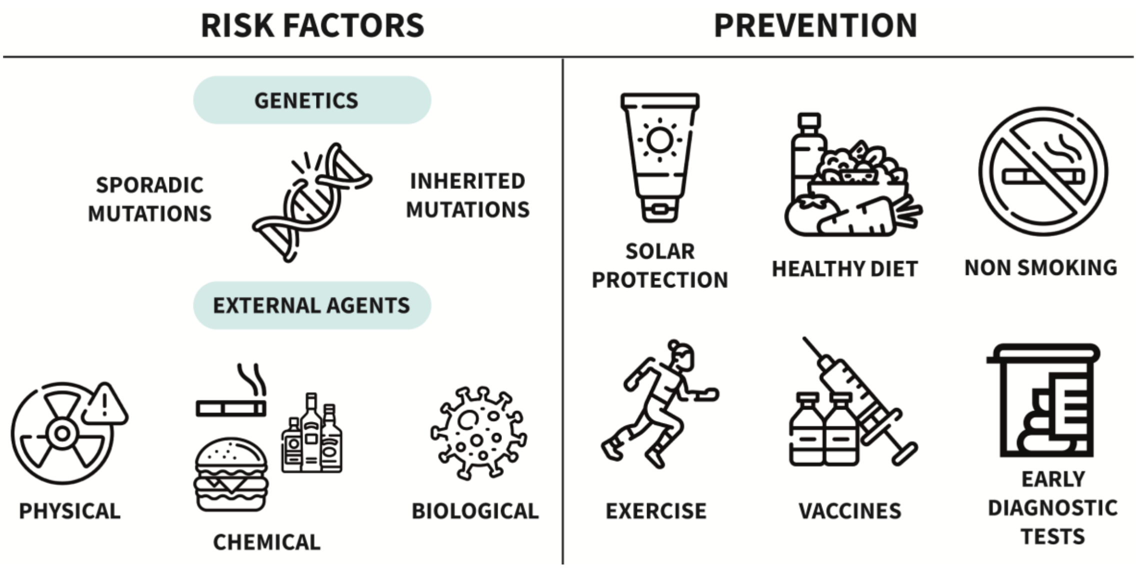

Figure I4. Risk factors and prevention in cancer. Risk factors include genetics (sporadic or inherited mutations), and external factors classified in physical (radiation), chemical (such as tobacco, food or alcohol) or biological (infections). Cancer can be prevented with a healthy life, protection from radiation and infections and early diagnosis tests. Original figure created by the author using symbols designed by Freepick (https://www.freepik.es/).

1.5. Diagnosis and treatment of cancer

Figure I5 presents a comprehensive diagram depicting the essential steps followed by CRC patients in relation to the diagnosis and treatment of their disease.

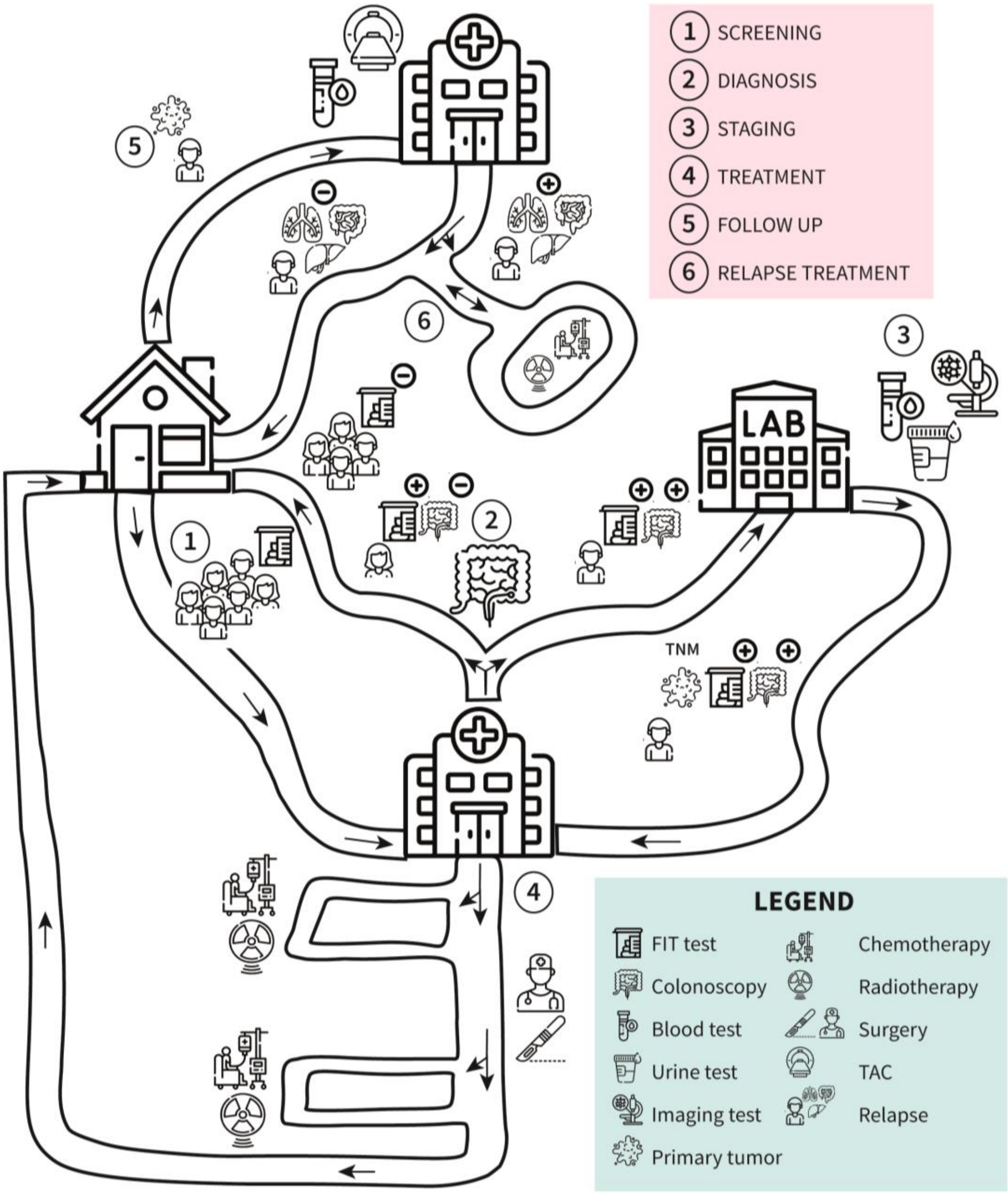

Figure I5. Diagram illustrating the key steps patients follow in the diagnosis and treatment of their disease. (1) FIT screening test (2) Colonoscopy test in cases of FIT positive (3) Blood and urine analysis in cases of FIT and colonoscopy positive tests (4) Treatment (5) Follow-up (6) Treatment in case of relapse. Two hospitals were represented in the figure to improve the flow of the diagram, although diagnosis, treatment and follow-up are usually carried out in the same center. Original figure created by the author using symbols designed by Freepick (https://www.freepik.es/).

Cancer is usually detected in preventive screenings, as an impact from another illness or because the patient suffers physical changes such as lumps, loss of weight, fatigue, bleeding, skin color changes, etc. In CRC, the FIT screening test is routinely performed on the general population between 50 and 69 years old (Figure I5 - 1) followed by a colonoscopy in the cases of presence of blood in the FIT sample (Figure I5 - 2). In the case of negative FIT and/or colonoscopy results, patients can wait until the next screening appointment. However, if FIT and colonoscopy are positive, some laboratory tests such as blood and urine tests and biopsy analysis are carried out to determine the stage of the tumor (Figure I5 - 3). The doctor will decide with these results which treatment is required for each patient (Figure I5 - 4). It usually consists of chemotherapy or radiotherapy treatment before surgery (neoadjuvant treatment). Depending on the stage of the tumor, the doctor could consider doing only surgery and could decide whether adjuvant (post-surgery) chemotherapy or radiotherapy is needed or not. Some other treatments may also apply, such as monoclonal antibodies. Every 3-6 months, a follow-up will be conducted to check for any signs of relapse (Figure I5 - 5). Blood and image tests such as PET are the most common techniques. Finally, if the results are negative, the patient is released until the next appointment but if a relapse is detected, a new treatment will be initiated (Figure I5 - 6).

1.5.1. Diagnosis of CRC

The most widely used techniques for early detection in CRC is FIT followed by colonoscopy. If preoperative endoscopy was incomplete due to tumor obstruction, visualization should be completed by computed tomography (CT) prior to surgery. Furthermore, CT/PET is usually performed for a complete pre-surgical staging and to identify occult metastatic disease. There are other techniques such as magnetic resonance imaging (MRI) or colonography that may also be used [53,54]. Whenever colonoscopy is not possible, colonography is used as a substitute. It is a non-invasive X-ray technique that allows for the detection of polyps and CRC tumors by providing a tridimensional view of the inner lining of the colon. MRI is a non-ionizing radiation technique that provides high-resolution images of soft tissues, making it particularly useful in CRC for visualizing internal layers and lymph nodes. However, its higher cost and longer scan time make it less frequently used than CT. Moreover, CT is generally more effective at detecting metastasis.

Currently, the development of omics technologies (genomics, transcriptomics, proteomics, etc.) has allowed scientists to identify biomarkers useful for diagnosis of CRC which can be used to assign a more efficient personalized treatment [55,56]. High serum levels of CEA and CA19.9 are used in CRC as biomarkers, but they are not exclusive. CEA is not specific, as it is expressed by other cancers such as ovarian tumors. CA19.9 is a specific marker of the digestive system tumors, including pancreatic cancer. Moreover, the sensitivity (the percentage of real positives that the model predicted as positive) is estimated to be 87 and 96%, respectively. However, the specificity (the percentage of negative samples that the model was able to predict as negative) is estimated to be 36 and 23%, respectively [57]. These sensibility and specificity data show that although elevated CEA or CA19.9 levels strongly indicate a high likelihood of cancer, low levels do not reliably rule it out, as many of these patients may still develop cancer.

1.5.2. Treatment of CRC

Standard conventional treatments for CRC are surgery, chemotherapy and radiotherapy. Monoclonal antibodies are also provided in some cases. There are defined clinical practice guidelines for diagnosis, treatment and follow-up of CRC, [58–60]. Patients diagnosed with CRC usually require multiple sequential treatments after detection of relapse, referred to as first line, second line, etc. A diagram of the treatment usually administered for patients in different stages at diagnosis is shown in Figure I6.

Surgery is the main curative treatment for patients with non-metastasized CRC. However, the outcome is strongly related to the quality of surgery and pre-operative staging but also treatment selection. Laparoscopic resection of CRC has been shown to be as safe as open surgery [61]. Once resected, the tumor can be used to judge the quality of the surgery. If the margin around it is free of cancer cells, the surgery is considered high quality [62]. Moreover, removal of lymph nodes is needed to provide adequate histopathological staging. The cure rate, defined as the percentage of patients who remain disease-free for life, is high in N0M0 cancer patients when treated with surgery, reaching 92% [35]. However, the clinical guidelines recommend considering neoadjuvant therapy in high-risk cases (poorly differentiated tumors, less than 12 lymph nodes resected, tumor invasion, obstruction or large tumors). Neoadjuvant treatment can reduce the tumor, and it might be recommended to optimize the chances for a successful resection [63].

The most frequent neoadjuvant treatment consists of a short-course radiotherapy or chemoradiotherapy with 5-Fluorouracil (5-FU) or capecitabine (prodrug of 5-FU). 5-FU is an antimetabolite with fluorine instead of hydrogen in the C-5 position of uracil. It inhibits the activity of thymidylate synthetase, affecting the pyrimidine synthesis and arresting the cell cycle in S phase [64].

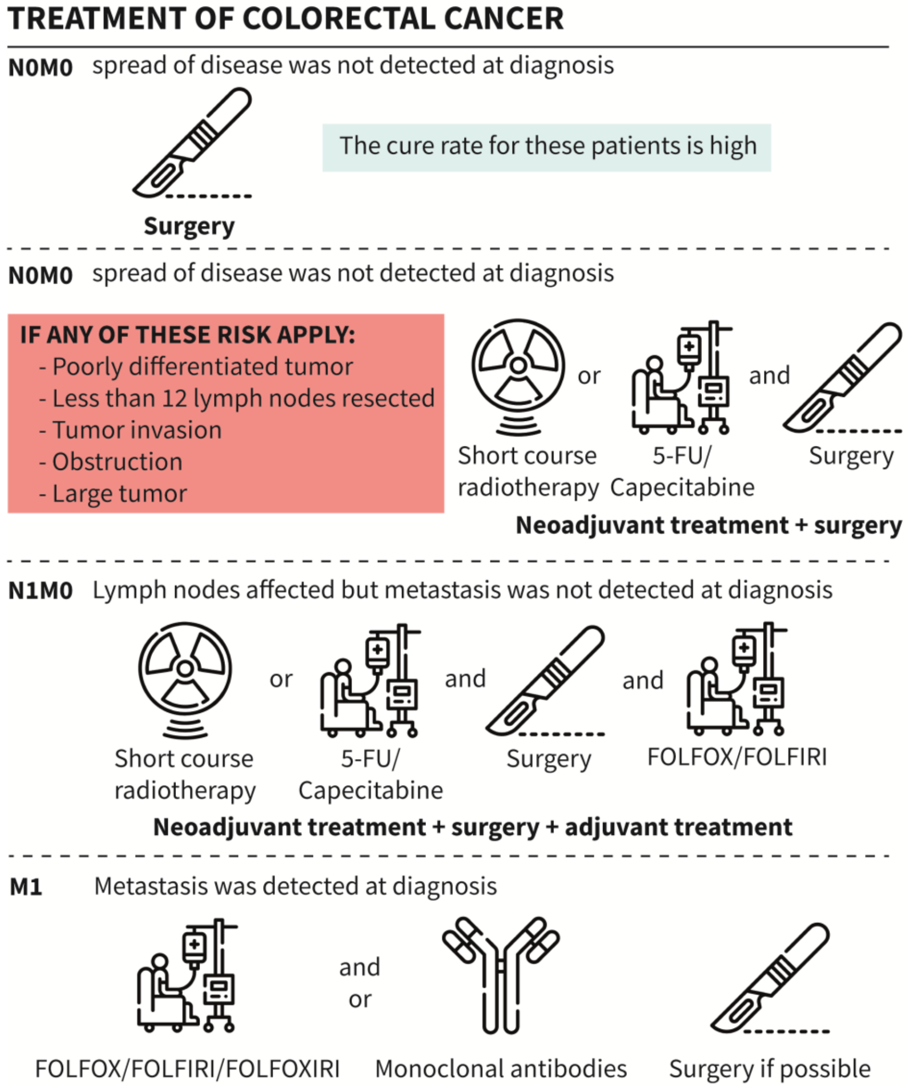

Figure I6. Diagram depicting the treatments usually administered in CRC patients. N0M0 patients (patients without nodules affected nor metastasis at diagnosis) receive surgery in most cases except for risk cases in which neoadjuvant treatment will also be considered. N1M0 patients (patients with nodules affected but without metastasis at diagnosis) will also receive adjuvant treatment in addition to neoadjuvant treatment and surgery. Finally, M1 patients (patients with metastasis at diagnosis) will receive a more active chemotherapy treatment usually combined with monoclonal antibodies. Surgery will be carried out if a resectable status is achieved. Original figure created by the author using symbols designed by Freepick (https://www.freepik.es/).

In cases in which any node is affected (any T, N1-2, M0), adjuvant treatment should also be considered. When considering metastatic patients, they are usually treated with the most active chemotherapy to reduce tumors. If a resectable state is achieved, surgery should be considered. The chemotherapy backbone for first-line treatment of spread disease is typically a combination of 5-FU, and either oxaliplatin (FOLFOX) or irinotecan (FOLFIRI). In the first case, 5-FU is usually replaced by capecitabine, as it is oral administrated, which is preferred by patients. Oxaliplatin is an alkylating agent that produces DNA damage through DNA crosslinking and irinotecan is a derivative of camptothecin which binds to topoisomerase I, preventing the relegation of the single-strand breaks in the DNA. A triplet (three chemotherapeutic agents), FOLFOXIRI, is usually administered in metastatic patients, which confers high response rate (approximately 60%) [65].

Finally, some targeted agents are used for metastatic CRC treatment. These include three major groups: monoclonal antibodies against epidermal growth factor (EGFR), such as cetuximab and panitumimab; monoclonal antibodies against vascular endothelial growth factor (VEGF-A), such as bevacizumab; and fusion proteins that target multiple proangiogenic growth factors (for example, aflibercept) or small molecule-based multikinase inhibitors (for example, regorafenib).

Overexpression of EGFR correlates with reduced survival and increased risk of metastasis. Cetuximab and Panitumimab are monoclonal antibodies that block EGFR, and hence EGF signaling, showing efficacy in chemotherapy-naïve patients as well as in patients whose tumors are refractory to chemotherapy, thus improving the overall response rate of tumors. A prerequisite for the efficacy of these agents is that the tumors do not harbor mutations in KRAS and NRAS [66,67].

The proangiogenic VEGF-A factor is produced by many tumor cells and promotes proliferation and migration of cells. Bevamizumab, a monoclonal antibody that block VEGF-A, has demonstrated efficacy in combination with chemotherapy in metastatic patients; it is usually combined with 5-FU or irinotecan, improving the overall survival of patients [68]. Aflibercept also targets angiogenesis and has been shown to improve overall survival when used in combination with FOLFIRI in the second line setting of treatment for metastatic disease [69].

Finally, regorafenib is an orally available multikinase inhibitor that has shown efficacy in patients who had previously been treated with all available therapies, becoming the standard in pre-treated patients [70].

1.5.2.1. Methods to measure the efficacy of anticancer treatment

Cell viability assays have become instrumental tools for assessing the efficacy of anticancer agents, particularly in the early stages of treatment studies. Diverse methodologies have emerged, driven by the need for more comprehensive understanding of cellular responses. Traditionally focused on survival or death, these studies have expanded to analyze specific aspects of cellular function. Advanced methods, including thiazolyl blue tetrazolium bromide (MTT) assays, fluorescence microscopy, real-time cell analysis (RTCA) and flow cytometry, offer detailed assessments, particularly relevant in studying tumor cells and therapeutic compound efficacy [71]. The MTT assay, a widely used technique, is a reliable and cost-efficient method that measures cellular metabolic activity. It relies on the reduction of MTT to formazan in the mitochondria, providing a colorimetric measure of cell viability by absorbance [72]. The fact that cell lines grow exponentially has been widely studied since 1961 [73–75] and continues to be used in recent studies [76,77]. In these experiments, absorbance is proportional to the number of cells and is usually normalized to the control cell line, so the evaluation of exponential growth in controls is essential to determine that there are no external factors that affect the culture. The percentage of viability is usually calculated as follows:

$$\text{Cell viability (\%)} = \frac{\text{Population sample}}{\text{Population control}} \text{ x } 100 = \frac{\text{Absorbance sample}}{\text{Absorbance control}} \text{ x } 100 \quad \text{(Eq.1)}$$

Among these assays, the discovery of parameters to analyze and compare the efficacy of treatments is essential. The determination of the half-maximal inhibitory concentration (IC_{50}) has gained prominence in this point. The IC_{50} value denotes the concentration of a compound at which 50% of cell viability is inhibited, serving as a parameter to assess the efficacy of potential therapeutic compounds [78]. Its use in cell viability assays offers a quantitative measure, enabling researchers to compare the efficacy of different compounds and make informed decisions in the development of cancer treatments [79,80]. Determining IC_{50} from dose-response curves is essential but challenging. Various methods, including nonlinear regression and statistical models, navigate curve complexities [78,81,82].

However, a major drawback of the IC_{50} index is its time-dependent nature. Since both the sample and control cell populations evolve over time at different growth rates, performing the same assay with a different end-point can result in a different IC_{50} value. In optimal laboratory conditions cell cultures produce colonies with exponentially growing populations, and a well-defined growth rate. When these conditions are changed, for example, to study chemotoxicity, some cells die and others reproduce at a slower rate; however, the change in colony population growth dynamics can be modeled as proportional to colony population itself (because it is a multiplicative phenomenon). The proportional constant is a new effective growth rate, i.e., the growth rate that can be measured for a population under given conditions. The colony population is an exponential function of time. If the number of individuals in the population increases, the growth rate is positive, but if the number of living individuals decreases, this corresponds to a negative growth rate (exponential with negative exponent) [83,84]. Thus, a direct way to evaluate the effect of a chemotoxic compound on cell culture is to study the change in cells' growth rate.

In this PhD thesis we propose a new method to measure the efficacy of chemotherapeutics based on the calculation of the effective growth rate for a range of drug concentrations. We also propose the use of two new parameters to evaluate the response of cell lines to different treatments that are not time dependent. One of them is the IC_{r0} index, a value that represents the drug concentration at which the effective growth rate is zero. The other, the IC_{rmed}, is the value that corresponds to the drug concentration that reduces the control population's growth rate by half. Advancements in growth rate analysis offer a precise alternative for IC_{50} determination without succumbing to curve intricacies. These advancements will be useful for a more accurate measurement of resistance to chemotherapy mainly caused by CSCs.

1.6. Cancer stem cells

According to the traditional model of carcinogenesis, a tumor may arise from any differentiated cell. This model is called the stochastic model and assumes that every cell can become tumorigenic via somatic mutations in genes that control proliferation and thus originate a tumor [85]. More recently, our understanding of tumorigenesis has shifted to the CSC model. The CSC model suggests that only a subpopulation of cells within the tumor, the CSCs, are able to proliferate and drive tumorigenic growth [86,87]. Both models are represented in Figure I7.

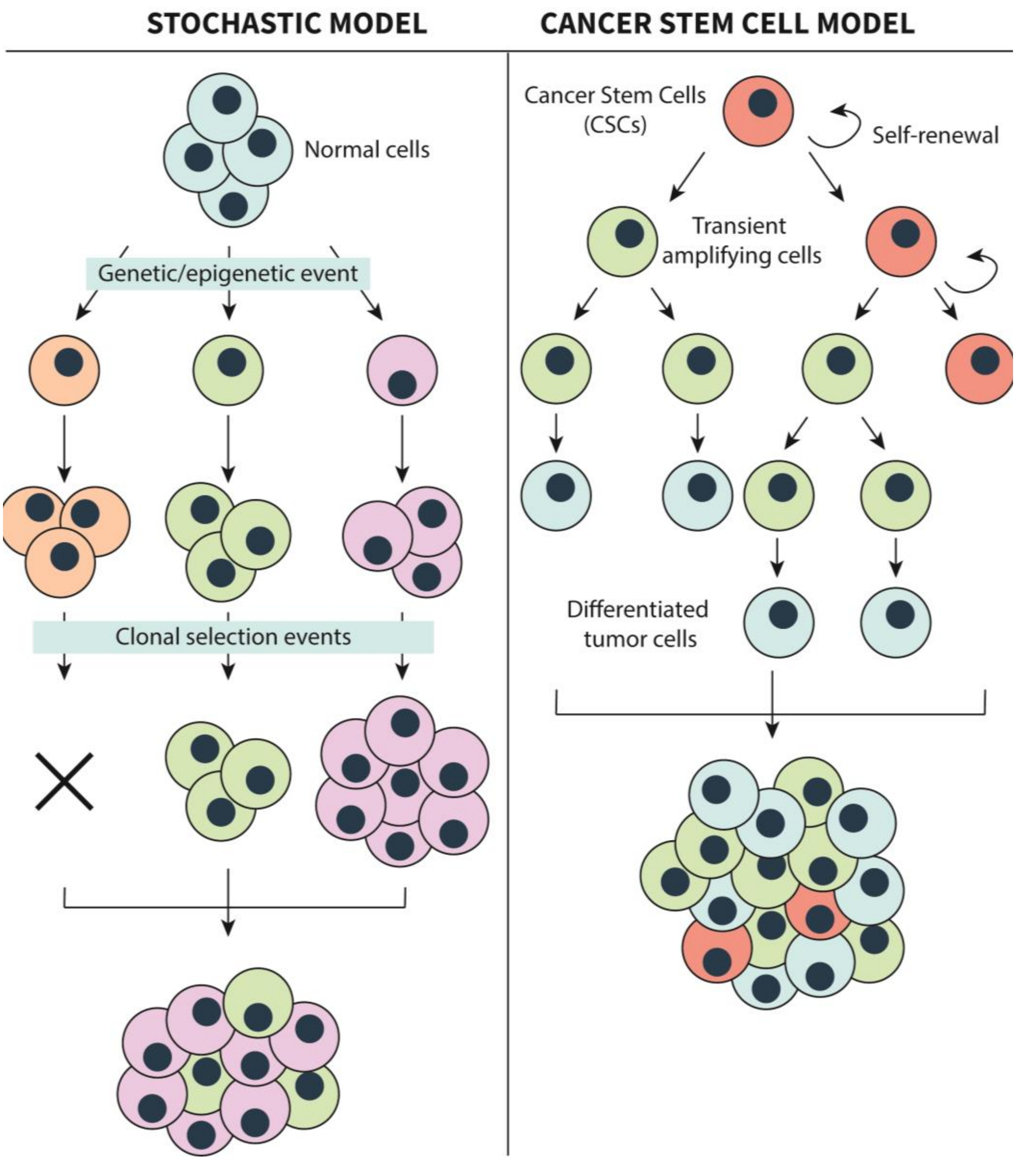

Figure I7. Models of carcinogenesis. The stochastic model assumes that tumor cells come from somatic mutations in any cell (left panel) whereas the CSC model suggests that a subpopulation can proliferate and promote tumorigenic growth (right panel). Image reproduced from [88].

CSCs were first identified in 1994 in the blood of leukemia patients as a small subpopulation of cells which could initiate leukemia in immune-deficient mice [89]. Although the origin of CSCs remains unclear, many studies have demonstrated that they display self-renewal, unlimited proliferative potential, multipotency, angiogenic and immune evasion features [86]. Moreover, CSCs are thought to be responsible for tumor initiation as well as local relapses or metastasis and resistance to therapy, thus playing an important role in tumor progression [90].

Metastasis is produced when CSCs detach from the tumor, undergo epithelial-mesenchymal transition (EMT) and invade into surrounding tissue. After entering the bloodstream and lymphatic circulation, the cells seed into a new organ site, commonly liver or lung in CRC [91].

In metastasis, the cell-cell junction architecture is compromised through downregulation of cadherins and other cell adhesion molecules, as well as from the secretion of enzymes that degrade the extracellular matrix (ECM) and the loss of apicobasal polarity due to poor cadherin-catenin anchorage to the cytoskeleton [92]. Mesenchymal characteristics are upregulated in EMT and controlled by pathways that also regulate stemness of CSCs, which refers to their ability to self-renew and differentiate into multiple cell types.

Despite the emergence of new targeted therapies and the use of therapeutic combinations, the fact that relapse remains high in cancer patients supports the model of CSCs. Most of the currently available therapies are directed against DNA integrity and the arrest of the proliferative cell cycle. So, they only target the high proliferating cells and not the CSCs, which will be able to re-initiate tumor growth and causing both relapse and metastasis (Figure I8) [87,93].

Resistance of CSCs to conventional treatment can be explained due to quiescence (a dormant or slow-cycling state in which cells remain metabolically inactive) and self-renewal processes, the expression of detoxifying enzymes, drug transporters and DNA repair mechanisms [94]. Although several treatments have demonstrated efficacy against CSCs in preclinical studies [95], they have not had success in clinical trial phases to date. For this reason, no CSC-targeted treatment has been approved by the FDA nor the European Medicines Agency (EMA). One example is Verastem (NCT05074810 and NCT06072781), an inhibitor of the PI3K/mTOR pathway that did not meet the efficacy standards required to progress to later phases. This also happened to Tarextumab (NCT01859741 and NCT01647828), which targets the receptors Notch2 and Notch3 or Rosmantuzumab (NCT02482441), which targets the protein R-spondine-3, a key regulator of the Wnt pathway. Further investigations are needed to make progress in this area, with multiple groups working towards this goal. For example, the Vall D´Hebron Hospital has recently administered the first treatment of AOP208 in a clinical study in phase I called SERONCO-1, which blocks the receptor of serotonin 1B, a surface protein of CSCs [96].

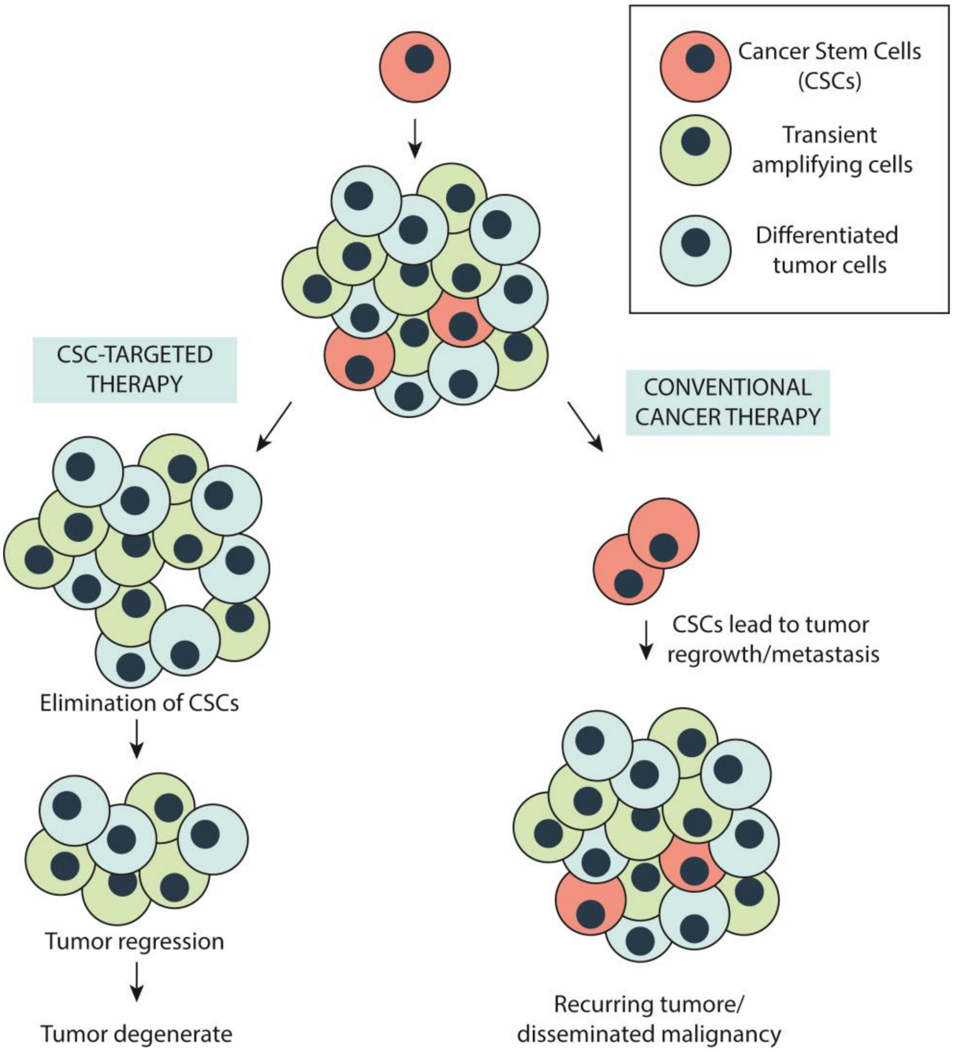

Figure I8. Conventional versus CSC-targeted therapy. Conventional treatment eliminates tumoral cells while CSCs remain resistant, originating the tumor again. However, CSC-targeted therapy eliminates CSCs and tumor degenerates. Image reproduced from [88].

1.6.1. Biomarkers for CSCs

Both relapse and resistance to conventional treatments make further study of CSCs essential. Surface markers play a crucial role in detecting CSCs, which could help early diagnosis of relapse and predict resistance to treatments. Many surface markers have been identified in CSCs in the last years. CD133, CD44, EpCAM and ALDH are commonly expressed in CSCs in most cancers [16]. Beyond these widely recognized markers, CSCs also display other markers such as LGR5, CD166, DCAMKL1, TROP2, etc [16–19]. The main function of some relevant CSCs biomarkers is shown in Figure I9 and further information for each one is included below.

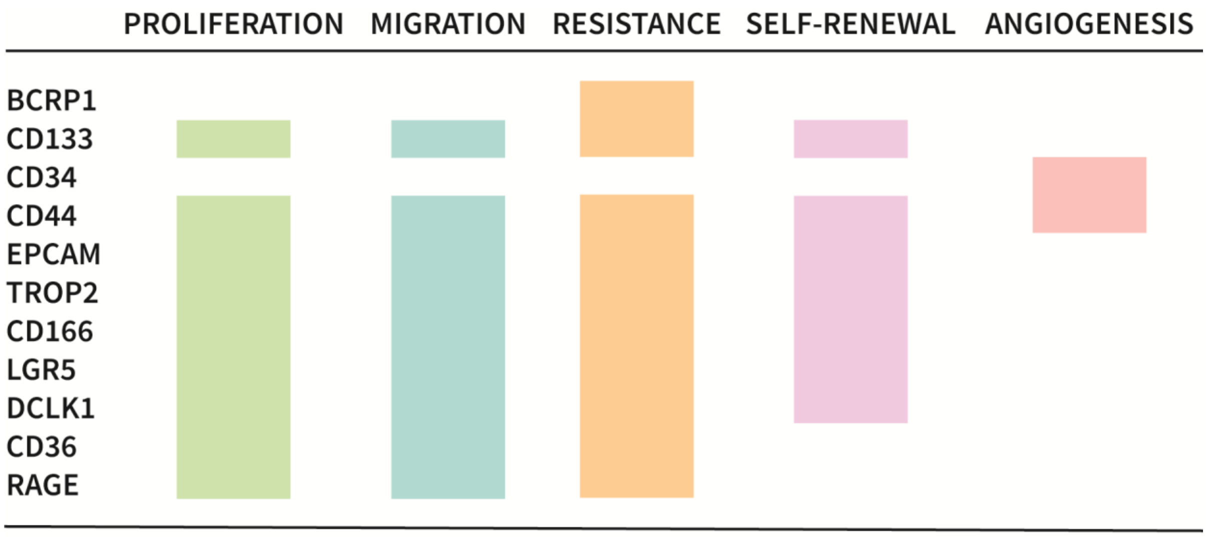

Figure I9. Implication of different biomarkers in proliferation, migration, resistance and angiogenesis. Colored wells indicate implication of the biomarker in the corresponding pathway, blank spaces indicate the biomarker is not involved in the pathway. Original figure created by the author.

- **BCRP1**

Breast cancer resistance protein isoform 1 (BCRP1, gene symbol *ABCG2*) is an ATP-binding cassette (ABC) efflux transporter. Its name derives from its initial cloning from a multidrug-resistant breast cancer cell line. The first hematopoietic stem cell population, called side population (SP), was characterized by the capacity to expel Hoechst from the cell. The experiments validated it by blocking the transporter with Verapamil, an inhibitor of BCRP1. When Verapamil was added, the side population decreased [97]. The SP population was also detected in tumor cell lines. So, BCRP1 is responsible for the efflux of drugs in CSCs [98].

Moreover, it is reported as a CSC marker [99] and has demonstrated to confer resistance to chemotherapeutic agents such as mitoxantrone, topotecan, and SN-38. Several drugs currently used for CRC treatment, such as irinotecan and 5-FU, are also substrates of BCRP1 [100,101]. Furthermore, the expression of BCRP1 in the membrane is associated with both lymph node and distant metastasis in CRC [102]. Together, these data suggest that BCRP1 may serve as a promising biomarker for CSCs detection, as Giampieri et al showed in a study where a CSC gene profile including *ABCG2* was found to predict relapse [103].

- **CD133**

Prominin 1 (CD133, gene symbol *PROM1*) is a membrane glycoprotein, thus may play a role in cell structure, cell recognition and cell signaling. CD133 is one of the most well-known stem cell biomarkers and has been postulated to identify the CSC population in numerous solid tumor types including colon cancer. Some studies have found that a small population of CD133+ cells was able to generate a tumor when transplanted in immunodeficient mice, whereas CD133- population could not, revealing CD133's role in cancer initiation [104,105]. Moreover, knockdown of CD133 impairs cell migration, suggesting a role in CRC metastasis [106]. Elevated CD133 expression also enhances resistance to chemotherapy and radiotherapy [107,108]. CD133 can also modulate autophagy and apoptosis, so it is associated with enhanced self-renewal, migration, invasion and survival in CRC [109].

Despite all these discoveries, the usage of CD133 as a colon CSC marker has continued to be controversial, as some studies have found other markers to be much more robust identifying CSCs [16] and some others determine a complex regulation of CD133, with the AC133 epitope being responsible for CSC identification and an adverse prognostic factor in CRC patients [110]. Nevertheless, numerous recent clinical studies have indicated that CD133 exhibits significant prognostic value for predicting patient survival in CRC [111,112] and some others are developing CD133 targeting therapies [113–115].

- **CD34**

CD34 (gene symbol *CD34*) is a transmembrane glycoprotein which is mainly implied in angiogenesis. Angiogenesis is the formation of new blood vessels from pre-existing ones, enabling tumor growth and spread. CD34 has been identified in different types of cancer including CRC, although its correlation with overall survival remains controversial [116,117]. Regarding CSCs, the first evidence was demonstrated in acute myeloid leukemia. Interestingly, they were identified as a CD34+CD38- subpopulation [118]. However, its expression in many cell types makes its study necessary in combination with other biomarkers.

Moreover, CD34+ hematopoietic progenitor cells have been used for transplantations where hematopoietic reconstitution was needed. It has been used to treat heart-related diseases, cancer and autoimmune diseases [119]. Furthermore, a novel CD34/CD3 bi-specific T-cell engager (BTE), a small molecule that consists of two single chain variable fragments that redirect T cells against tumors, has been developed for leukemia treatment [120].

- **CD44 (and variant CD44v6)**

CD44 (gene symbol *CD44*) is a type I transmembrane glycoprotein whose overexpression has been related to poor prognosis for many cancer types including CRC [121]. Its implication in cell proliferation, migration, angiogenesis, EMT and drug resistance has been described [122–125].

Although CD44 has been identified as a CSC biomarker both by itself and in combination with other biomarkers [126,127], its expression is not only limited to this cell population. CD44 is also expressed in many cell types including all hematopoietic cells, epithelial cells, fibroblasts and endothelial cells [128]. The full-length CD44 consists of 20 exons, but the 10 central ones undergo alternative splicing leading to different variants (CD44v). CD44v6 contains the variant exon 11 and is more specifically expressed in CSCs. It plays a role in cancer progression as it is involved in EMT, cell migration and invasion, and accounts for the metastatic phenotype of tumors. It has also been correlated with poor prognosis in many cancers including CRC [129–132]. All these results have drawn much attention, and some CD44v6-targeting strategies are being developed [133–135].

- **EPCAM**

Epithelial cell adhesion molecule (EPCAM, gene symbol *EPCAM*) is a type I transmembrane glycoprotein, member of the TACSTD gene family. Although its role in cell migration, adhesion and invasiveness has been described [136,137], other authors state that the loss, rather than overexpression, of membranous EpCAM is linked to tumor progression [138]. Further study of the cleavage of this protein could aid in understanding this controversy. After cleavage, the N-terminal remains bound to the protein by a disulphide bond preventing dimerization and the interaction with claudin that is implied in ERK signalling [139]. Furthermore, an intracellular fragment could also be cleaved, which is translocated from the cytoplasm into the nucleus and form complexes that stimulate the expression of genes involved in cell proliferation [140].

Regarding CSCs, EPCAM is expressed in this population but also in differentiated cancer cells of epithelial origin. That is why a combination of biomarkers will be preferred to detect CSCs [141]. Nevertheless, some EPCAM-targeting antibodies and CAR-T cells are being studied [142,143]. Finally, a diagnostic technology called CellSearch ® was developed by Veridex. Briefly, it detects the CSCs by detecting EPCAM positive and CD45 negative cells. More details of this technology are described in section 1.7.

- **TROP2**

Trophoblast cell-surface antigen 2 (TROP2, gene symbol *TACSTD2*) is a cell surface glycoprotein which is a member of the TACSTD gene family. TROP2 and EPCAM (TACSTD1, member of the same family) share 49% identity and 67% similarity of the aminoacid sequence. Moreover, the distribution of hydrophobic and hydrophilic residues is highly similar, and both dimerize. Thus, it is reasonable to think that they present similar functions. However, some structural differences have been found, such as TROP2 cleavage is possible in dimeric state while EPCAM dimer destabilization is required prior to cleavage [144].

Similar to EPCAM, the overexpression of TROP2 has been observed in many types of carcinomas including CRC [17]. However, other authors show that TROP2 is downregulated in some other carcinomas such as lung [145]. It has also been associated with cell proliferation, migration, drug resistance and EMT to varying degrees, depending on cancer type [146–151]. Finally, membrane-localized TROP2 has been associated with lower survival rates while intracellular TROP2 has been associated with higher survival rates in breast cancer patients [152]. That is why the cellular localization of TROP2 could be important to understand the role of TROP2 in tumorigenesis.

TROP2 undergoes intramembrane proteolysis, so the intracellular domain is released and accumulates in the nucleus in a complex with β-catenin, this allows for activated transcription of some genes that induces EMT [147,153]. Moreover, the extracellular domain is cleaved by tumor necrosis factor-α converting enzyme (TACE) [154]. Further studies on the mechanism that regulates this processing and its role in tumorigenesis are needed.

Notwithstanding the debate regarding the role of TROP2 in some cancer types and cell localizations, a TROP2-targeted antibody containing SN-38 (an active metabolite of irinotecan) has been developed. It has been FDA approved in 2016 for breast cancer treatment although other clinical trials (NCT02574455) are being developed [155,156]. Finally, several other TROP2-targeting immunotherapies are being tested such as nanoparticles conjugated with anti-TROP2 antibody [157] or TROP2 bispecific antibodies that target T cells to tumors [158].

- **CD166**

Activated leukocyte cell adhesion molecule (CD166, gene symbol *ALCAM*) is a protein that belongs to the immunoglobulin superfamily involved in cell-cell interactions. CD166 has been identified as a CSC marker in various solid tumors including CRC. Moreover, a combination of biomarkers including CD166 have been found to better detect CSC populations [141,159,160].

In CRC patients, there is controversy about the role of CD166 in tumor prognosis. Its high expression has been associated with short overall survival [161]. However, reduced expression on the cell surface, cytoplasm accumulation or shedding has been related to adverse outcome [138,162,163].

Although the role of CD166 in CRC prognosis remains unclear and further investigation is needed, a CD166 targeting CAR-T cell therapy has been proved in preclinical assays for CRC with promising results targeting CSCs [164]. Finally, there are some clinical trials (NCT04596150) that study the effect of a CD166 monoclonal antibody conjugated to a cytotoxic agent called ravtansine for the treatment of some tumoral types [165].

- **LGR5**

Leucine-rich repeat-containing G-protein coupled receptor (LGR5, gene symbol *LGR5*) is a seven-transmembrane protein family of G protein-coupled receptors (GPCRs). Thus, its main function is to transduce signals, in this case, involving the Wnt pathway [166]. Many studies have demonstrated its role in proliferation, tumor growth, migration, invasion and chemoresistance in CRC [167–170]. Moreover, LGR5+ populations have the capacity for colony formation, self-renewal, differentiation and tumorigenicity and present higher expression of stemness genes. All these together make LGR5 a CSC biomarker [171,172].

The plasticity of LGR5 in CRC has been studied, demonstrating that the majority of disseminating cells are LGR5- cells, which have the capacity to re-establish the LGR5+ population in the metastatic tumor requiring LGR5- cells for long-term metastatic growth [173–175]. Finally, some Spanish researchers have recently developed a LGR5-targeted antibody that carries 5-FU or oxaliplatin as a targeted therapy for CRC [176].

- **DCLK1**

Doublecortin-like kinase 1 (DCLK1, gene symbol *DCLK1*) is a protein kinase that is overexpressed in numerous cancers including CRC [19]. It is involved in cancer growth, EMT, metastasis, cancer cell self-renewal and therapy resistance [177–180]. Recent studies demonstrate that DCLK1 is able to distinguish between tumor and normal stem cells, being downregulated in the bulk tumor and upregulated in CSCs [181,182]. Thus, it is a good target for CSC-targeted therapy development. In relation to that, several kinase inhibitor, siRNA, monoclonal antibodies and CAR-T cells are being developed against DCLK1 [183–187].

- **CD36**

CD36 (gene symbol *CD36*) is a transmembrane glycoprotein, well known for its role in lipid homeostasis. Tumor cells demand lipid energy uptake from the tumor microenvironment (TME) despite fatty acid biosynthesis. Several authors have described that CD36 mediates the uptake of lipid metabolites secreted by cancer-associated fibroblasts (CAFs) in cancer cells [188,189]. Furthermore, CSCs have been proven to use the CD36 receptor to promote their maintenance, and this protein's expression has been negatively correlated to patient prognosis [190]. Moreover, CD36 expression has been related to tumor growth, cell proliferation, invasion and metastasis in CRC [191–193]. However, Fang et al have recently described the role of CD36 as a tumor suppressor [194]. Thus, more studies are needed to confirm its role in CRC tumorigenesis.

Due to its broad expression patterns in many cancer types, some CD36-targeting therapies have been developed, most of which are small molecule inhibitors or lipid mimetics [191]. VT1021 is the only CD36-targeting agent in trials for cancer therapy, which binds to CD36 and CD47 and induce tumor apoptosis as well as immune modulation in the TME [195].

- **RAGE**

Receptor for advanced glycation end products (RAGE, gene symbol *RAGE*) is a receptor from the immunoglobulin family involved in inflammation-related disorders. Some studies have determined that RAGE is overexpressed in some cancers including CRC. Some others have shown the direct relation between high expression of the receptor and a bad prognosis in CRC [196–198]. It has also been demonstrated that there are several RAGE ligands. These activate multiple signaling pathways that are involved in chemoresistance [199], angiogenesis [200], migration [201], invasion and proliferation [202–204]. Furthermore, an aptamer (a short, single-stranded DNA or RNA molecule that can bind to a specific target) against RAGE has been demonstrated to inhibit proangiogenic and proliferative features of CRC, providing knowledge for development of novel drugs for CRC treatment [205].

1.7. Detection of circulating tumor cells (CTCs)

CTCs are tumoral cells that detach from a primary or metastatic solid tumor and intravasate into the bloodstream. These cells provides a valuable tool for studying cancer progression, metastasis and response to treatment, however their concentration in blood is very low (1-10 cells per ml of blood), making their detection difficult. Several studies have demonstrated that the detection of CTCs predicts survival in different cancer types such as breast, prostate, colon, rectal, lung and pancreatic cancer [206–209].

The CellSearch® system, developed by Veridex, is the first and only technology approved by the FDA for detecting CTCs in patients with metastatic tumors. It consists of an immunomagnetic separation of the cells by using nanoparticles coated with antibodies targeting EPCAM and a negative selection of CD45 cells. CD45 is a membrane protein that is usually expressed in immune cells. Although the CellSearch® system can detect CTCs in several metastatic carcinomas [210], some studies have shown that a combination of biomarkers improves the detection of CTCs [211].

Moreover, the CellSearch® system is not approved for evaluation of solid tumors, as in the early stages of the disease the number of CTCs is low, and the system is not able to detect them. For this reason, new methods that combine biomarkers and improve sensitivity and specificity of the method should be studied. Finally, methods that detect CTCs in early stages of cancer are also relevant for detecting cancer relapse sooner [212].

Previous work carried out in the Cancer Stem Cell laboratory focused on developing a new method (CMT) to detect CTCs both in metastatic and early stages of cancer [213]. This study was based on the quantification of the expression of four biomarkers: BCRP1, CD133, AC133 and EPCAM, along with a ROC curve to determine a threshold of high and low expression of biomarkers. 44 samples were then analyzed with this new method and with the approved CellSearch® system to predict relapse.

Several metrics were used to compare the methods:

- Positive predictive value (PPV). It expresses the percentage of correctly classified instances in the set of the instances classified as positive.

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \in [0,1]$$

- Negative predictive value (NPV). It expresses the percentage of correctly classified instances in the set of the instances classified as negative.

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \in [0,1]$$

- Sensitivity. It expresses the percentage of correctly classified instances in the set of the truly positive instances.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \in [0,1]$$

- Specificity. It expresses the percentage of correctly classified instances in the set of the truly negative instances.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \in [0,1]$$

Where TN represents the true negative instances, TP the true positive instances, FN the false negative instances and FP the false positive instances.

The CellSearch® system was 93.8% successful in detecting positive samples (positive predictive value, PPV) although its negative predictive value (NPV, the likelihood that a sample that is considered negative by a method is in fact negative) was of only 42.9%. In contrast, the newly established method achieved positive predictive values of 81.1% and negative predictive values of 85.7%.

Moreover, if we analyze the percentage of real positives that the model predicted as positive (sensitivity) it was 48.4% for CellSearch® system and 96.8% for the new method. This result shows that the new method is able to assign a positive result (patient will relapse) to almost all positive patients for relapse. However, if we analyze the specificity of the model, or the percentage of negative samples that the model was able to predict as negative, it was 92.3% for the CellSearch® system and 46.2% for the new method. This is the weak point of the method, and further investigation is needed to improve the specificity.

All these results suggest that the laboratory's prior prediction model (CMT) is an improvement over the established method and is more likely to guess the correct outcome. However, further investigation is needed to accurately identify patients who will not relapse, as the model predicted relapse in 53.8% of negative cases (patients that will not relapse). This new method improves the CellSearch® detection model to predict relapses independently of the stage of the tumor, but requires some improvements in sensitivity and specificity.

In the current study, we broaden this approach by developing an updated method that increases the number of samples and biomarkers analyzed, while integrating AI algorithms to improve the parameters of sensitivity and specificity.

1.8. Machine Learning as a tool for biomedicine

Mathematical tools play a crucial role in biomedicine, enabling the modeling and understanding of complex biological systems. These tools have been widely used to simulate several biological processes, including disease dynamics, cellular interactions, and treatment responses [214,215]. Moreover, AI has revolutionized biomedicine, particularly in oncology, by providing advanced tools for cancer diagnosis, treatment and monitoring [216,217].

In this PhD thesis, we build upon these advancements to develop a mathematical model for predicting relapses in CRC patients. To achieve this, we used ML, a subfield of AI, which consists of a collection of techniques and technologies that enable machines to simulate human intelligence. Among AI technologies, ML focuses on enabling systems to learn from data and improve their capabilities over time without direct human intervention. This thesis will focus on two main types of ML: supervised ML and unsupervised ML.

On the one hand, supervised ML models (Figure I10) learn from labelled data, that is, a variable is the target, and we already know the result. The model will be trained with this data and the database is divided into two (training and prediction databases) to train and evaluate the model several times.

The main supervised ML techniques are classification and regression with the principal difference between them being that classification predicts a categorical variable while regression predicts a numerical variable.

In our study, an example of a classification problem arises when data includes the biomarker expression levels and a target called relapse, with outcomes categorized as True/False (categorical). In this case, the model uses the biomarker expression data to classify samples and predict the likelihood of relapse. Conversely, an example of a regression problem occurs when the data contains biomarker expression levels, but the target variable is numerical, such as the number of days to relapse. Here, the model aims to predict a continuous outcome based on the biomarker data.

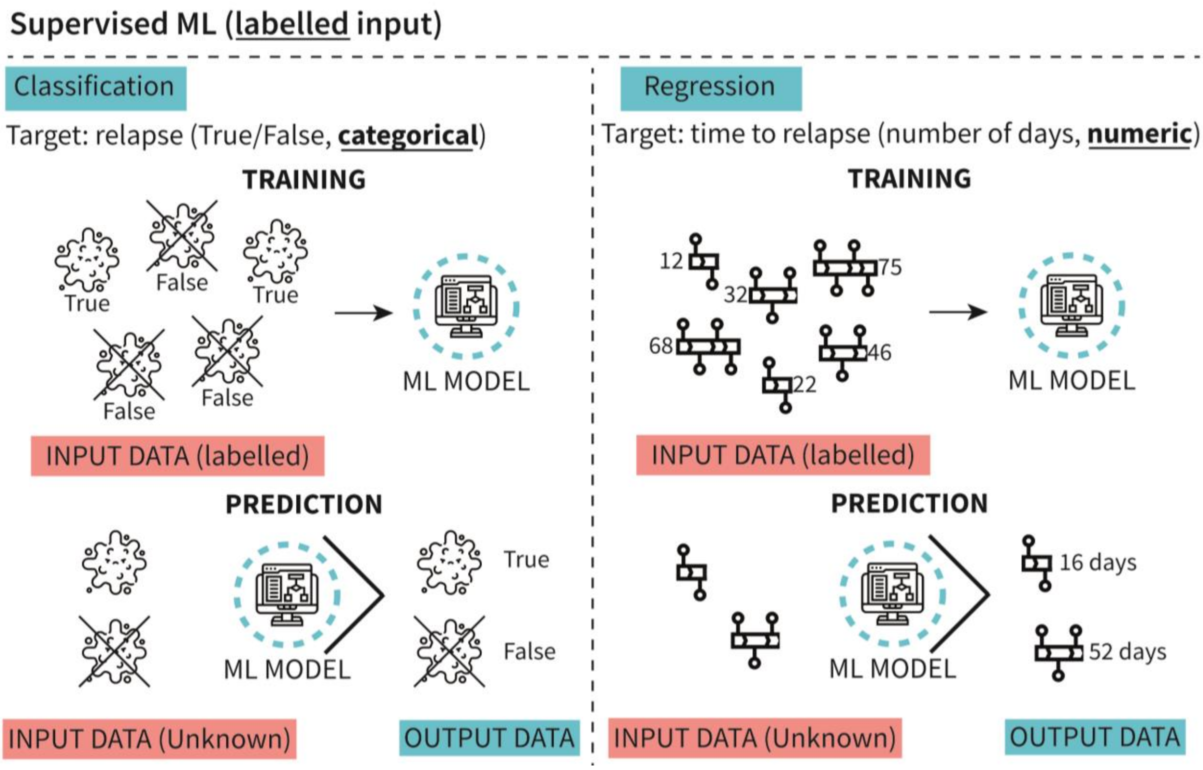

Figure I10. Supervised ML. Supervised ML includes classification and regression models. In both cases, the input consists of labelled data. A training dataset is used to generate a model which can then be used to predict the value of the target for any unknown data. In the classification models the target is a categorical variable while in the regression models it is a numerical variable. Original figure created by the author using symbols designed by Freepick (https://www.freepik.es/).

On the other hand, unsupervised ML models (Figure I11) learn from unlabeled data by identifying patterns or structures within it. There are different types of unsupervised ML, such as clustering or dimensionality reduction. Clustering is an unsupervised technique that is used to group data in clusters based on similarity or proximity between data.

An example of clustering is to generate groups of samples of patients depending on the expression of biomarkers. In this case, we do not have any target, but we can cluster the samples and then analyze the characteristics of each cluster. For example, we can evaluate if the generated clusters separate patients that are receiving treatment from those who are not.

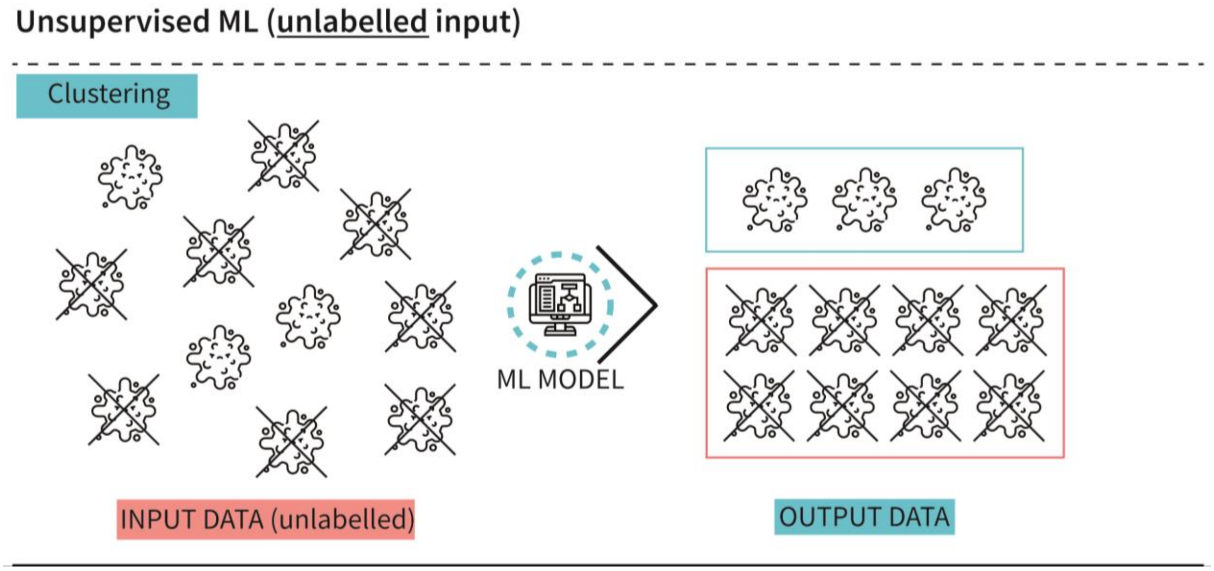

Figure I11. Unsupervised ML. In the case of unsupervised ML, the input data are unlabeled. Clustering is an unsupervised ML model that groups data in clusters based on similarity or distance between data. Original figure created by the author using symbols designed by Freepick (https://www.freepik.es/).

Different models have been developed for each type of ML (Figure I12) [218]:

- Supervised ML models:
  - Logistic Regression. It is usually employed for binary classification (true/false, yes/no) by predicting probabilities using a logistic function (sigmoid curve).
  - Decision tree. It splits the data based on the variables values to maximize the information, i.e., finding at each node a threshold for a feature that provides the maximum entropy reduction.
  - Random Forest. It builds multiple decision trees and combines the outputs. It reduces the risk of overfitting when compared to a single decision tree.
  - Support Vector Machine (SVM). It tries to find the optimal hyperplane that separates data points of different classes with the maximum margin.

  - K-Nearest Neighbors (kNN). It is a simple non-parametric algorithm that classifies data points based on the class of the nearest points in the variable space. It assumes that similar data are likely to belong to the same class.
  - Linear regression. It is a simple algorithm that assumes a linear relationship between input variables and the output variable (target).
  - Gradient boosting. It is an ensemble method that builds a series of weak learners iteratively and each new learner corrects the errors of the previous ones by minimizing a loss of function.
  - Neural Network. It consists of layers of interconnected nodes that process input data and learn complex relationships between variables and the target.
- Unsupervised ML models:
  - K-means. It generates K cluster centroids and assigns each data point to the nearest centroid based on a distance metric (commonly Euclidean distance). Then, centroids are recalculated by considering the mean of all the points assigned for each cluster. This process is iterated several times until the centroids no longer change significantly.
  - Hierarchical clustering. It builds a hierarchy of clusters using a tree-like structure, called dendrogram. It first calculates the pairwise distances between all data points using a distance metric (commonly Euclidean distance) and saves the results in a matrix. It is iterated by updating the clusters based on their distance. The main difference with K-means is that the number of clusters is not predesigned, a dendrogram is plotted and you can decide where to set a threshold. If the threshold is defined at a higher level, fewer clusters are defined.

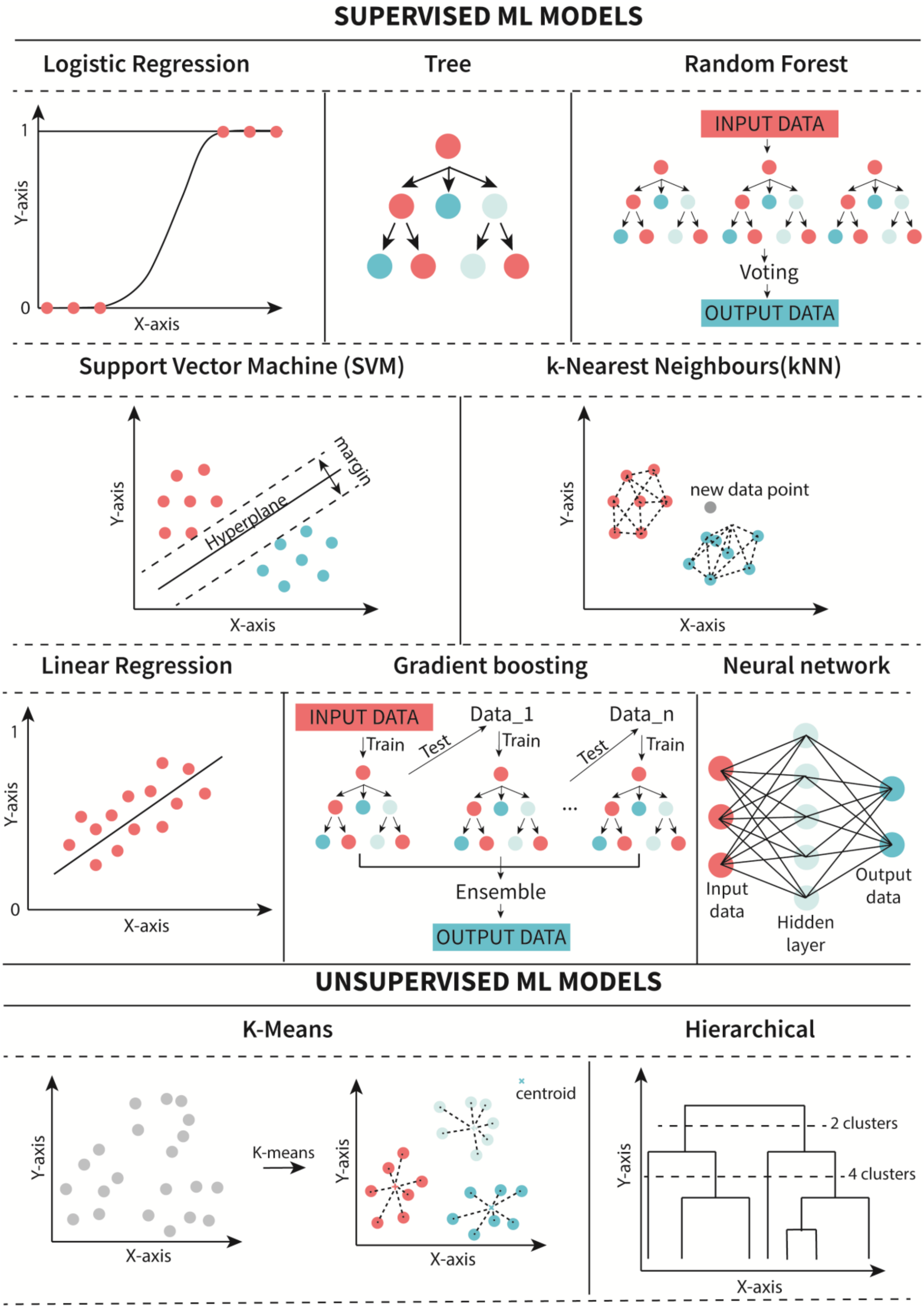

Figure I12. Supervised and Unsupervised ML models. Logistic Regression, Tree, Random Forest, SVM, kNN, Linear regression, Gradient boosting, Neural network, k-Means and hierarchical clustering methods are represented. Original figure created by the author.

Finally, several metrics are used for evaluating the different models [219]:

- Classification accuracy (CA). It expresses the percentage of correctly classified instances in the set of all the instances.

$$\text{CA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \in [0,1]$$

- Precision (Prec). It expresses the percentage of correctly classified instances in the set of the instances classified as positive.

$$\text{Prec} = \frac{\text{TP}}{\text{TP} + \text{FP}} \in [0,1]$$

- Sensitivity (usually referred as recall). It expresses the percentage of correctly classified instances in the set of the truly positive instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \in [0,1]$$

- F1 Score. It is an harmonic mean of precision and recall.

$$\text{F1} = \frac{2 \times \text{Prec} \times \text{Recall}}{\text{Prec} + \text{Recall}} \in [0,1]$$

- Specificity (Spec). It expresses the percentage of correctly classified instances in the set of the truly negative instances.

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \in [0,1]$$

- Area under the curve (AUC). It is the area under the receiver operating characteristic (ROC) curve, obtained by plotting sensitivity against the false positive rate (1-specificity) at all possible threshold values. The closer to the diagonal reference line, the lower the predictive capacity. The area under the ROC curve, AUC, is another evaluation metric with values from 0 to 1.

2. Objectives

The main objective of this PhD thesis was to develop a mathematical model for the early prediction of relapses in CRC patients, with the goal of implementing a novel diagnostic tool that integrates the expression levels of multiple CSCs biomarkers measured in blood samples. Additionally, this work also focused on exploring resistance to chemotherapy by developing an innovative analytical strategy to evaluate treatment efficacy and identifying a new biomarker candidate for resistance detection.

To achieve these main objectives, we established several specific objectives for each:

1. Development of a mathematical model for the early prediction of relapses in CRC patients.
    1.1. Selection of CSCs biomarkers previously described in the literature for this population.
    1.2. Expression of selected CSCs biomarkers in blood samples from patients previously diagnosed with CRC.
    1.3. Correlation analysis between clinicopathological characteristics of CRC diagnosed patients and the expression of CSCs biomarkers.
    1.4. Training ML models to develop a mathematical model for prediction of relapses in CRC patients.
    1.5. Evaluation of the obtained ML predictive model.
2. Study of resistance to chemotherapy in CRC.
    2.1. Generation of patient-derived xenograft tumors.
    2.2. Development of a novel strategy to evaluate treatment efficacy.
    2.3. Identification of novel biomarker candidates for resistance detection.

3. Materials and methods

This section provides a detailed description of the experimental procedures employed in this PhD thesis. It is divided into two main subsections, corresponding to the main objectives of the PhD thesis: a) development of a mathematical model to predict relapses in CRC patients, and b) study of resistance to chemotherapy. Further sections were included to structure the conducted experiments and address specific objectives of the thesis.

3.1. Mathematical model to predict relapses in CRC patients

Briefly, this section describes the procedures employed to assess the development of a mathematical model for early prediction of relapses in CRC patients by measuring the expression of multiple CSCs biomarkers. To achieve this, we first selected the biomarkers for study and analyzed their expression in CRC cell lines. Then, we processed blood samples from patients with CRC from different centers and from healthy people, isolated the cells and measured the expression of the selected biomarkers. Finally, we processed and analyzed all the clinicopathological information together with the experimental data in different databases.

3.1.1. Selection of CSC biomarkers

We first searched for references on PubMed to identify biomarkers previously reported in the literature as CSC markers, focusing specifically on markers of CRC, given its relevance to this study. For this, we conducted a PubMed search using the terms “CSC biomarkers” and “CRC”, reviewing a total of 116 articles. From these, 13 biomarkers were selected, prioritizing membrane proteins, as flow cytometry was the chosen technique in this PhD thesis. We considered those better characterized specifically in CRC and included previously used biomarkers from the CMT method. Then, we quantified their expression in established CRC cell lines before proceeding with patient samples.

3.1.1.1. Cell lines

Five human CRC cell lines were employed: SW-480 (CCL-228), SW-620 (CCL-227), DLD-1 (CCL-221), HCT116 (CCL-247), HT29 (HTB-38). All of them were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). The experiments were conducted immediately after purchase, ensuring the quality of the cell lines. We included CRC cell lines with diverse mutational backgrounds (Table M1) to better capture the intratumoral heterogeneity observed in patients, thus providing a more accurate representation of the disease.

The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Corning, New York, NY, USA) supplemented with 10% fetal bovine serum (FBS, PAN Biotech, Aidenbach, Germany), 1% L-glutamine (Cytiva, Washington, DC, USA), and 1% penicillin/streptomycin (Corning, New York, NY, USA) in a 37ºC humidified incubator (Series II Water Jacket, Thermo Scientific, Waltham, MA, USA) with 5% CO_2.

Table M1. Characteristics of CRC cell lines. Data obtained from [220].

Characteristic	SW480	SW620	DLD-1	HCT116	HT29
Tissue of origin	Colon	Colon	Colon	Colon	Colon
Type	Primary	Metastatic	Primary	Primary	Primary
Dukes' stage	B	C	C	C	C
Year of establishment	1976	1976	1979	1979	1964
Patient's sex	Male	Male	Male	Male	Female
Patient's age	50	50	58	48	44
CEA expression	High	High	High	Moderate	High
Tumorigenic capacity	Yes	Yes	Yes	Yes	Yes
Heterogeneity in culture	Moderate	Moderate	High	Low	High
MSI status	MSS	MSS	MSI	MSI	MSS
APC mutation	Mutated	Mutated	Mutated	Wild-type	Wild-type
p53 mutation	Mutated	Mutated	Mutated	Mutated	Mutated
KRAS mutation	Mutated	Mutated	Mutated	Mutated	Wild-type
BRAF mutation	Wild-type	Wild-type	Wild-type	Wild-type	Mutated
Nuclear β-catenin localization	Yes	Yes	Yes	No	Yes
c-MYC expression	High	Moderate	High	Moderate	High

3.1.1.2. Gene expression analysis by RT-qPCR

Cells were harvested with trypsin for 5 min at 37ºC and centrifuged at 300 *g* for 5 min. Total RNA from cultured cells was extracted using the RNeasy Mini kit (Qiagen, Hilden, Germany) and used immediately for reverse transcriptase reactions or stored at -80ºC until use. RNA concentration was measured in a Nanodrop spectrophotometer (ThermoFisher, Walthman, MA, USA). cDNA synthesis was performed using the RevertAid First Strand cDNA Synthesis kit (Fermentas, Waltham, MA, USA). The reaction was prepared according to the manufacturer's instructions: a mixture containing 1 µg of RNA, 1 µL of random hexamers and Milli-Q H_2O until 12 µL was incubated at 70 ºC for 5 min. Then, 4 µL of buffer 5X, 1 µL ribonuclease inhibitor, 2 µL 10 mM dNTP and 1 µL reverse transcriptase were added to each tube. The reaction mixture was then incubated at 25 ºC for 5 min, 42ºC for 1 h and inactivated at 70 ºC for 10 min. All cDNAs were stored at -20 ºC until further use.

The quantitative PCR reaction was performed using the KiCqStart SYBR Green kit (Sigma, St. Louis, MO, USA) according to the manufacturer's instructions. KiCqStart SYBR Green predesigned primers (Sigma, St. Louis, MO, USA) were employed for the genes listed in Table M2 except *CD44V6* and *RAGE* that were designed in Primer3 Plus [221]. All primers were used in a final concentration of 300 nM, and 10 ng of cDNA were used per well, for a total volume of 10 µL. All cDNA samples were measured in triplicate in a 96-well plate covered with adhesive seals in the thermocycler Roche LightCycler 480 (Roche, Basel, Switzerland). Reactions started with 10 min at 95ºC, followed by 45 cycles of 15 s at 95 ºC, 1 min at 60ºC and 10 s at 72ºC. The $2^{-\Delta CT}$ method was used for calculating the normalized mRNA expression. *β-ACTIN* was used as a housekeeping gene (a gene that is constitutively expressed across all cells and is usually employed to normalize the quantity of genetic material between samples).

Table M2. Oligonucleotides used for RT-qPCR.

Gene	**Oligo name**	**Tm (ºC)**	**Sequence (5'→3')**
ABCG2	FH1_ABCG2	57.6	AAAGCCACAGAGATCATAGAG
	RH1_ABCG2	56.1	GATCTTCTTCTTCTTCTCACC
PROM1	FH3_PROM1	60.0	GTAAGAACCCGGATCAAAAG
	RH3_PROM1	59.0	GCTCTGGAGTTTCATTCAAG
LGR5	FH1_LGR5	58.3	AAATGCCTTATGCTTACCAG
	RH1_LGR5	58.3	ATCTTGAGCCTGAAACATTC
ALCAM	FH2_ALCAM	58.8	GATACCATTATCATACCTTGCC
	RH2_ALCAM	61.6	ATCGGGCTTTTCATATTTCC
CD34	FH1_CD34	58.2	ACCAATCTGACCTGAAAAAG
	RH1_CD34	60.1	GATTCATCAGGAAATAGCCAG
CD36	FH1_CD36	58.4	AGCTTTCCAATGATTAGACG
	RH1_CD36	55.6	GTTTCTACAAGCTCTGGTTC
EPCAM	FH1_EPCAM	56.3	GTATGAGAAGGCTGAGATAAAG
	RH1_EPCAM	58.9	CTTCAAAGATGTCTTCGTCC
TACSTD2	FH2_TACSTD2	59.2	AAGTAACTGAATCCATTGCG
	RH2_TACSTD2	58.9	ACAAACGATGGTAAATGAGG
CD44V6	FW_CD44V6	63.4	AACAGCTACCCAGAAGGAACAG
	RV_CD44V6	63.4	CTTTGGGTGTTTGGCGATA
DCLK1	FH1_DCLK1	58.7	GAATGTATTGGATAAGACCACC
	RH1_DCLK1	59.3	GGTCTTATTAAAAGGGCGAG
CD44	FH1_CD44	56.6	TTATCAGGAGACCAAGACAC
	RH1_CD44	61.5	ATCAGCCATTCTGGAATTTG
RAGE	FH1_RAGE	63.5	GACTCTTAGCTGGCACTTGGAT
	RH1_RAGE	64.0	GTCTCCTGGTCTGTTCCTTCAC
β-ACTIN	FH1_ACTB	59.7	GACGACATGGAGAAAATCTG
	RH1_ACTB	58.0	ATGATCTGGGTCATCTTCTC

3.1.2. Blood samples processing

3.1.2.1. Ethical committee evaluations

Evaluation of the project by the corresponding ethical committees was requested and approved for all five centers: Jimenez Diaz Hospital (JDH), Ramon y Cajal Hospital (RCH), Ciudad Real Hospital (CRH), Toledo University Hospital (TUH) and Daroca Health Center (DHC). Ethical approval of the project was also obtained from the Universidad Politécnica de Madrid as samples were processed on the university's premises. All the approved ethical evaluations are attached to Annex A1.

3.1.2.2. Samples

Samples were obtained from patients previously diagnosed with CRC. The inclusion criteria for this project were:

- Adult patients (>18 years old), previously diagnosed with colon or rectal cancer.
- Accepted and signed consent form for each patient.
- Availability for periodic extraction of blood samples.
- Absence of any other neoplasia.

The exclusion criteria for this project were:

- Prior infection with hepatitis B, human immunodeficiency virus (HIV), etc.
- Aplastic anemia.

Blood samples were collected in 4 ml EDTA conventional tubes (Grenier Bio-One, Frickenhausen, Germany) from patients previously diagnosed with CRC and from healthy people (25 samples). We received samples from five centers: JDH, RCH, CRH, TUH and DHC. The number of patients and samples are shown in Table M3.

Table M3. Number of patients and samples included in the study.

Center	Number of patients	Number of samples
JDH	16	17
RCH	11	31
CRH	54	85
TUH	23	31
DHC	88	225
TOTAL	**192**	**389**

The samples were extracted from patients during their routine hospital visits and sent by courier to the laboratory for analysis. Samples were stored at room temperature in agitation and were processed within the 24 hours post-extraction.

3.1.2.3. Peripheral blood mononuclear cells (PBMCs) isolation

A Ficoll gradient was performed to isolate the PBMCs from blood samples. Blood was diluted 1:1 in PBS 1X (Cytiva, Marlborough, MA) and added slowly on a 2 ml Ficoll® (Cytiva, Marlborough, MA) layer dispensed in a 15 ml tube. Then, it was centrifuged at 400 *g* for 30 min without acceleration nor braking. The layer of cells was extracted by pipetting without disturbing the gradient and centrifuged at 400 *g* for 5 min. The pellet was washed twice in PBS 1X and then resuspended in PBS with 0.5% bovine serum albumin (BSA) (Sigma, St. Louis, MO, USA). A schematic workflow is shown in Figure M1. The cell yield in this procedure ranged from 2 to 3 million PBMCs.

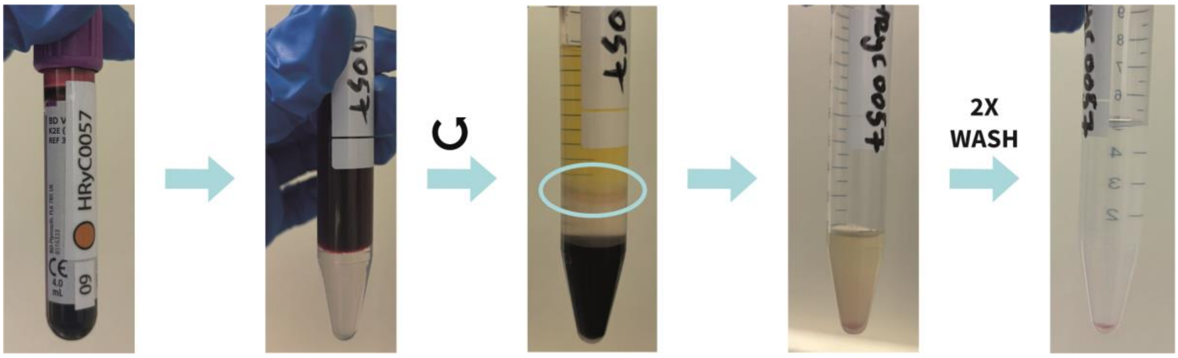

Figure M1. Schematic workflow to isolate PBMCs. Blood was diluted in PBS and disposed on a Ficoll layer. After centrifugation, the cells were isolated by pipetting, washed and resuspended in PBS-BSA 0.5%.

3.1.2.4. Immunostaining and flow cytometry

The expression of the different biomarkers was analyzed by flow cytometry. Approximately 250,000 cells per condition were incubated with the corresponding antibody in a volume of 50 µl in PBS-BSA 0.5% for 15 min at room temperature in darkness and agitation. Required dilution and commercial brand for each antibody are shown in Table M4. Next, the antibody was removed by centrifugation at 400 *g* for 5 min and fixed with Glyo-fixx™ (Thermo Scientific, Waltham, MA, USA) in a 1:1 proportion with PBS-BSA 0.5% for a final volume of 50 µl per tube. Cells were finally detected in a FACS Canto cytometer (Becton Dickinson, Franklin Lakes, NJ, USA). All the data was analyzed with FlowJo™ v10.8 Software (BD Life Science, Franklin Lakes, NJ, USA).

Table M4. Antibody dilution and commercial brand for flow cytometry application.

Antibody	Commercial brand	Dilution	Antibody	Commercial brand	Dilution
Anti-BCRP1	Miltenyi Biotec	1:10	Anti-TROP2	Becton Dickinson	1:100
Anti-CD133	Miltenyi Biotec	1:50	Anti-CD166	Becton Dickinson	1:10
Anti-AC133	Miltenyi Biotec	1:50	Anti-LGR5	Miltenyi Biotec S.L.	1:50
Anti-CD34	Miltenyi Biotec	1:50	Anti-DCAMKL1	Abcam	1:50
Anti-CD44	Becton Dickinson	1:50	Anti-CD36	Miltenyi Biotec	1:50
Anti-CD44V6	Miltenyi Biotec	1:50	Anti-RAGE	Abcam	1:500
Anti-EPCAM	Miltenyi Biotec	1:50			

3.1.3. Data processing

Two different databases were created, one for the biomarkers' expression and the other for clinicopathological characteristics.

- Biomarkers' expression database.

All the samples were acquired in the flow cytometer and analyzed by the FlowJo® program. Figure M2 shows an example of the analysis of a sample that corresponds to the BCRP1 biomarker in a patient from CRH. First, we selected lymphocytes population by its forward scatter (FSC) and side scatter (SSC) parameters, as the CSC population is expected to have similar shape and size. Then, single cells were selected by representing the FSC-A (area) and FSC-H (height) parameters. As single cells have the same area (FSC-A) and height (FSC-H), cells in the 45° angle were selected.

Next, we established a threshold from which expression will be considered positive by using the negative control (sample without staining) as a reference. The same threshold is applied to the rest of the biomarkers, and the cells above this threshold are considered positive. In the example, 41.6% of the population is positive for BCRP1.

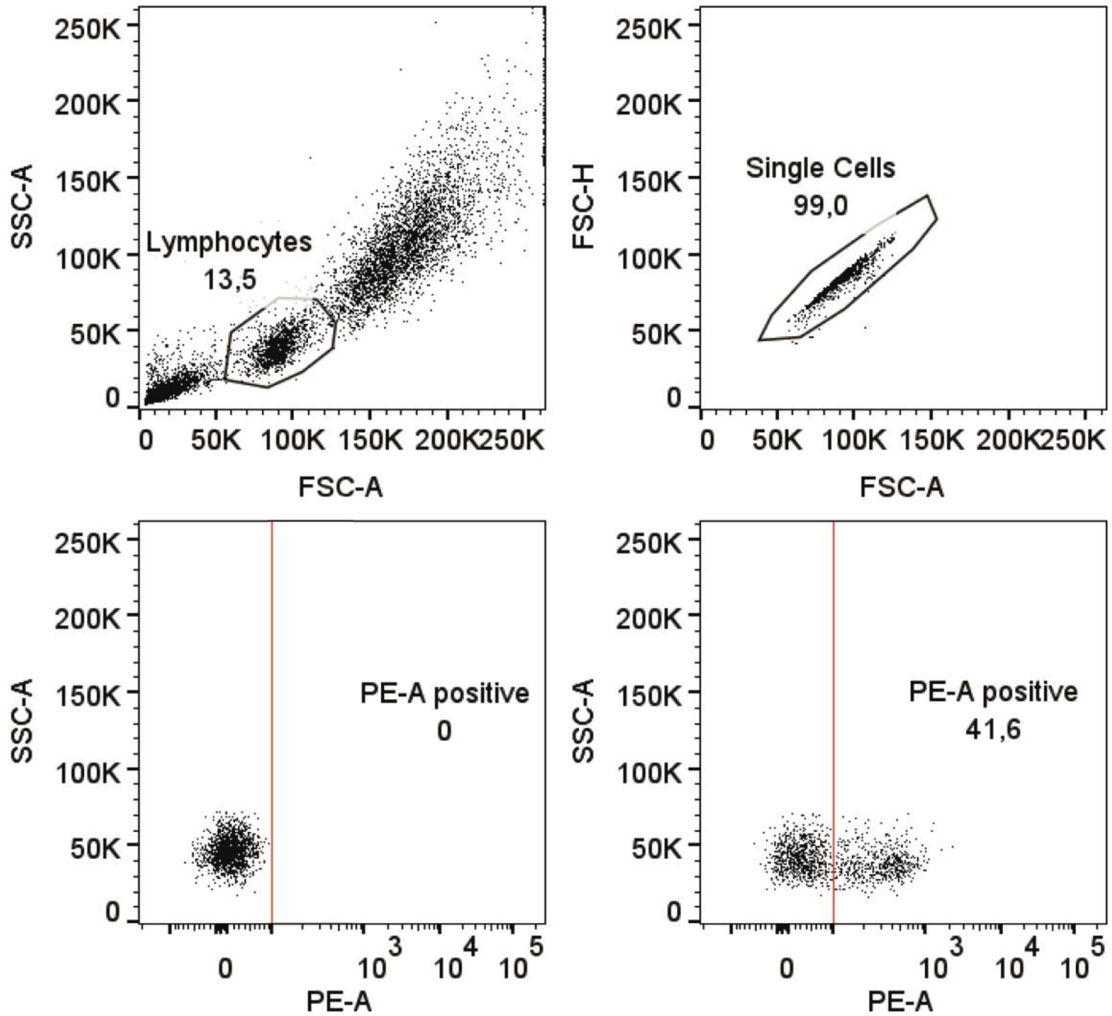

Figure M2. Analysis of biomarkers' expression by FlowJo®. Lymphocytes were first selected considering their size (FSC) and complexity (SSC). Then, single cells were selected considering they would present the same area (FSC-A) as height (FSC-H). Residual fluorescence is shown in the control (unstained) sample was used to establish a threshold from which samples were considered positive. The red line represents this threshold, and a positive sample is shown as an example on the right. All the cells that present higher fluorescence intensity than the control were considered positive. The percentage of cells above this threshold is also shown in the graphic.

Data was exported in two Excel sheets, one containing the threshold defined for each sample and the other containing the percentage of positive cells from the total of single cells, as well as the mean and median fluorescence for each biomarker. A third Excel file was used to register sample number, date of sample extraction, and corresponding patient ID.

A python script was developed to read the registration document along with the documents exported from FlowJo®, to process the data and to generate three corresponding Excel documents with filenames that include *cito*, *BD* and *report* as a prefix and which contain the following information:

- *Cito*: Combines the information provided by the three input Excel documents. It contains the sample ID, sample extraction date and the expression data for each biomarker, including the threshold, percentage of expression, mean and median of fluorescence. One document is generated for each patient which includes all the corresponding samples. The file contains eight rows per sample as each acquisition tube in the flow cytometer corresponds to a row. As we measure 13 biomarkers (in pairs) and acquire a negative control, each sample vial is processed in 8 acquisition tubes. This high number of rows per sample in the file complicates the analysis and requires further processing, via the *BD* file.
- *BD*: The eight rows per sample of the previous file are consolidated into a single row per sample, preserving all the information, as this format enhances data analysis. One document is generated for each patient (including all the samples for this patient).
- *Report*: A representation of the percentage of expression, mean and median is exported for each patient with all the samples included on it. The information in this output report is of use for medical practitioners to check the evolution of the biomarkers over time for each patient.

This script uses the *pandas*, *numpy*, and *os* libraries.

- Clinicopathological characteristics database.

A database including all the clinicopathological characteristics was needed at this point of the project. We prepared a template to ensure that all doctors recorded the data in a consistent manner, using dropdown menus. This approach prevented manual entry, which could lead to inconsistencies when different doctors registered certain data. The information registered in this database is cumulative and is continuously updated.

The following considerations were taken into account:

- Due to the participation of several centers and doctors, an easy to use and guided document was needed.
- Further computational analysis could not consider free text, only concrete options.
- Treatment information could vary across samples over the course of the study.
- Patients could relapse and be treated several times over the course of the study.

The database was thus generated as follows:

- Each column corresponds to a variable of interest and each row corresponds to a sample.
- The first 50 rows are the options to select from a dropdown list for each column.
- The options are standardized across all centers, but additional ones can be added if needed, making it a flexible database.
- Some columns are designed to add comments in case any information that is not considered in the dropdown options is needed.
- There are some columns that do not vary between samples. They are situated at the beginning of the table. Examples: stage at diagnosis, smoker, localization of primary tumor, etc.
- There are some columns that do vary between samples, as the different lines of treatment. Different columns are displayed for the different lines of treatment, to collect all the data.
- All the data from diagnosis and treatment are registered by date.

Once the two databases were created, one for the biomarkers' expression and another for the clinicopathological characteristics, we developed two more Python scripts that integrate the data from all the hospital centers.

- Patients database. This database contains the last sample for each patient (as it contains the most complete information). We selected variables that remain constant across samples, which will be used to analyze correlations with the variable *relapse*. These variables include the *sex, body mass index (BMI), family background, smoker, stage, T, N, M, localization* and *differentiation*. T, N and M columns were processed to generate three cohorts: N0M0 contains patients without nodules affected nor metastasis at diagnosis, N1M0 contains patients with nodules affected but without metastasis at diagnosis and M1 contains patients with metastasis at diagnosis.

  Samples database. This database contains all the samples of each patient (one sample per row) and integrates the data from all the hospital centers. It includes time-dependent variables, such as diagnostic and treatment information. To study possible correlations between the biomarker expression levels (which also vary over time) and the time-dependent variables, we defined and calculated several variables:

  - *Relapse*. We compared the date of the sample with the one for the diagnosis of each line of treatment for each sample. The lowest positive value was registered in a variable called *T_for_relapse*. This calculation was carried out in days, and another variable was generated to calculate in trimester and semester. Finally, a variable called *Relapse* (True/False) was generated. True was registered in cases in which *T_for_relapse* exist and False in cases in which it does not exist.
  - *T_since_surgery*. For each sample, we compared the date of the surgery of each line with the date of the sample. The lowest positive value was registered in a variable called *T_since_surgery*.
  - *T_since_treatment* (chemotherapy, radiotherapy or biological treatment). For each sample and line of treatment, the date of the treatment (beginning and end) was subtracted from the sample date, and these were consigned in the variables *T_since_treatment_beginning* and *T_since_treatment_end*.

In the case of chemotherapy, this was carried out both for neoadjuvant and adjuvant treatment. We only considered the lowest positive *T_since_treatment_beginning*, as it is the nearest to the sample and we do not considered negatives because it means that the treatment was given after sample processing. Different possibilities emerged:

- If *T_since_treatment_end* values for the different lines of treatment were all positive, the lowest value was registered in *T_since_treatment*.
- If *T_since_treatment_end* values for the different lines of treatment were all negative, the highest value was registered in *T_since_treatment*.
- If *T_since_tratment_end* values for the different lines of treatment had negative and positive values, the lowest value of positives was compared to the higher negative value. We compared it in absolute value and register the lowest in *T_since_treatment*.

Treatment in this part corresponds to chemotherapy, radiotherapy and biological treatment, as the same procedure was carried out. So, *T_since_chemotherapy*, *T_since_radiotherapy* and *T_since_biological* variables were generated.

- Treatment (categories). We created a new variable called *Treatment_sample* with three categories: *Sx (surgery) alone, Sx + other, other without Sx*. We considered different possibilities:
  - If *T_since_surgery* had a value and the rest were empty, we assigned *Sx alone* to this variable.
  - If *T_since_surgery* had a value and minimum one of the rest had a value, we assigned *Sx + others*.
  - If *T_since_surgery* did not have a value and minimum one of the rest had a value, we assigned *others without Sx*.
- *In_treatment* (True/False). If *T_since_treatment_end* was higher than the date of the sample we consigned True and if it was lower, we consigned False. We proceed the same for chemotherapy, radiotherapy and biological treatment.

If one of them was True, we assigned the value True to the variable *In_treatment*. If all of them were false, we assigned the value False to the variable.

- *In_treatment_since_last_sample* (True/False). We first calculated the time between a sample and the previous one for the same patient. If this difference was higher than the *T_since_treatment*, with treatment corresponding to surgery, chemotherapy, radiotherapy or biological treatment, the value True was assigned. If it was lower or there was no value in this variable, the value False was assigned. Moreover, the difference between the percentage of positive cells in that sample and the previous sample was calculated for each biomarker. The mean (*dif_mark_mean*), maximum (*dif_mark_max)* and minimum (*dif_mark_min*) value of these differences were also calculated.

The Python scripts and an example of the databases are available in github (https://github.com/JJ-Lab/CSC_ONCOREC). Data was now prepared to analyze correlations between variables and to generate a model for relapse prediction.

3.1.3.1. Correlation between variables

We first analyzed the variables that do not change between samples, that is *sex, BMI, family background, smoker, stage, cohort TNM, localization* and *differentiation*. We wanted to know if there was any correlation between all of them and relapse. For that purpose, we used the patient database and generated a boxplot in Graphpad Prism and a Chi Square test was carried out to analyze significant differences between groups.

Then, we analyzed the variable that does change between samples, which is mainly the treatment. The aim was to check whether any of the variables influenced the biomarkers' expression. We carried out the following procedures:

- Type of treatment: we evaluated if the type of treatment received until the date of the sample influenced biomarker expression. For that, we used the variable *Treatment_sample* which contains three groups: *Sx alone, Sx + other or other without Sx* (for more details read the previous subsection). In this case we used both supervised ML (taking the variable *Treatment_sample* as a target) and unsupervised ML (clustering). Further details are provided below.

- In treatment: we evaluated if biomarker expression was influenced if the patient was in treatment at the date of the sample or not. For that, we used both supervised ML (taking the variable In_treatment as a target) and unsupervised ML (clustering). Further details are provided below.
- *In_treatment_previous_sample*: we evaluated if the difference in biomarker expression between a given sample and the previous one was influenced if the patient had received any treatment during this time. For that, we used the variables *difference_biomarkers* and the variable *In_treatment_previous_sample* (for more details read the previous subsection). In this case we used both supervised ML (taking the variable *In_treatment_previous_sample* as a target) and unsupervised ML (clustering). Further details are provided below.

Finally, we developed a mathematical model for relapse prediction. For that, we used different supervised ML models (more details below). The input was the biomarker´ expression and the target was the variable *relapse* (True/False). We also included the variable *status* (*Treatment, Discharge, Follow-up*) as input to check whether it improves the model.

We cropped the database into two new ones, the training database and the prediction database. The last one included 25 random samples and the training one the rest of the samples. This was used to train and probe the model and was carried out in a Python script by using the *pandas* and *random* libraries.

Finally, we also checked whether we were able to predict the time to the next relapse, using the same input but establishing the variable *Time_to_relapse* as the target (for more details read the next subsection). All these models were generated in Orange software [222] and will be explained in the following subsection.

3.1.3.2. Supervised ML

The supervised ML models were carried out in Orange software. A workflow diagram is shown in Figure M3. Data was first imported from the database generated by the Python script. A data table was generated to check whether the results were well imported and to consult them in case needed in the analysis. Then, a column selection was performed. Input and target variables were selected in this widget. Normalization of the data was also carried out before testing and scoring different supervised ML based algorithms. SVM, kNN, Logistic regression, Random Forest, Gradient Boosting and Tree models were used for that purpose.

We performed stratified 5-fold cross-validation to evaluate the model's performance. The evaluation of the results was performed with the score widget, a confusion matrix and a ROC curve. The score widget calculates all the metrics described above in the introduction section: AUC, CA, F1, Prec, Recall and Spec. For more information refer to section 1.8.

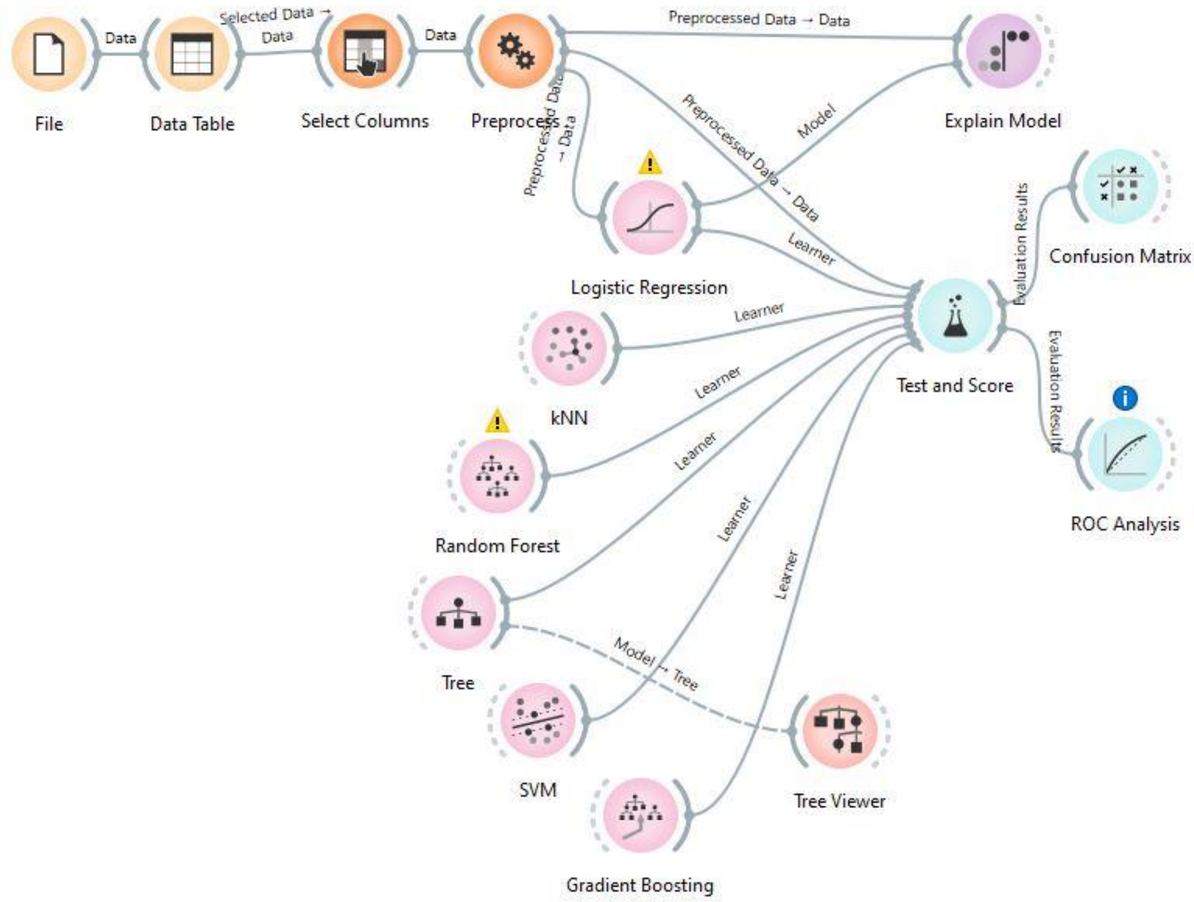

Figure M3. Workflow diagram to generate a mathematical model based on Supervised ML. Data was imported from the database and normalized. Then, different models (SVM, kNN, Logistic regression, Random Forest and Tree) were tested, and confusion matrix and a ROC curve were generated to evaluate the results.

3.1.3.3. Unsupervised ML

The unsupervised ML models were carried out in Orange software. In particular, we focused on clustering model. We tried to generate clusters with different inputs as described in the previous subsection and evaluated whether the clusters generated were defined based on the variable of interest. A workflow diagram is shown in Figure M4.

Data was first imported from the database generated by the Python script. A data table was generated to check whether the results were well imported and to consult them in case needed in the analysis. Then, a column selection was performed. Input variables were selected in this widget. The k-Means model was used to cluster the samples based on the selected input. This widget permits selection of the number of clusters based on the Silhouette score. Additionally, we applied Hierarchical Clustering, which first measures distances between samples to construct a clustering structure. Unlike k-Means, where the number of clusters is predefined, Hierarchical Clustering generates a dendrogram, allowing clusters to be selected based on a distance threshold rather than a fixed number. Results were visualized in boxplot and scatter plots for the variable of interest and the data that corresponded to each cluster was shown in a data table by first selecting the rows of interest in case we wanted to select a single cluster.

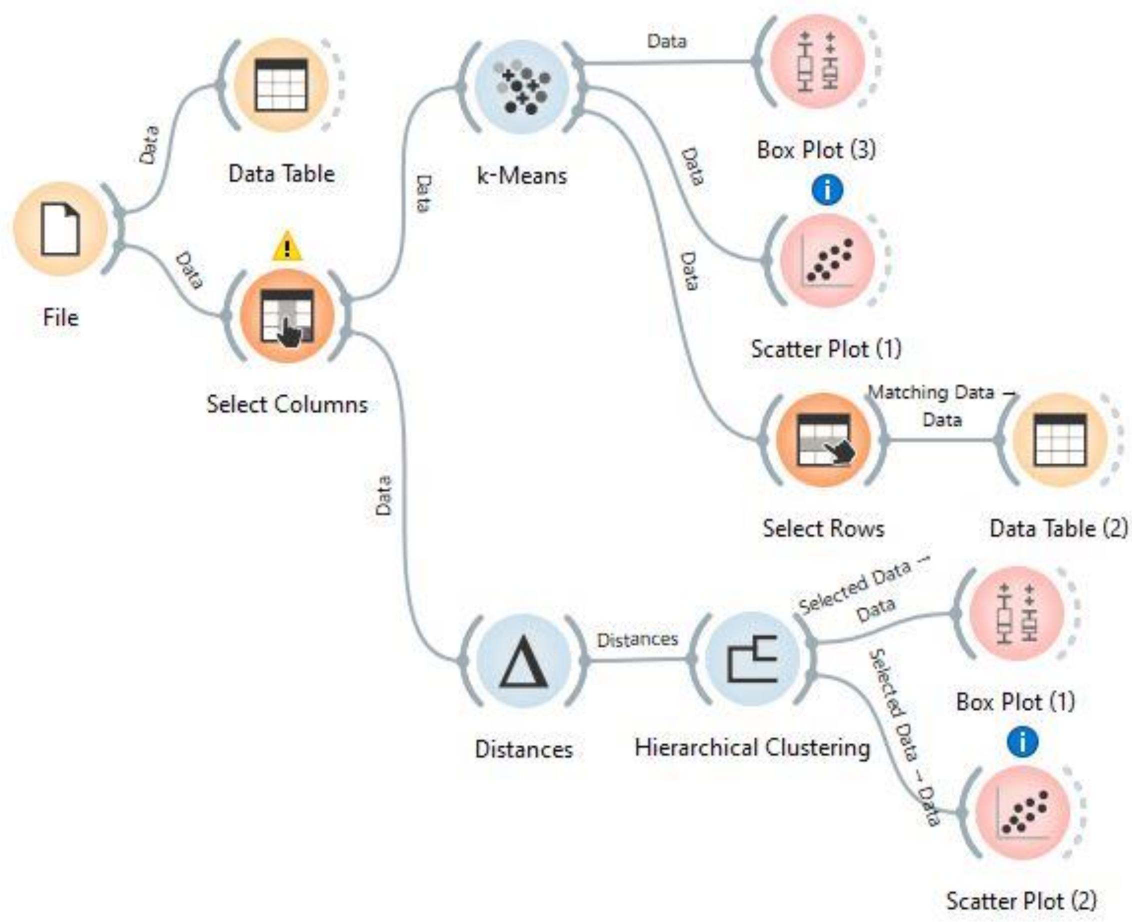

Figure M4. Workflow diagram to cluster samples based on Unsupervised ML. Data was imported from the database. Then, K-means model was used to generate clusters. The Silhouette score was used to select the number of clusters and a boxplot, scatter plot or data table were used to visualize the results.

3.2. Resistance to chemotherapy in CRC

The second main objective of the PhD thesis was to evaluate resistance to chemotherapy in CRC. To achieve this, we first analyzed the resistance and toxicity profiles of the patients included in the study. Subsequently, we cultured patient-derived tumors using various techniques and animal models. Furthermore, we developed a novel method to assess the efficacy of chemotherapy, addressing the limitations of the conventional IC_{50} method. Lastly, we investigated the role of the biomarkers selected for the relapse prediction model in chemotherapy resistance, aiming to identify new biomarker candidates for resistance detection.

3.2.1. Evaluation of resistance and toxicity in patients

We analyzed the variables *resistance* and *toxicity* (defined in the clinicopathological database) for the different lines of treatment and chemotherapeutic agents. For that purpose, we used the patient database and generated a boxplot in Graphpad Prism. A Chi Square test was carried out to analyze significant differences between groups.

3.2.2. Culture of patient-derived tumors

The purpose of this section was to generate patient-derived tumors using cells isolated from blood of CRC patients. They could then be used to carry out viability assays with different chemotherapy agents and determine the treatment that best fits each patient (precision medicine). Moreover, they could also serve to discover new biomarkers for resistance detection. To achieve this specific objective, we first selected those patients whose tumor samples were more likely to progress. Then, we enriched the CSCs population by immunomagnetic separation and finally injected the cells into different animal models that lack active immune systems, such as nude and NSG mice or chicken embryos. The tumors were then dissected and cultured in special conditions to establish patient-derived cultures. Finally, we also tried to grow tumors directly in culture without the use of animal models.

3.2.2.1. Selection of patients

A total of five patients were selected to generate patient-derived cultures. The main consideration for selection was the stage of the primary tumor, with stage IV patients preferred. Another consideration was if the patient relapsed in the last year. These requirements aimed to increase the probability of success as we considered more CSCs would be present in blood.

The blood was processed as previously described in section 3.1.2, and PBMCs were isolated. Then, an immunomagnetic separation was performed before seeding or injecting the cells into immune deficient mice.

3.2.2.2. Immunomagnetic separation

To enrich the population of CSCs in the sample, we selected CD45 leukocyte common antigen negative cells by immunomagnetic separation. PBMCs were incubated with CD45 antibody (Miltenyi Biotec, Bergisch Gladbach, Germany) in a 1:10 dilution in PBS for 15 min at 4ºC. Then, two washes with 2 ml PBS and centrifuging at 400 *g* for 10 min were carried out. After that, the cells were resuspended in 80 µl of PBS and 20 µl of anti-mouse magnetic beads were added (Miltenyi Biotec, Bergisch Gladbach, Germany). They were incubated for 15 min at 4ºC and washed twice in 2 ml PBS and centrifuged at 400 *g* for 10 min. The pellet was resuspended in 1 ml of PBS.

The immunomagnetic separation was performed in a MACS® Manual separator (Miltenyi Biotec, Bergisch Gladbach, Germany) with MS Columns (Miltenyi Biotec, Bergisch Gladbach, Germany). First, the column was washed with 500 µl of PBS. Then, 1 ml of cells resuspended in PBS were added and allowed to pass through it. A tube was positioned below the column to recover the CD45 negative cells. Three washes with 500 µl of PBS were carried out. The washes were collected in the same CD45 negative cells tube. Finally, the column was removed from the magnet and 1 ml of PBS was added to elute with the syringe into a new CD45 positive cells tube.

3.2.2.3. Confocal microscopy

After selection of CD45 negative cells to enrich the CSC population, the separation was checked by confocal microscopy. Fixed cells were seeded in 9-well microscope slides (Thermo Scientific, Waltham, MA, USA) with poly-L-Lysine in a dilution 1/20 (Sigma, St. Louis, MO, USA).

After an incubation of 15 min at room temperature (RT), the supernatant was removed, and a mix of antibodies was added. Hoescht was used in a 1:1000 dilution to stain the nuclei, CD45 and CD68 conjugated to FITC (Miltenyi Biotec, Bergisch Gladbach, Germany) in a 1:50 dilution were used to stain the cytoplasm of immune cells and pan-cytokeratin and keratin-8 cojugated to PE (Miltenyi Biotec, Bergisch Gladbach, Germany) in a 1:50 dilution were used to stain the cytoplasm of possible tumoral cells. Microscope slides were incubated for 30 min at RT. Images were acquired in a Stellaris 8 Laser Scanning Confocal Microscope (Leica, Wetzlar, Germany) with Depletion Lasers (STED) at 592, 660, and 775 nm.

3.2.2.4. Mice handling

NSG and nude mice (Charles River, Wilmington, MA, USA) were maintained at the facility in the UCM School of Biology. Nude mice were used because they are immunodeficient rodents lacking thymus, resulting in little to no T-cell function. This makes them incapable of rejecting foreign cells, which is a requirement for this experiment as human cells are injected. We also used NSG mice, a strain characterized by its highly immunodeficiency. They lack T cells, B cells, and natural killers, which confers a higher success rate in the generation of human tumors. Mice were housed in cages placed on racks with assisted ventilation. They were bred in cages with water and sterile food pellets (Envigo, Huntingdon, United Kingdom) ad libitum. Mice were exposed to light-dark cycles of 12:12 hours. Temperature and humidity were also controlled by the facility. All the procedures were carried out under the ethical committee number 365-3-21 respecting all the European and Spanish regulation for animal handling.

3.2.2.5. Injection of tumoral cells in immune deficient mice

PBMCs or CD45 positive and negative selected cells were injected into the flank of nude mice or into the flank and the vein tail of immune deficient NSG mice. Approximately 2 million PBMCs were isolated from blood. However, the number of CD45-selected cells was not counted, as their quantity was too low, and we aimed to prevent any loss during counting.

Before the placement in the laboratory of Dr. Frédéric Hollande, who helped us to optimize the process, we injected cells isolated from 24 patients in the flank of nude mice. After optimization, CD45-selected cells isolated from 5 patients were injected into NSG mice, both in the flank and in the tail vein. Three independent repetitions were carried out. The survival rate of the animals was 100% in all cases.

In the case of the flank injection, a total volume of 100 µl in 1:1 Matrigel® (Corning, New York, USA) and DMEM medium (Corning, New York, NY, USA) was injected with a 1 ml insulin syringes (Becton Dickinson, Franklin Lakes, NJ, USA). In the case of vein tail, the tail was first warmed with warm water to dilate the veins. Then, a total of 100 µl were injected with a 1 ml insulin syringe in the tail vein. A representative image of the procedure is shown in Figure M5.

Figure M5. Injection of tumoral cells in immune deficient mice. Cells were injected both in the flank and in the tail vein. A total volume of 100 µl was employed.

3.2.2.6. Tumor cell seeding onto the chorioallantoic membrane of chicken embryos

All experiments were performed in accordance with the ethical committee number 2022-112. Fertilized eggs were provided by the farm of the National Institute for Agricultural and Food Research and Technology (INIA-CSIC) and incubated in a 96 eggs automatic rotary incubator (Expondo, Berlin, Germany) at 38ºC and 60-70% humidity. On day 10 of incubation, a small hole was perforated in the blunt end, where the air is usually localized. Then, another hole was perforated at 90 degrees, and a small window was made to introduce the PBMCs over the chorioallantoic membrane (CAM) [223,224]. Approximately 2 million PBMCs were isolated from blood and seeded onto the CAM using both the drop method and a sterile ring. The DLD1 cell line was used as a positive control, with 250,000 cells seeded and following the same procedure. The windows were then sealed, and the embryos were returned to the incubator. Daily monitoring was performed to detect potential tumor growth. After seven days, the embryos were macroscopically examined to assess tumor development. A representative image of the procedure is shown in Figure M6.

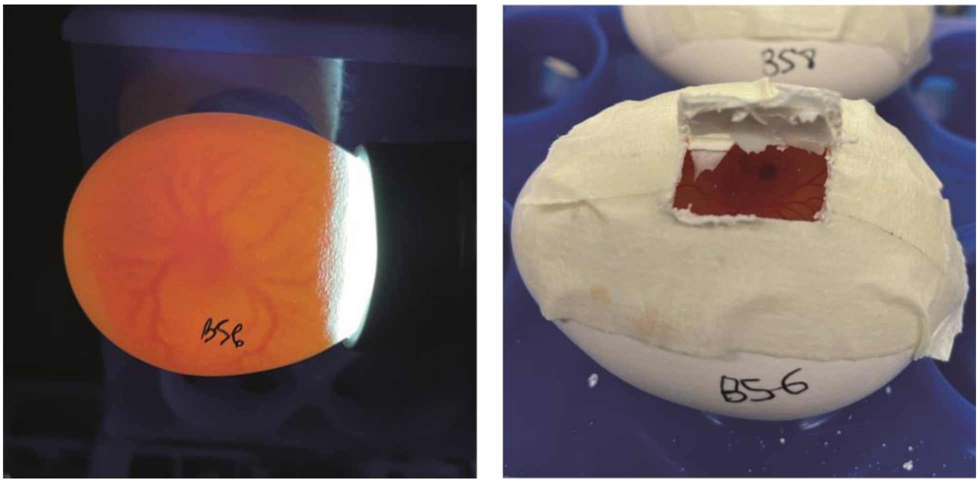

Figure M6. Injection of tumoral cells in chicken embryos. Cells were injected over the CAM after opening a small window in the shell.

3.2.2.7. Tumor dissection and culture

Tumors were dissected from the mice, washed in PBS 1X, disaggregated into small pieces using a scalpel and added to a 50 ml Falcon tube containing 10 ml of dissociation medium and incubated at 37ºC for 1h 30 min. The dissociation medium contained DMEM-F12 (Corning, New York, NY, USA) and collagenase (Thermo Scientific, Waltham, MA, USA) at 1 mg/ml. Every 20 min, we mechanically dissociated by vortex at 4ºC during 1 min. Then, DNase (Qiagen, Hilden, Germany was added to the suspension and incubated for 1 min at RT. The reaction was stopped by adding 10 ml of DMEM-F12 + 10% FBS, filtered in a 100 µm cell strainer and centrifuged at 220 *g* for 5 min. After removing the supernatant, we resuspended the pellet in DMEM-F12 and seeded in a dome of 40 µl in a proportion 70% Matrigel – 30% medium on a 24-well plate. Plates were incubated for 45 minutes and 500 µl of COM were added over the domes. Finally, the cells were incubated in a 37ºC humidified incubator (Series II Water Jacket, Thermo Scientific, Waltham, MA, USA) with 5% CO_2 until growing. Medium was changed every 3 days. This procedure is shown in Figure M7.

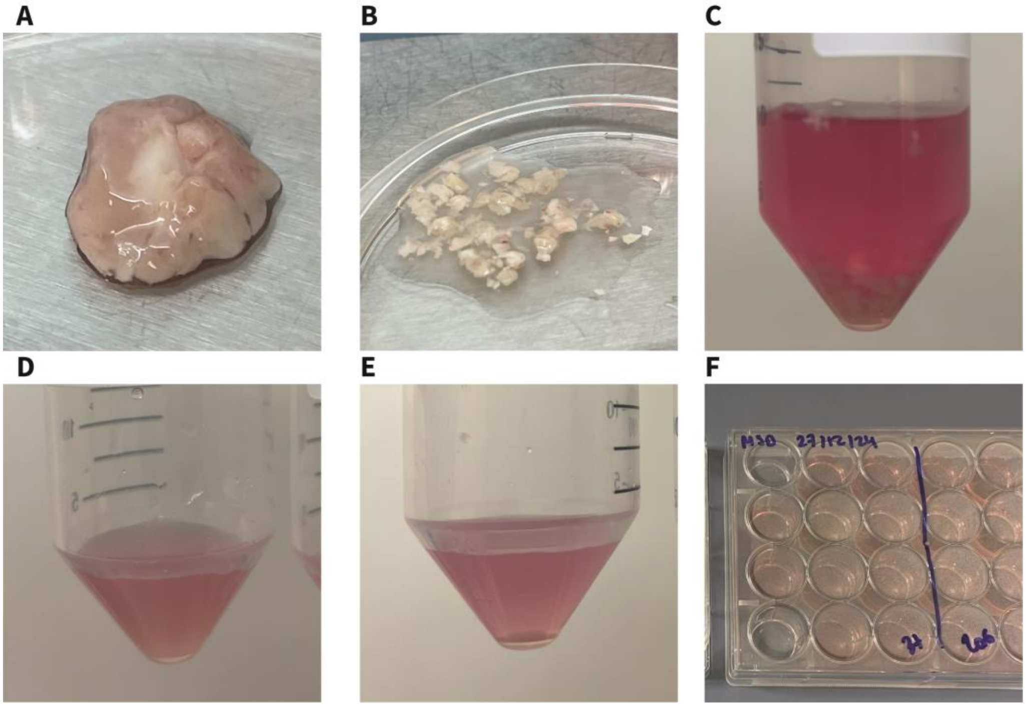

Figure M7. Dissection and culture of patient-derived tumors. (A) Tumor dissected from a NSG mouse (B) Tumor disaggregated with a scalpel into small pieces (C) Tumor in dissociation medium before enzymatic and mechanical dissociation (D) Tumor after enzymatic and mechanical dissociation (E) Pelleted cells after filtering (F) Seeding on a 24-well plate.

3.2.2.8. Patient-derived cell culture

We also tried to culture patient-derived cells directly in cell culture by two methods: in suspension and in Matrigel® growing.

In the suspension method, the CD45 positive and negative cells were seeded in M12 Medium (100 µl/well) in ultralow attachment 96-well plates (Clearline, Paris, France). M12 medium contains DMEM-F12 2 mM of L-glutamine, 100 Unit/ml of penicillin and streptomycin, N2 supplement (Pan-Biotech, Aidenbach, Germany), 20 ng/ml of epidermal growth factor (EGF, Corning, New York, NY, USA) and 10 ng/ml of fibroblast growth factor (FGF, Corning, New York, NY, USA). Cells were then maintained in a 37ºC humidified incubator (Series II Water Jacket, Thermo Scientific, Waltham, MA, USA).

In the Matrigel® method, the CD45 positive and negative cells were resuspended on 1:1 Matrigel® and BOM medium in a total volume of 40 µl. BOM medium contains DMEM-F12, 2 mM of L-glutamine and 100 Unit/ml of penicillin and streptomycin. Matrigel® was always stored on ice until seeded in a pre-warmed 24-well plate.

A sphere of Matrigel® and cells was formed and left to rest in a 37ºC humidified incubator for 30-45 min before adding 500 µl/well of pre-warmed COM medium. COM medium contains BOM medium, B27 supplement 2X (Thermo Scientific, Waltham, MA, USA), NAC 1 mM (MedChem Express, Shenzhen, China), YP27632 10 µM (MedChem Express, Shenzhen, China),, SB202190 10 nM (MedChem Express, Shenzhen, China), Gastrin 1 µg/ml (MedChem Express, Shenzhen, China), A8301 500 nM (MedChem Express, Shenzhen, China), and EGF 50 ng/ml (Corning, New York, NY, USA). Cells were then maintained in a 37ºC humidified incubator.

3.2.3. New growth rate-based indices to measure resistance to chemotherapy

We also developed a method to evaluate the efficacy of chemotherapy, with the aim of studying resistance. Our method serves as an alternative meant to address the main limitation of the conventional IC_{50} method: its time-dependent nature. We applied this method in the experiments of the last section, whose purpose is to discover new candidates as biomarkers for resistance detection.

To achieve this specific objective, we first performed viability assays for different cell lines and end-time points. Then, we analyzed the data by fitting the effective growth rate in an exponential and calculating different parameters to measure resistance, called IC_{50}, IC_{r0} and IC_{rmed}.

3.2.3.1. Cell lines

Five human CRC cell lines were employed: SW-480 (CCL-228), SW-620 (CCL-227), DLD-1 (CCL-221), HCT116 (CCL-247), HT29 (HTB-38). All of them were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Corning, New York, NY, USA) supplemented with 10% fetal bovine serum (FBS, PAN Biotech, Aidenbach, Germany), 1% L-glutamine (Cytiva, Washington, DC, USA), and 1% penicillin/streptomycin (Corning, New York, NY, USA) at a 37ºC humidified incubator (Series II Water Jacket, Thermo Scientific, Waltham, MA, USA) with 5% CO_2.

3.2.3.2. Viability assays

First, Thiazolyl blue tetrazolium bromide (MTT, BioChem, PanreacApplichem, Barcelona, Spain) was used to detect the viability to different concentrations of oxaliplatin (50 µg/mL, 10 serial 1:2 dilutions starting from the maximum doses) or irinotecan (20 µg/mL, 4 serial 1:2 dilutions starting from the maximum doses). Oxaliplatin and irinotecan were purchased from Accord Healthcare S.L.U., Barcelona, Spain. Briefly, cells were seeded in 96-well plates (Deltalab S.L, Barcelona, Spain) at a concentration of 100,000 cells/mL in a volume of 100 µL. Next, the chosen chemotherapeutic drug was added to each well. Three replicates per condition were performed on each plate with three independent experiments in total (conducted separately in time and materials to ensure reproducibility). After 0 (seeding time), 24, 48 and 72 h of the treatment, the medium was removed, and 50 µL of 0.5 mg/mL MTT was added. Then, plates were incubated for 4 h at 37 ºC, and finally, the medium was removed and resuspended in 100 µL of dimethyl sulfoxide (DMSO, Labkem, Barcelona, Spain) before detection of absorbance in a BIOBASE-EL 10A (Biobase, Shandong, China) spectrophotometer at 546 nm.

3.2.3.3. Mathematical model

As shown in Figure M8, colony population can be fitted as an exponential growth for short times, with an effective growth rate that can be positive or negative (indicating a population that increases or declines over time, respectively).

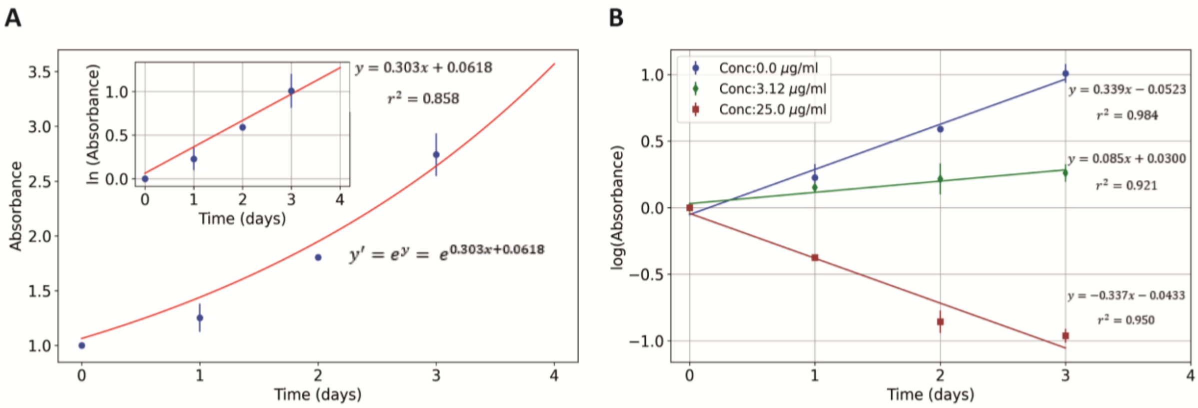

Figure M8. Exponential growth in HCT116 cell line in control and treatment conditions. (A) MTT absorbance reading over time fits with an exponential curve in control conditions. The inset shows this growth on a semi-logarithmic scale. (B) Exponential fitting of MTT absorbance readings on a semilogarithmic scale for different oxaliplatin concentrations (0, 3.12, and 25 µg/ml).

In cell viability experiments, the population of a colony subjected to a given concentration of chemotherapy drug, $\boldsymbol{N_C}$, will also follow an initial exponential growth, whose effective growth rate could be positive or negative. Then, cell viability can also be written as an exponential growth with an exponent equal to the difference between the sample growth rate and the control growth rate:

$$\boldsymbol{V_c(t)} = \frac{\boldsymbol{N_C(t)}}{\boldsymbol{N^*(t)}} = \frac{\boldsymbol{N_{0C} \cdot e^{r_C t}}}{\boldsymbol{N_0^* \cdot e^{r_0 t}}} = \frac{\boldsymbol{N_{0C}}}{\boldsymbol{N_0^*}} \cdot \boldsymbol{e^{(r_C - r_0) \cdot t}} \text{ (Eq. 1)}$$

where $\boldsymbol{N_{0C}}$ and $\boldsymbol{N_0^*}$ are the respective initial populations of the colony under treatment and the control (c=0); and $\boldsymbol{r_C}$ and $\boldsymbol{r_0}$ are the effective growth rates of the colony under treatment and the control, respectively. In this expression one can see the time-dependence of the viability. The greater the difference between r_c and r_0, the faster the viability will change over time.

In any case, from this equation the sample growth rate can be calculated for any viability value:

$$r_C = r_0 + \frac{1}{t} \cdot ln\left(V_c(t) \cdot \frac{N_0^*}{N_{0C}}\right) \text{ (Eq. 2)}$$

The absorbances measured from controls and treated cells were used to calculate the growth rates in each condition. First, the data were normalized to the absorbance obtained in the seeding time (t0). Then, the absorbances from 24h (t1), 48h (t2) and 72h (t3) were plotted and the effective growth rates were obtained from linear fits by the least squares method. To estimate the confidence intervals of the effective growth rate bootstrap methods with 1,000 bootstrap samples were used [225].

Bootstrapping is a technique to estimate the properties of a population statistic from the properties of a sampling data from an approximating distribution. The bootstrap idea is to use the sample distribution, instead of the population distribution. The sample distribution is made empirically by resampling from the original data. For each bootstrap sampling the same amount of data is randomly selected with replacement from the original sample. With each bootstrap sample the statistic of interest is calculated and finally a distribution of the statistic is obtained.

When data distribution is known or can be estimated, bootstrap sampling can be drawn from this distribution. In this case the bootstrap method is called parametric. In regression problems, with a covariate and a response variable, observed data are formed by pairs (X_i, Y_i); in this case the empirical bootstrap method (also called nonparametric bootstrap) consists of randomly sampling with replacement the pairs from the observed data (in the regression model empirical bootstrap is also called paired bootstrap). In an empirical bootstrap, covariate data values are independently randomly selected, so if one needs to sample all the predictors as fixed values, the residual bootstrap model may be used. In this instance one fits the observed data and determines the residuals for each observation. With the residual bootstrap technique, one selects bootstrap samples from the residuals and the new y_i^* is determined by $y_i^* = \hat{y}_i + e_b$, where $\hat{y}_i$ is the fitted value for x_i and e_b is an independently randomly selected bootstrap residual [226].

The effective growth rate is obtained from the slope distribution of all linear fits of the logarithm of MTT values versus time (one fit for each bootstrap sample), from a total of 1,000 bootstrap samples. This slope distribution can be fitted to a Gaussian distribution. From the bootstrap distribution one can obtain the effective growth rate (its mean, $\bar{r}$) and its standard error (SE). After verifying the statistic is roughly normally distributed a 100 (1-α) % confidence interval can be estimated:

$$r = \bar{r} \pm z_{\alpha/2} SE \text{ (Eq.3)}$$

where $z_{\alpha/2}$ is the standard normal value with probability α/2 to the right.

Curve fittings simulations were carried out in Python (Python Software Foundation, Wilmington, DE, USA) with the *scipy* [227] and *pandas* [228] libraries. Graphics were carried out with the *matplotlib* library [229]. The Python scripts are available in github (https://github.com/JJ-Lab/CSC_IC50).

3.2.3.4. Calculation of the IC_{50}, IC_{r0} and IC_{rmed} indices

As can be seen in Figure M9, the plot of growth rate vs. drug concentration shows a rapid initial decrease followed by a slower decrease for high concentrations. In most of the experiments (shown in semi-logarithmic scale and shifted to avoid negative values), one can observe that the logarithm of the growth rate decreases approximately linearly with time, so it can be fitted by linear regression (see inset of Figure M9).

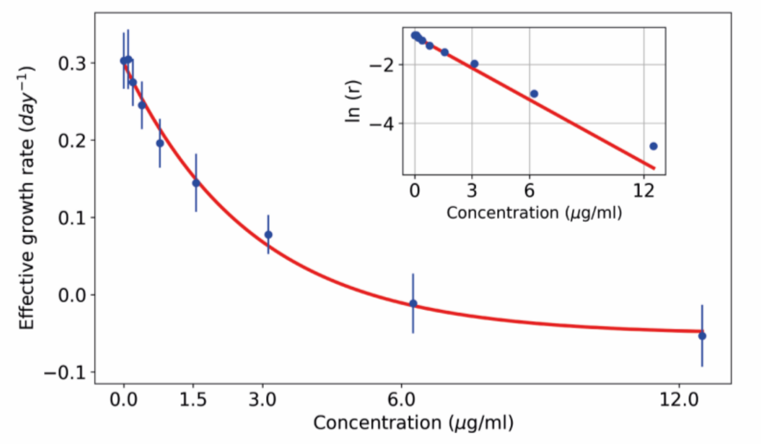

Figure M9. Concentration dependence of the effective growth rate in HCT116 cell line. The inset shows the same plot on a semilogarithmic scale.

When this fit is adequate, a mathematical model of the dependence of the growth rate on the drug concentration can be obtained. With this model the concentration for a given growth rate can be calculated:

$$r(C) = A \cdot e^{-\alpha \cdot C} + r_\infty \qquad \text{(Eq.4)}$$

where r_∞ is the negative shifted value (at very high drug concentrations the growth rate is negative) and α is the exponent. By applying logarithms and inverting Eq.4, the drug concentration for a given effective growth rate can be determined as follows:

$$Ln\left(\frac{r(C)-r_\infty}{A}\right) = -\alpha \cdot C \qquad \text{(Eq.5)}$$

$$C(r) = \frac{-1}{\alpha} ln\left(\frac{r-r_\infty}{A}\right) \qquad \text{(Eq.6)}$$

Let us maintain that we have calculated an effective growth rate for different drug concentrations using the bootstrapping technique. The bootstrapping technique provides a distribution of the effective growth rate from which we can obtain its mean and confidence intervals. Now we can also use the bootstrapping technique in the exponential fit of growth rates across different drug concentrations with this empirical distribution. In this instance we have performed a mixed bootstrap regression model: we have fixed the regressor values (as in the residual bootstrap method) and the response variable, r_i^* is generated by Monte Carlo simulation from the slope distribution as a parametric bootstrap (obtained in the previous effective growth rate calculation).

From each bootstrapping regression we get the fitting parameters (A, α, r_∞) needed in Eq. 6 to obtain the corresponding concentration for the IC_{50} and other indices. With the N_b bootstrap simulations an index distribution can be obtained, from which the expected value and the confidence intervals can be calculated. Monte Carlo simulations were carried out with the same libraries as described in the Bootstrap method.

We have found an expression (derived from Eq.2) for the effective growth rate corresponding to 50% viability, r_{IC50} (Eq.7).

$$r_{IC50} = r_0 + \frac{1}{t} \cdot ln\left(0.5 \cdot \frac{N_0^*}{N_{0C}}\right) \text{ (Eq. 7)}$$

The IC_{50} index is defined as the drug concentration that will produce 50% cell viability, so Eq.6 can be used to calculate IC_{50} (substituting r by r_{IC50}).

$$C(r_{IC50}) = \frac{-1}{\alpha} ln\left(\frac{r_{IC50} - r_\infty}{A}\right) \qquad \text{(Eq.8)}$$

The IC_{r0} index is a value that represents the drug concentration at which the effective growth rate is zero. The other, the IC_{rmed}, is the value that corresponds to the drug concentration that reduces the control population's growth rate by half. Similarly, IC_{r0} and IC_{rmed} can be calculated substituting in Eq.6 r by 0 and $r_C/2$, respectively.

$$C(0) = \frac{-1}{\alpha} ln\left(\frac{0 - r_\infty}{A}\right) \qquad \text{(Eq.9)}$$

$$C(r_{c/2}) = \frac{-1}{\alpha} ln\left(\frac{r_{c/2} - r_\infty}{A}\right) \qquad \text{(Eq.10)}$$

In summary, in order to obtain the IC_{50} index (and the other new indices) the following steps are required: i) determine the distribution of the effective growth rates for each concentration (previous section); ii) from the growth rate distribution get a sample of r for each drug concentration and perform a shifted exponential fit; iii) determine the corresponding effective growth rate for that index; iv) using Eq.8-10 calculate the concentration for $C(r_{IC50})$ or $C(0)$ or $C(r_{c/2})$; v) repeat steps ii) to iv) a large number of times (1000-5000 times is recommended [230]) to get a distribution of concentrations for r_{IC50}. From this distribution we can obtain the mean and confidence intervals.

In order to compare the IC_{50} index value obtained with this new procedure, the IC_{50} parameter was calculated by two conventional methods: GraphPad Prism 8.0 (GraphPad Software Inc., San Diego, CA, USA) and the IC_{50} calculator web program (https://www.aatbio.com/tools/ic50-calculator).

Both tools perform data fitting by assuming that the response curve follows a symmetrical sigmoidal shape, using a four-parameter logistic regression model (this value is also referred to as the relative IC_{50} parameter):

$$Y = Y_{min} + \frac{Y_{max} - Y_{min}}{1 + \left(\frac{X}{IC50}\right)^{H}} \text{ (Eq.11)}$$

where Y is the dependent variable, viability in this case. Y_{min} and Y_{max} are the minimum and maximum values of Y. X is the independent variable, in this case the concentration of a drug. IC_{50} is the half-maximal inhibitory concentration, or the value of X at which Y is halfway between Y_{min} and Y_{max}. Finally, H is the Hill coefficient, which determines the steepness of the curve.

3.2.4. Novel biomarker candidate for resistance detection

We finally searched for novel candidates as biomarkers for resistance detection. We first evaluated the expression of the biomarkers used in the study and found that TROP2 correlates with the resistance to chemotherapy calculated by the new proposed method IC_{rmed}. Then, we carried out some experiments to confirm this correlation, such as a chronic treatment with oxaliplatin to generate resistant cells and probe whether the resistant cells increased the levels of TROP2. We evaluated its expression both at gene and protein expression levels. Lastly, we silenced the gene to evaluate if the absence of the gene originated a cell culture that was more sensitive to chemotherapy.

3.2.4.1. Chronic treatment with oxaliplatin

The cells were grown in p100 plates for several days until exponential growth, to avoid the lag phase. Then, two different plates were seeded: a control plate containing non-treated cells (DMEM without oxaliplatin (Accord Healthcare S.L.U., Barcelona, Spain)) and a plate with treated cells (DMEM with 10 µg/mL oxaliplatin). The cells were exposed to oxaliplatin at 0.5 and 5 µg/ml for two weeks in the preliminary assay and at 10 µg/ml for 8 weeks in the broadened study. These treated and non-treated cells were then evaluated in viability (details in section 3.2.3.2) and gene/protein expression assays (for more information see below).

3.2.4.2. Silencing of *TACSTD2*

For *TACSTD2* silencing, SW480, SW620, DLD1, HCT116 and HT29 human CRC cell lines were transfected with a siRNA designed by Sigma to target that gene (Sigma, St. Louis, MO, USA). The reference SASI_Hs02_00333686 was used to silence *TACSTD2* and the reference SIC001, consisting of siRNA that does not bind any sequence, was used as a control of the transfection. Cells were plated in a 6-well plate at a concentration of 250,000 cells/ml in 2 ml of total volume in DMEM. 24 hours later, 90 pmol of siRNA were prepared in 150 µl DMEM per well. This volume was then added to a mixture containing 9 µl of lipofectamine RNAimax (Thermo Scientific, Waltham, MA, USA) and 150 µl of DMEM per well. After incubation of 5 min, 150 µl were added per well and 72-hours post-incubation, the cells were trypsinized for their use in viability assays (details in section 3.2.3.2).

3.2.4.3. Gene expression by RT-qPCR

Cells were harvested with trypsin for 5 min at 37ºC and centrifuged at 300 *g* for 5 min. Total RNA from cultured cells was extracted using the RNeasy Mini kit (Qiagen, Hilden, Germany) and used immediately for reverse transcriptase reactions or stored at -80ºC until use. RNA concentration was measured in a Nanodrop spectrophotometer (ThermoFisher, Walthman, MA, USA). cDNA synthesis was performed using the RevertAid First Strand cDNA Synthesis kit (Fermentas, Waltham, MA, USA). The reaction was prepared according to the manufacturer's instructions: a mixture containing 1 µg of RNA, 1 µL of random hexamers and Milli-Q -H_2O until 12 µL was incubated at 70 °C for 5 min. Then, 4 µL of buffer 5X, 1 µL ribonuclease inhibitor, 2 µL 10 mM dNTP and 1 µL reverse transcriptase were added to each tube. The reaction mixture was then incubated at 25 °C for 5 min, 42ºC for 1 h and inactivated at 70 °C for 10 min.

The quantitative PCR reaction was performed for *TACSTD2* using the KiCqStart SYBR Green kit (Sigma, St. Louis, MO, USA) according to the manufacturer's instructions. KiCqStart SYBR Green predesigned primers (Sigma, St. Louis, MO, USA) for *TACSTD2* and *BACT* were employed, see Table M1 for details of the sequence. All primers were used in a final concentration of 300 nM, and 10 ng of cDNA were used per well, for a total volume of 10 µL. All cDNA samples were measured in triplicate in a 96-well plate covered with adhesive seals in the thermocycler Roche LightCycler 480 (Roche, Basel, Switzerland).

Reactions started with 10 min at 95ºC, followed by 45 cycles of 15 s at 95 ºC, 1 min at 60ºC and 10 s at 72ºC. The $2^{-\Delta\Delta CT}$ or $2^{-\Delta CT}$ method was used for calculating the normalized mRNA expression. *β-ACTIN* was used as a housekeeping. In the case of $2^{-\Delta\Delta CT}$ analysis, treated samples were also normalized to control samples.

3.2.4.4. Protein expression by flow cytometry

The cells were incubated with anti-TROP2 conjugated to PE antibody (Becton Dickinson, Franklin Lakes, NJ, USA) in a dilution 1:50 and a total volume of 50 µl for 15 min at room temperature in darkness and agitation. Next, the antibody was removed by centrifugation at 400 *g* for 5 min and fixed with Glyo-fixx™ (Thermo Scientific, Waltham, MA, USA) in 1:1 proportion with PBS-BSA 0.5% for a final volume of 50 µl per tube. Cells were finally detected in a FACS Canto cytometer (Becton Dickinson, Franklin Lakes, NJ, USA). All the data was analyzed with FlowJo® program.

3.2.4.5. Protein expression by confocal microscopy

Control cancer cell lines and chronic treated cells were seeded in 9-well microscope slides (Thermo Scientific, Waltham, MA, USA) with poly-L-Lysine in a 1:20 dilution (Sigma, St. Louis, MO, USA). After an incubation of 15 min at RT, the supernatant was removed, and a mix of antibodies was added. Hoescht was used in a dilution 1:1000 to stain the nuclei, anti-TROP2 conjugated to PE (Becton Dickinson, Franklin Lakes, NJ, USA) also used in a 1:50 dilution and pan-cytokeratin and keratin-8 cojugated to APC (Miltenyi Biotec, Bergisch Gladbach, Germany) in a 1:50 dilution were used to stain the cytoplasm of the cells. Microscope slides were incubated for 30 min at RT. Images were acquired in a Stellaris 8 Laser Scanning Confocal Microscope (Leica, Wetzlar, Germany) with Depletion Lasers (STED) at 592, 660, and 775 nm.

3.3. Statistical analysis

The t-Student test was used to compare two independent groups, while the Wilcoxon signed-rank test was applied for paired data, such as in MTT assays where different drug concentrations were tested within the same groups. ANOVA was employed to compare more than two groups, with Tukey's test used to identify significant differences among all groups and Dunnett's test to compare each group against the control. Finally, the Chi-square test was applied to analyze categorical variables. All statistical analyses were performed using GraphPad Prism 8.

4. Results

The purpose of this section is to expose the main results obtained throughout this PhD thesis. As it consists of a multidisciplinary thesis, some explanations are included. We divided the section into two subsections. The first one is focused on developing a mathematical model to predict relapses in patients diagnosed with CRC. For that, we selected thirteen biomarkers from literature. After testing their expression in CRC cell lines, we measured their expression in cells isolated from blood of CRC. The clinicopathological characteristics of all the patients were also registered, which allowed us to study the correlation between variables. We finally developed a mathematical model for relapses.

Once we were able to detect relapses in CRC patients, the aim was to evaluate the resistance to chemotherapeutics. First, we developed tumors derived from patient cells. We used several techniques such as injections in different animal models (nude mice, NSG mice and chicken embryos) and culture directly in plates (Matrigel and in suspension cultures). We also improved the method by depleting CD45+ cells, thereby enriching the CSC population. While we successfully generated two patient-derived tumors, no primary patient CRC cultures are yet available for use in resistance experiments as they require additional time to establish a stable culture. For that reason, we conducted resistance studies using five human CRC cell lines.

For that purpose, we developed a new growth rate-based method and established two new parameters that overcome the disadvantages of the standard IC_{50} method, which is both method and time-dependent. Then, we carried out a preliminary assay to check whether any of the selected biomarkers were candidates to measure resistance and discovered that TROP2 was overexpressed in a more resistant cell line. We broadened this result and studied the correlation between resistance and the expression of this biomarker by the new developed method. Moreover, we established chronic cultures to develop resistant cell lines and evaluate the differences in expression of this biomarker. Finally, we silenced the gene to check differences in resistance. All these results suggest that TROP2 is a good candidate to detect resistance although more experiments will be carried out in the future to confirm this, especially in patient samples.

4.1. Development of a mathematical model to predict relapses in CRC patients

4.1.1. Selection of CSC biomarkers

The aim of this section is to develop a mathematical model which predicts relapses in CRC patients. Several biomarkers have been described in the literature for CSCs, although none of them are exclusive for this population. That is why in this study we propose the use of a combination of biomarkers. First, we conducted a selection of biomarkers from the literature, identifying those that meet the following criteria:

- The expression should be related to CSCs as our purpose is to predict relapses by finding this population of cells in blood from CRC patients that may have evaded from the primary tumor.
- The expression should be related to CRC since we are focused on this type of cancer in this study. The idea is to develop the model and then try to extrapolate to other cancer types, with some modifications that may apply and will be considered in the future.
- They should be membrane proteins as the main technique that we are using to detect the CSC population is flow cytometry.

Considering these requirements, we selected a total of thirteen biomarkers. The gene and protein symbols are shown in Table R1.

Table R1. Gene and protein symbols of the thirteen biomarkers selected for this study.

Gene Symbol	Protein Symbol	Gene Symbol	Protein Symbol
ABCG2	BCRP1	*TACSTD2*	TROP2
PROM1	CD133	*ALCAM*	CD166
PROM1	AC133	*LGR5*	LGR5
CD34	CD34	*DCLK1*	DCLK1
CD44	CD44	*CD36*	CD36
CD44v6	CD44v6	*RAGE*	RAGE
EPCAM	EPCAM		

4.1.2. Expression of CSC biomarkers in CRC cell lines

Then, the expression of all of them was quantified by RT-qPCR in five human CRC cell lines: SW480, SW620, DLD1, HCT116 and HT29. *β-ACTIN* was used as a housekeeping gene to normalize the data. Results were analyzed by the $2^{-\Delta CT}$ method and are shown in Figure R1.

ABCG2 was highly expressed in HT29 cell line, which is consistent with the fact that it is a more resistant cell line. Significant differences were determined between SW480 and HT29 cell lines in relation to *PROM1* expression, being HT29 the one with more expression. In relation to *CD34* expression, it was lowly expressed in all the cell lines, although significant differences were detected in SW480 cell line when compared to the rest.

On the contrary, *CD44* was highly expressed in all the cell lines, HCT116 and HT29 being the ones with more expression. A similar pattern was found in *CD44V6,* a variant of the previous one, although less expressed.

Moreover, *EPCAM* and *ALCAM* were both expressed in the five cell lines although *EPCAM* expression was higher, and no significant differences were found between the cell lines in both cases. Regarding *TACSTD2*, the SW620 cell line did not express this gene. The rest of the cell lines expressed it, with HT29 showing higher expression. Further analysis of this will be carried out in section 4.2.3 as a relation to resistance could be considered.

In relation to *LGR5*, it was highly expressed in the SW620 and DLD1 cell lines. Additionally, *DCLK1* was highly expressed in the SW480 cell line when compared to the rest of the cell lines. Finally, *CD36* was more expressed in the SW480, HCT116 and HT29 cell lines.

Based on these results, we decided to further consider all of them for immunodetection in blood samples from patients previously diagnosed with CRC.

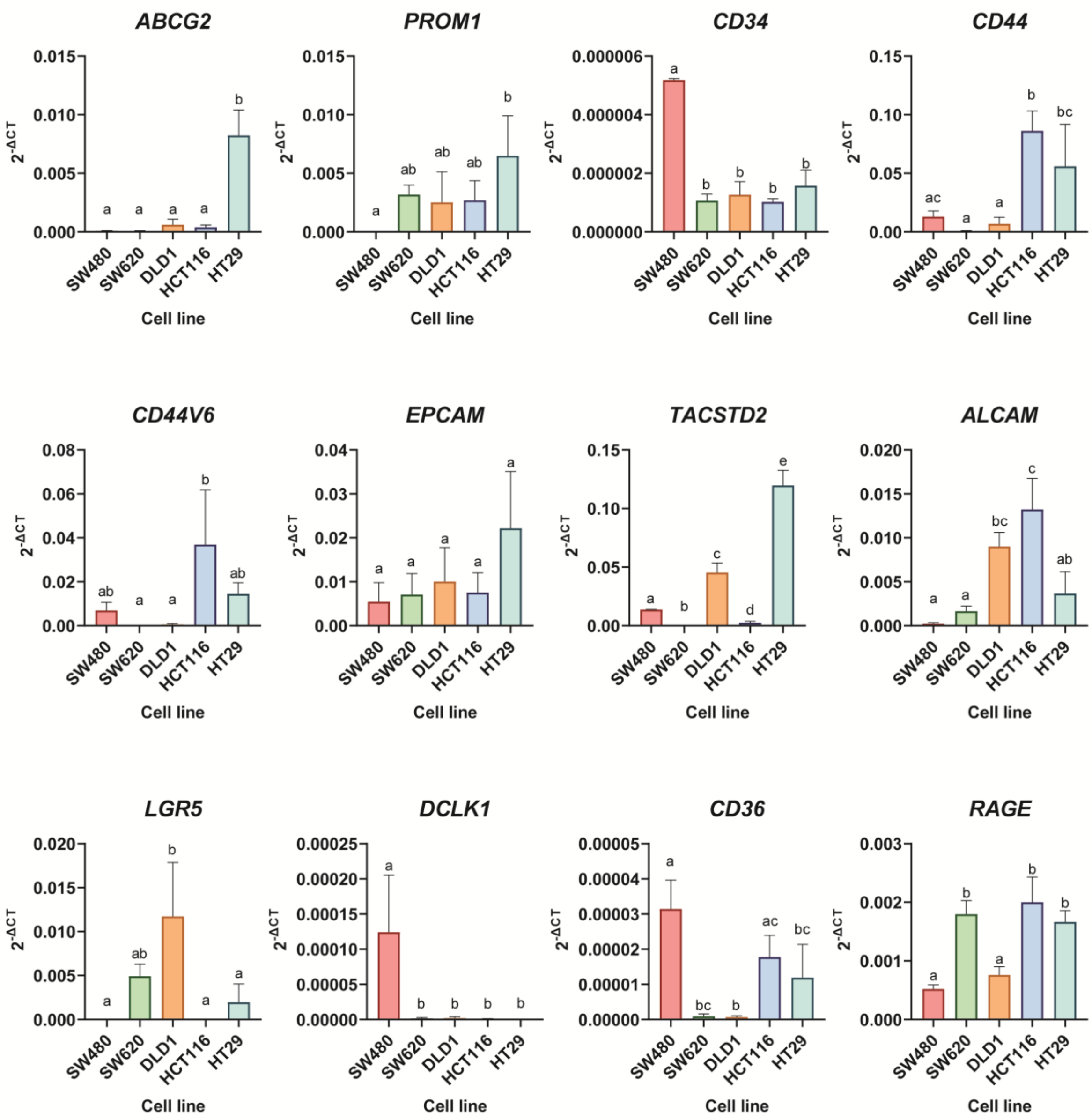

Figure R1. Gene expression of CSC biomarkers in five human CRC cell lines. *ABCG2, PROM1, CD34, CD44, CD44V6, EPCAM, TACSTD2, ALCAM, LGR5, DCLK1, CD36* and *RAGE* gene expression were measured in SW480, SW620, DLD1, HCT116 and HT29 CRC cell lines. RNA was extracted from cell lines and retrotranscribed prior to performing a qPCR. *β-ACTIN* was used as a housekeeping gene. Results were analyzed by the $2^{-\Delta CT}$ method. One-Way ANOVA and Tukey´s multiple comparison test was carried out to evaluate differences between cell lines for each biomarker. Different letters indicate statistically significant differences between groups (p-value < 0.05). N= 3.

4.1.3. Expression of CSC biomarkers in patients previously diagnosed with CRC

A schematic workflow of this section is shown in Figure R2. Briefly, after the selection of thirteen biomarkers, we proceeded to immunodetect CSCs in blood from patients previously diagnosed with CRC. As this study includes the use of human samples, we first obtained approval from the ethics committee. It was approved in all four hospitals and the health center involved in this study, as well as by the Universidad Politécnica de Madrid, as it is the center where the samples were processed (Annex A1). Different centers' inclusion in the project was not simultaneous but progressive. Then, we selected patients to be included in the project in collaboration with the doctors from each center. The patients were required to fulfill the inclusion criteria and not meet any of the exclusion criteria (for more information, see the Materials and Methods section – 3.1.2.2).

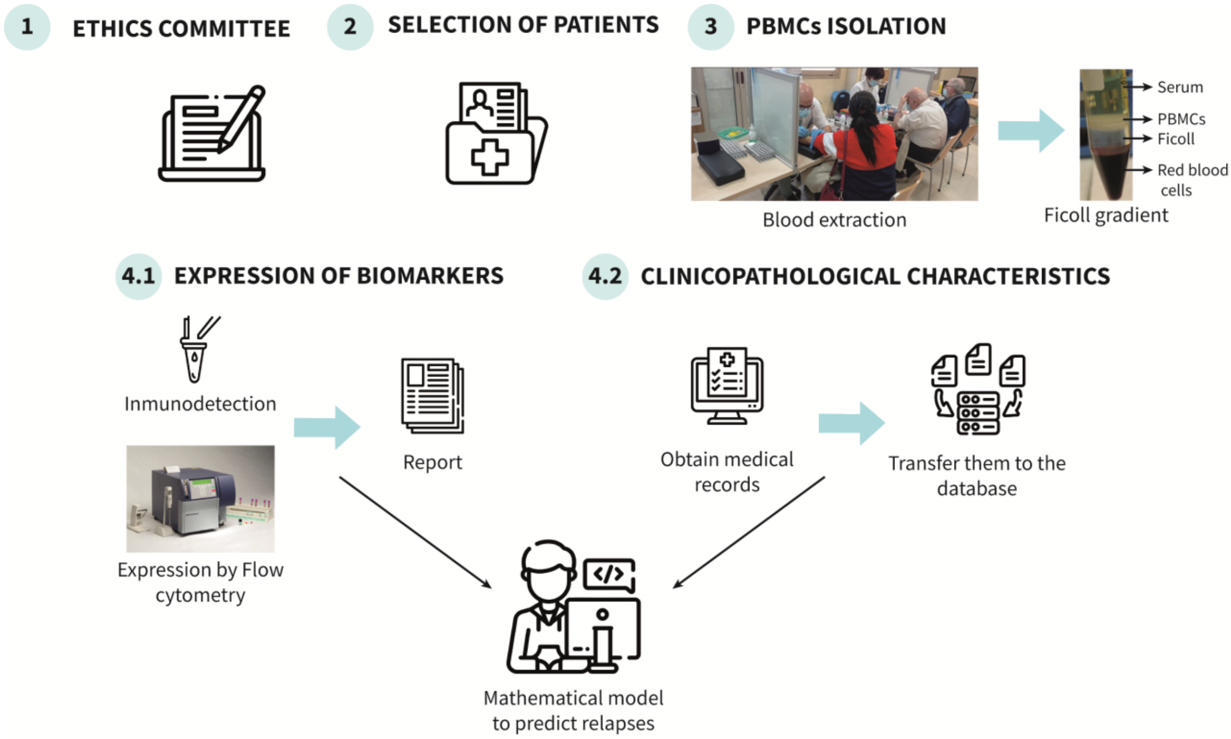

Figure R2. Schematic workflow to develop a mathematical model to predict relapses in CRC. After the ethics committee approval and the selection of patients, PBMCs were isolated by a Ficoll® gradient. The data from the expression of biomarkers together with the clinicopathological characteristics were considered to generate a mathematical model to predict relapses in CRC.

Once the ethics committees were approved and after selecting the patients, we scheduled appointments for them and extracted a 4 ml EDTA blood tube. PBMCs were isolated after a Ficoll® gradient and further inmunostained with antibodies for the thirteen biomarkers previously selected. An unstained control was also carried out. The expression of all the biomarkers was measured by flow cytometry and the results were processed by a Python script that was developed for this purpose. A database was also developed to register all the clinicopathological characteristics for each patient and variations between samples. Together, the expression of biomarkers and clinicopathological characteristics were considered to develop the mathematical model to predict relapses in CRC patients.

4.1.3.1. Detailed information about the patients included in the study

First, we are going to detail some information about the clinicopathological characteristics of the patients included in the study (Figure R3).

Regarding the age, the median age across all the centers was 71 years old, the youngest patient being 36 years old and the oldest one 92 years old, both from the DHC. Most of the population was above 50 years old, with only 7 patients below this age.

In relation to the BMI, the median was 26.9 kg/m², the patient with the lowest BMI had 16.4 kg/m², and the patient with the highest BMI had 46.9 kg/m². In the Spanish population, people are considered underweight at a BMI of less than 18.5, normal weight between 18.5 kg/m² and 25 kg/m², overweight between 25 kg/m² and 30 kg/m² and obesity from 30 kg/m². Considering these parameters, 3.7% of the population of study was underweight, 32.8% was of normal weight, 37% was overweight and 26.5% were obese.

Concerning the sex, 57.3% of the patients were male and 42.7% were female, so it was almost balanced and representative of the CRC patient's population. However, some differences were detected between centers. The JDH population were 56% male and 44% female, in the RCH they were almost all male (91% male versus 9% female), in CRH they were 65% male and 35% female, in TUH they were 48% male and 52% female and finally in DHC they were 51% male and 49% female, respectively.

Regarding the family background, 40% of the population of study presented relatives with cancer and 45% did not. The family background was unknown for the remaining 15%. So, the population was also balanced in this regard. However, some differences were detected between centers, as the relatives with cancer were higher in the JDH (50% vs 12.5%), RCH (54.5 vs 9.1%) and TUH (34.8 vs 13%) and lower in the CRH (35.2 vs 54.7%) and DHC (40.9 vs 55.7%).

We also registered if the patients were active smokers, exposed smokers (smokers in the past) or never smokers. Considering all the centers together, 45.8% had never smoked, 37.5% had previously smoked but are no longer currently smoking (exposed) and 8.3% were active smokers. We did not have data for the remaining 8.3%. This indicates that the population under study was smoke-free, as most of them had never smoked or decided to quit smoking. However, some differences were detected between centers, as the exposed smokers were higher than never smokers in the RCH (63.6% vs 9.1%) and TUH (39.1 vs 17.4%), similar in the CRH (44.4 vs 37%) and DHC (40.9 vs 55.7%) and lower in the JDH (31.3 vs 62.5%).

In relation to the hospital status of the patient, we classified it into three main groups: *in treatment, follow-up* or *discharge*. When considering all the centers, 34.6% of patients were in treatment, 58.6% of patients were in follow-up and 6.8% of patients were discharged. The proportion of follow-up patients was higher than in treatment patients in all the centers except for CRH where 77.8% of patients were in treatment and 22.2% were in follow-up. Note that all RCH patients and the majority of DHC patients were in follow-up.

As described in the clinical guides for diagnosis of CRC, the most widely used diagnosis technique in all the centers was colonoscopy (80%), followed in 8.3% of the cases by other techniques such as CT. As expected, the localization of the primary tumor was detected in a different location depending on the center because it depends on the patient. In the JDH, 56.3% of tumors were located in the descending colon, followed by 25% in the ascending colon and 18.8% in the rectum. In the case of RCH, 36.4% of tumors were located in the ascending colon followed by 27.3% both in the transverse colon and the rectum and 9.1% in the descending colon. CRH presented a 42.6% of tumors in the rectum followed by 27.8% of tumors in the descending colon, 13% in the ascending colon and 3.7% in the transverse colon. In the TUH, 52.2% were located in the rectum and 4.3% in the transverse colon. Finally, 26.1% of tumors were located in the rectum, 25% in the ascending colon, 9.1% in the descending colon and 1.1% in the transverse colon in the DHC.

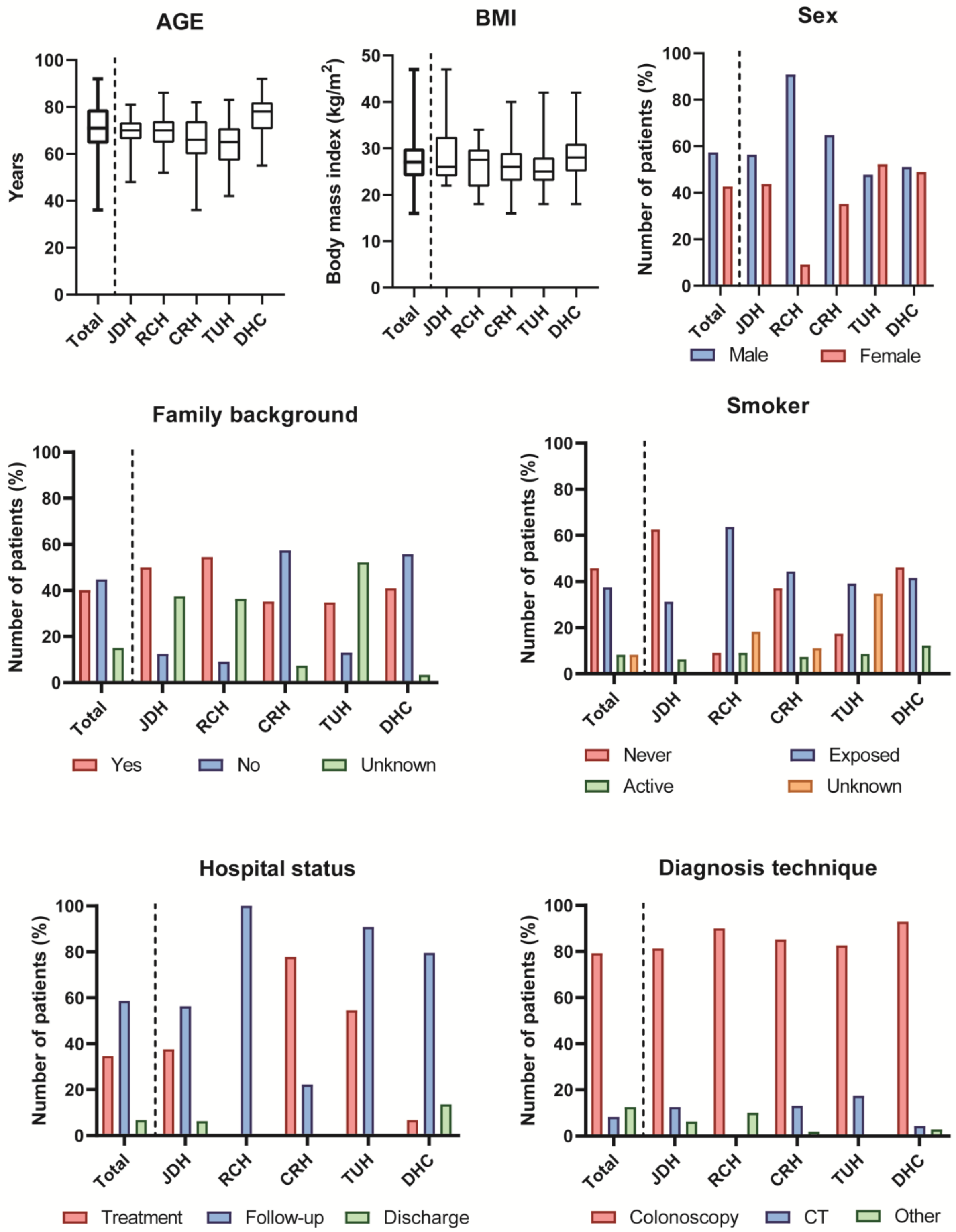

Figure R3a. Clinicopathological characteristics of the patients included in the study. Age and BMI were represented as boxplots. Sex, Family Background, Smoker, Hospital status and Diagnosis technique were represented as bar plots. Data is shown for each center and the total represents the mean of the data from all centers.

Moreover, the stage of the patients included was almost equal (approximately 40%) for N0M0 and N1M0 cohorts, not presenting affected lymph nodes or with affected lymph nodes, respectively. Cancer had not spread to distant organs in either case. However, 20% of patients presented metastasis at diagnosis (M1 cohort). These latter are those in which progression is more probable as CTCs are likely to evade the tumor, be present in blood and disseminate to other organs. The distribution of patients among these cohorts was different in each center, with CRH and TUH having a greater proportion of M1 patients, of 40.7% and 27.3%, respectively. Additionally, most patients presented large tumors at diagnosis, with 60.3% classified as stage T3 and 15.6% as stage T4.

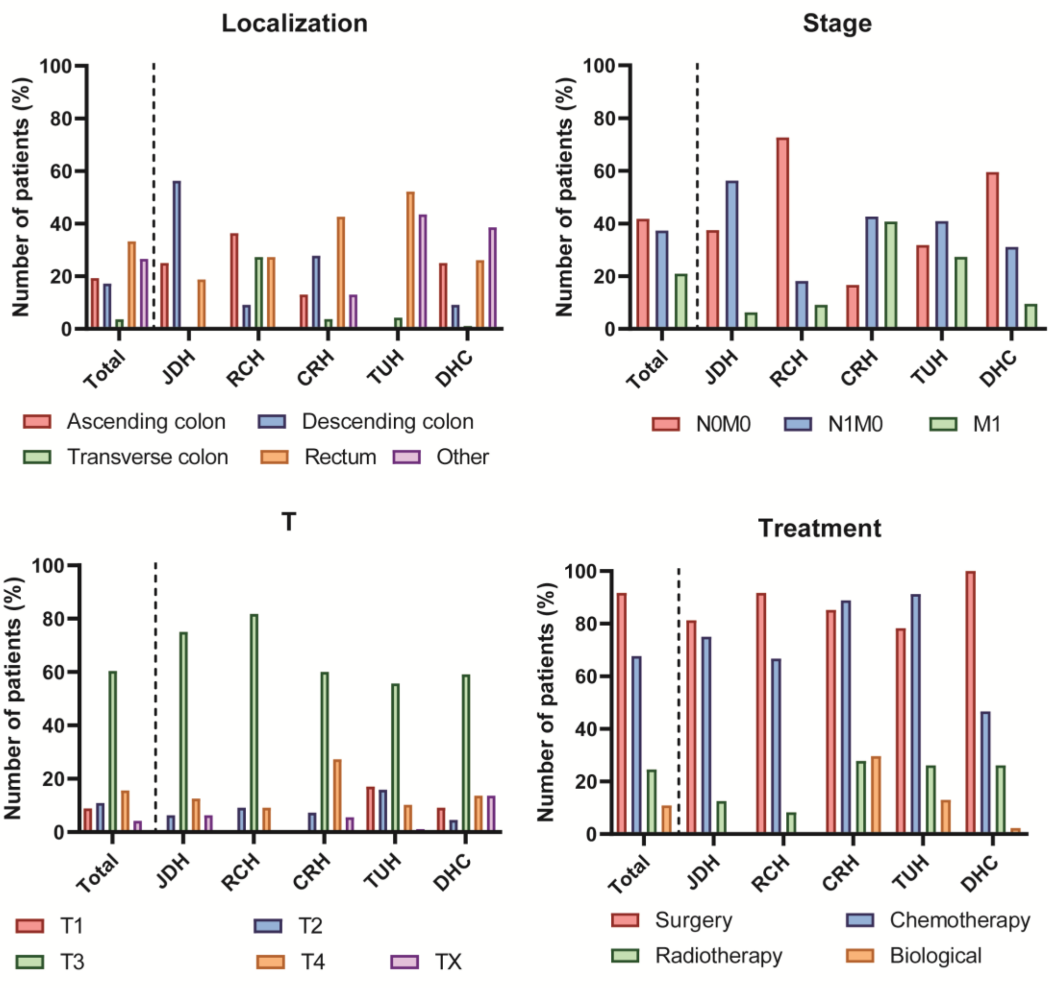

Figure R3b. Clinicopathological characteristics of the patients included in the study. Localization of the tumor, Treatment and TNM stage were represented as bar plots. Data is shown for each center and the total represents the mean of the data from all centers.

Finally, surgery was the most frequent treatment (91.7% of patients, almost all of them) followed in most cases by chemotherapy (67.7% of patients). In some cases, radiotherapy (24.5%) or biological (10.9%) treatment was added. This pattern was maintained for all the centers although biological treatment was most frequently provided in CRH (29.6%).

4.1.3.2. Expression of CSC biomarkers in PBMCs

To assess the expression of the selected CSC biomarkers, 4 ml of blood was extracted in EDTA tubes from all patients included in the study at various stages of disease progression. Moreover, blood from healthy people was also extracted as a control for the experiment. The percentage of total single cells (PBMCs) that expressed each biomarker is shown in Figure R4, both for the control (healthy subjects) and for the patient (previously diagnosed with CRC) samples.

The distribution of the percentage of positive cells presented a similar pattern in almost all the biomarkers, being zero for control samples and near zero but with a small number of samples with higher expression for patient samples. With one exception, as the CD44 biomarker was expressed in 100% of cells from control samples and the expression was distributed from 0 to 100% in the patient samples. Note that there are patients whose expressions of BCRP1, CD133 and CD34 varied between 0 and 100% but the patient with the maximum expression of AC133, EPCAM, CD166 and RAGE was around 60%. Moreover, the expression levels of CD44V6 and CD36 in patient samples ranged from 0 to 100%, with no samples clustering at 0%. Finally, the expression levels of LGR5 and DCLK1 were elevated in patient samples, with some patients expressing this protein in the 100% of cells, especially for DCLK1. However, the expression of DCLK1 was also higher in control samples when compared to other biomarkers, reaching a maximum of 40%.

The results show that all the biomarkers except DCLK1 and CD44 were not expressed in controls, but expression was detected in some samples from patients. These results demonstrate that these biomarkers are useful for the project, as the samples with higher expression of biomarkers could be the ones with less prognosis and with more probability of relapse. If this is the case, the model will determine whether the levels of these biomarkers predict relapses. Moreover, note that the expression of TROP2 is rather modest in patients when compared to control samples, which will be important for its resistance role explained in section 4.2.3.

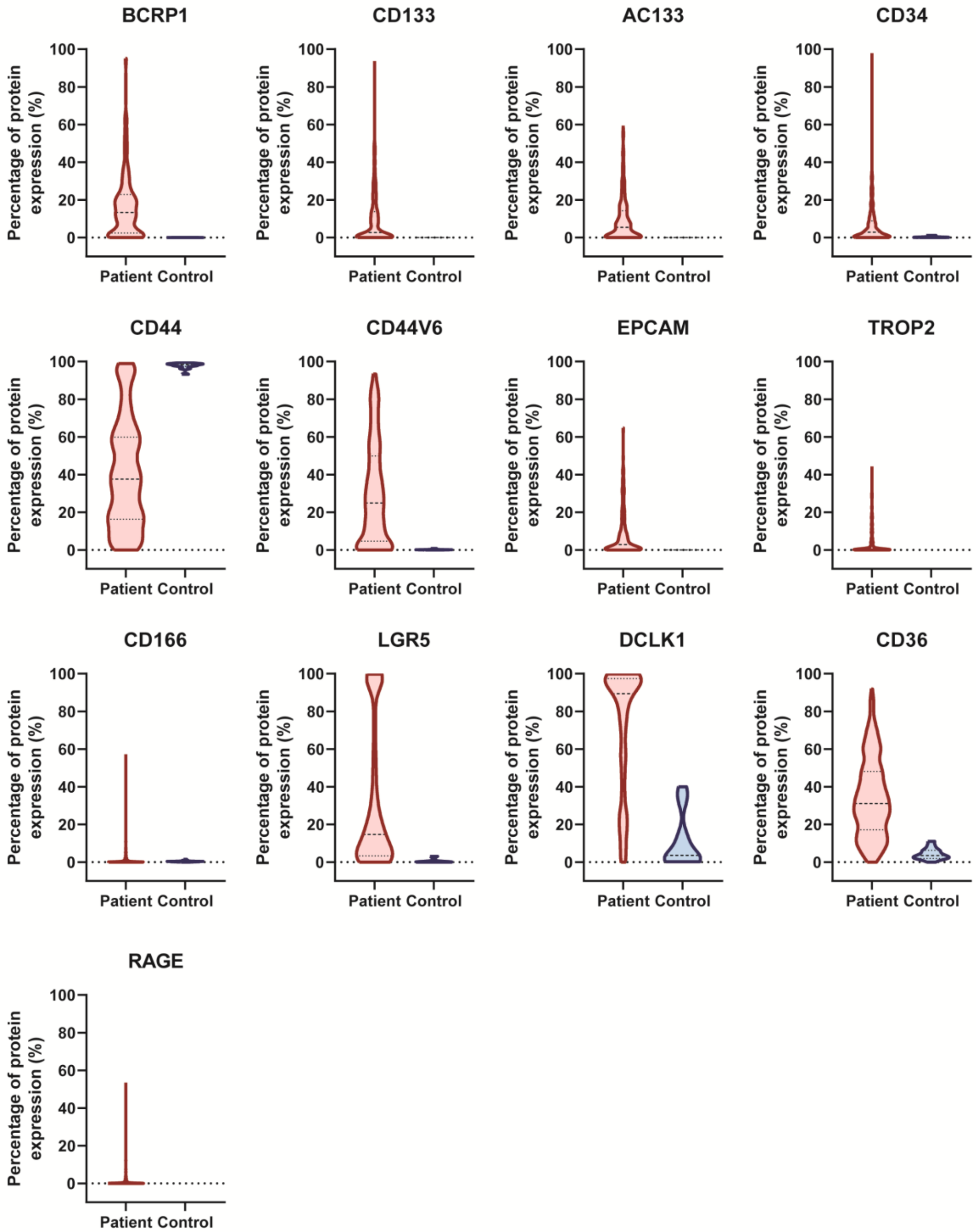

Figure R4. Expression of CSC biomarkers in PBMCs. The expressions of BCRP1, CD133, AC133, CD34, CD44, CD44V6, EPCAM, TROP2, CD166, LGR5, DCLK1, CD36 and RAGE were measured by flow cytometry in PBMCs isolated from blood samples by Ficoll® gradient. Expression is shown as the percentage of total single cells both in control (healthy subjects, in blue) and patient (previously diagnosed with CRC, in red) samples. N=192 patient samples and N=25 healthy control samples.

Before going to the development of a mathematical model for relapse prediction, we studied the correlation between variables, both clinicopathological characteristics and the biomarkers' expression.

4.1.3.3. Correlation analysis of the variables studied in this project

To analyze the correlation between variables, we first developed Python scripts to process all the clinicopathological characteristics and biomarkers' expression (for more information, see the Materials and Methods section – 3.1.3). Once the data were processed, we first studied the correlation between relapse (True/False) and the clinicopathological characteristics that do not vary between samples. We considered the variable *relapse* as True if the patient relapsed at any time from diagnosis until the end of the study period for the project. In Figure R5, each variable was represented for patients that relapsed (True, in red) and did not relapse (False, in blue).

We started evaluating personal characteristics such as sex, BMI, family background and smoking condition. In relation to sex, differences in relapse were not detected between males and females (41.8% of males and 36.6% of females relapsed). The BMI was also not different when comparing patients that relapsed and did not relapse (the median for patients that relapsed was 26.2 kg/m^2 and 27.9 kg/m^2 for the patients that did not relapse). Significant differences were not detected when evaluating the family background or smoking conditions. Patients without familiar backgrounds relapsed in 39.5% of patients and the ones with familiar backgrounds in 35.10% of the cases. Although there were no significant differences, a slight increase in relapse was observed in active smoker patients (50%) followed by exposed smokers in 38.9% of patients (patients who smoked in the past, but they are no longer smokers) and never smokers (33%).

We also evaluated characteristics of the tumor at diagnoses such as stage, localization and differentiation. Significant differences were detected in all of them between patients that relapsed or did not relapse. In relation to the stage, both in the stage I to IV and TNM, the patients who were less likely to relapse were the stage I and II or cohort N0M0 (20%, 19% and 23% respectively), which correspond to patients with more prognosis as lymph nodes nor distal disease were detected. Then, stage III patients or N1M0 patients presented a higher probability of relapse (39% and 37.9%, respectively). These patients present lymph nodes affected but disease has not spread to distal organs. Finally, patients with distal disease at diagnosis, stage IV and M1 cohort, presented a higher rate of relapses (83.3% and

83.8% respectively). The rate of relapses was also different depending on the location of the primary tumor, being higher for transverse colon (57.1% of patients with primary tumor located at the transverse colon relapsed), followed by rectum (50%), descending colon (39.4%), ascending colon (27%) and sigma (18.5%). Differentiation of the tumor was also studied. Patients with moderately differentiated tumors relapsed in a lower proportion (28.3%) when compared to poorly differentiated tumors (66.7%). However, patients with well differentiated tumors relapsed in a higher proportion (53.7%) when compared to moderately differentiated tumors (28.3%).

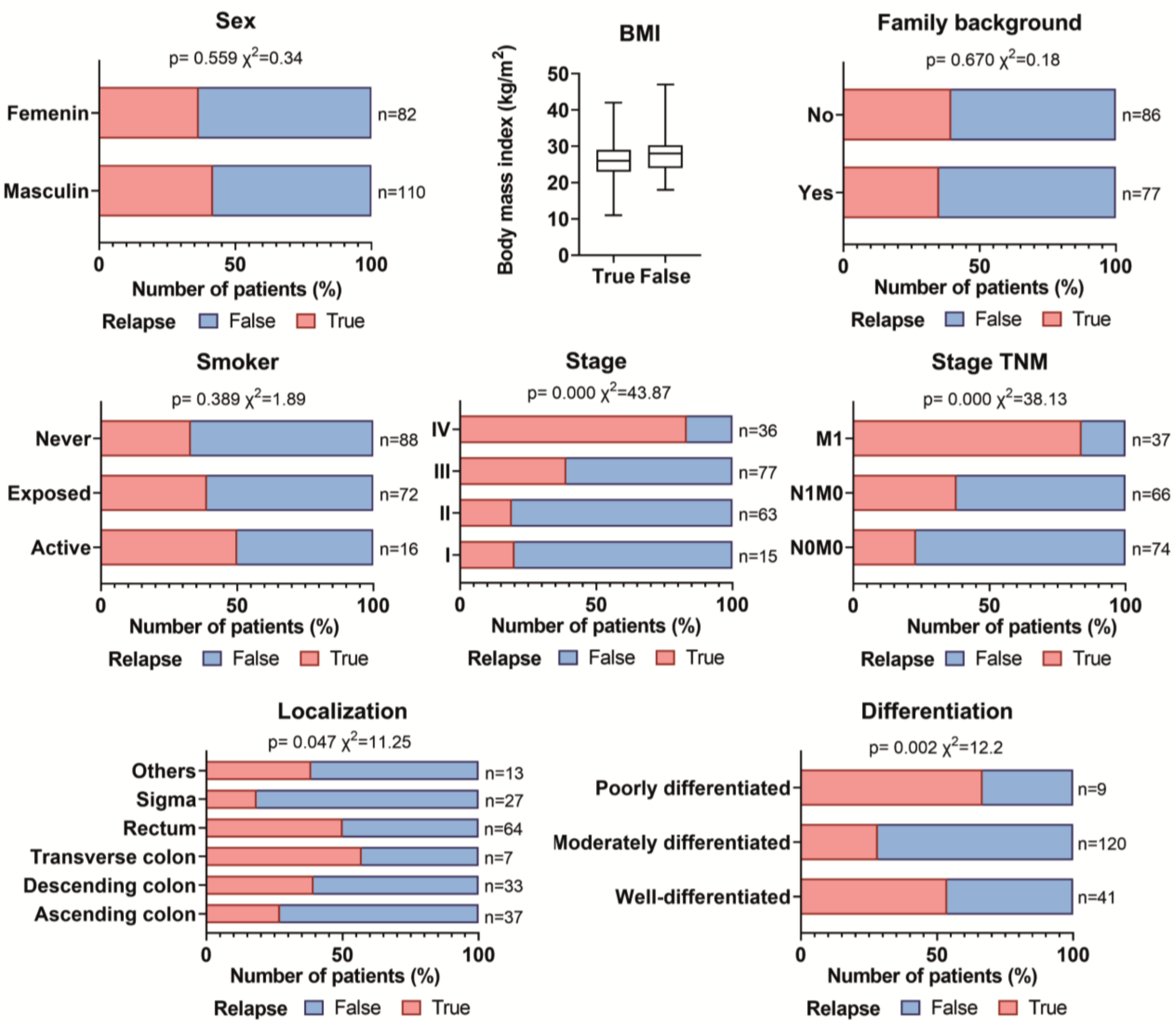

Figure R5. Correlation between the clinicopathological characteristics from patients and relapse variable. The percentage of patients for each condition was calculated and compared by a Chi Square test (p-values are shown in the figure). Each variable was represented for patients that relapsed (True, in red) and did not relapse (False, in blue). Relapse means patients who relapse at any time from diagnoses until the end of the study period for the project.

Until now, we found no correlations in our study between the sex, BMI, smoking conditions and family background. However, correlation between the stage, localization and differentiation of the tumor at diagnosis and if the patient relapsed was found. The probability of relapse was higher when the tumor was detected in a more advanced stage. Moreover, the rectum and the transverse colon were the locations with higher probability of relapse in our population of study. Finally, the poorly differentiated tumors were the ones with higher probability of relapse.

Then, we studied the correlation between the biomarkers' expression and the clinicopathological characteristics that may have changed between samples. This consists of treatment-related variables. First, we checked whether the type of treatment received until the date of the sample influenced in the biomarkers' expression. For that, we classified samples in three groups: *Sx alone*, *Sx + other treatment* and *other treatment without Sx*. Other treatment usually refers to chemotherapy, but radiotherapy and biological treatments were also included.

Supervised ML was used to check whether the percentage of biomarkers was able to predict which of these groups corresponds to the sample. SVM, kNN, Logistic Regression, Random Forest, Gradient Boosting and Tree models were employed. The scores for evaluating the models (including AUC, CA, F1, Pec, Recall and Spec), the confusion matrix for each model and the ROC curve, are shown in Figure R6.

Regarding the scores for evaluating the models, all of them were around 0.5 points, indicating that the models could not predict the type of treatment. Moreover, the confusion matrices revealed low predictive capacity, as the percentages along the diagonal were far from 100% (higher values on the diagonal indicate better model accuracy and, therefore, greater predictive capacity).

The ROC curve further confirms the lack of predictive ability, as the values are close to the diagonal reference curve (the closer to this line, the lower the predictive capacity). Based on these results, we conclude that the type of treatment received prior to sample collection was not correlated with the expression of the biomarkers.

A

Model	AUC	CA	F1	Prec	Spec	Recall
SVM	0.661	0.551	0.548	0.561	0.551	0.607
kNN	0.662	0.588	0.576	0.574	0.588	0.571
Logistic Regression	0.685	0.591	0.580	0.628	0.591	0.593
Random Forest	0.695	0.601	0.588	0.585	0.601	0.582
Gradient Boosting	0.678	0.606	0.582	0.576	0.606	0.563
Tree	0.611	0.556	0.547	0.543	0.556	0.559

B

SVM (Actual \ Predicted)	SX+OTHER	SX ALONE	OTHER W/O SX	N
SX+OTHER	59.1%	38.0%	3.0%	237
SX ALONE	51.2%	48.8%	3.9%	129
OTHER W/O SX	66.7%	23.3%	6.7%	30
N	226	160	10	396

kNN (Actual \ Predicted)	SX+OTHER	SX ALONE	OTHER W/O SX	N
SX+OTHER	65.8%	30.4%	3.8%	237
SX ALONE	62.0%	33.3%	4.7%	129
OTHER W/O SX	86.7%	3.3%	10.0%	30
N	261	116	18	396

Log Reg (Actual \ Predicted)	SX+OTHER	SX ALONE	OTHER W/O SX	N
SX+OTHER	72.6%	16.5%	11.0%	237
SX ALONE	69.8%	21.7%	8.5%	129
OTHER W/O SX	56.7%	0.0%	43.3%	30
N	276	67	50	396

Random Forest (Actual \ Predicted)	SX+OTHER	SX ALONE	OTHER W/O SX	N
SX+OTHER	71.7%	25.3%	3.0%	237
SX ALONE	65.1%	31.0%	3.9%	129
OTHER W/O SX	90.0%	3.3%	6.7%	30
N	281	101	14	396

GB (Actual \ Predicted)	SX+OTHER	SX ALONE	OTHER W/O SX	N
SX+OTHER	57.4%	38.4%	4.2%	237
SX ALONE	40.3%	59.7%	0.0%	129
OTHER W/O SX	70.0%	13.3%	16.7%	30
N	209	172	15	396

Tree (Actual \ Predicted)	SX+OTHER	SX ALONE	OTHER W/O SX	N
SX+OTHER	62.9%	32.1%	5.1%	237
SX ALONE	47.3%	47.3%	5.4%	129
OTHER W/O SX	56.7%	33.3%	10.0%	30
N	227	147	22	396

C

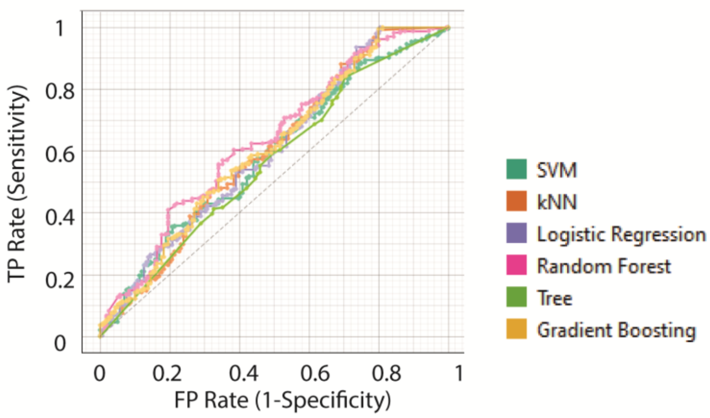

Figure R6. Correlation between biomarker expression and the type of treatment received before sample extraction analyzed by Supervised ML. SVM, kNN, Logistic Regression, Random Forest, Gradient Boosting and Tree models were used. The scores (A), the confusion matrix (B) and the ROC curve (C) for all these models are shown. AUC: Area under the curve, CA: classification accuracy, F1: F1 score, Prec: precision, Recall, Spec: specificity.

We also tried to cluster the samples depending on the biomarkers' expression and check whether the clusters generated were different in terms of the type of treatment received. Unsupervised ML was used for this purpose. K-means and Hierarchical clustering methods were employed, classifying the samples in two clusters. The percentage of samples which correspond to each cluster for each type of treatment was represented in Figure R7A. We found that the clusters did not separate samples depending on the type of treatment previously received. Although almost all the samples for other treatments without surgery were in cluster 2, there were a lot of samples defined in this cluster for the other two categories. This analysis also confirms that the type of treatment received before sample extraction did not influence biomarkers' expression.

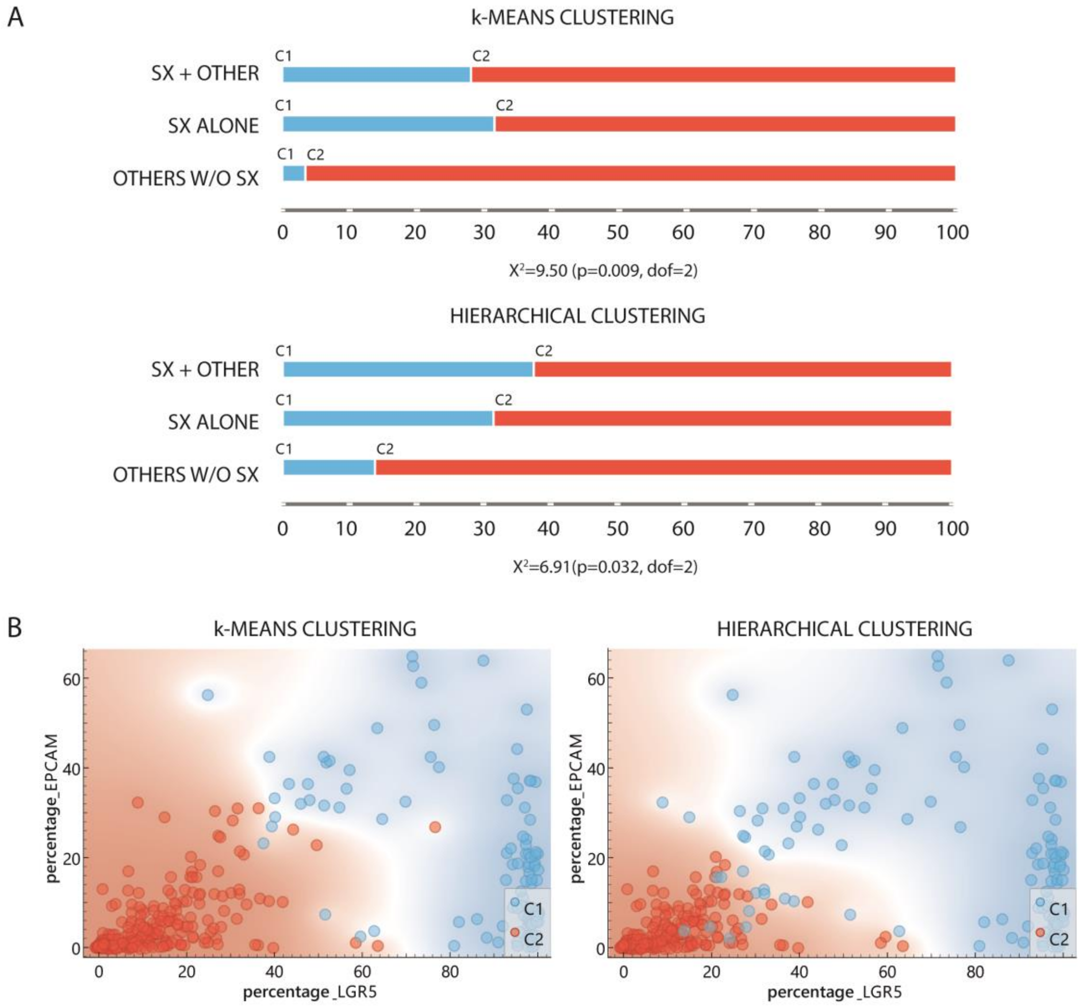

Figure R7. Correlation between biomarker expression and the type of treatment received before sample extraction analyzed by Unsupervised ML. Unsupervised ML was used to generate clusters depending on the expression of biomarkers. (A) The percentage of samples which correspond to each cluster for each type of treatment. (B) Difference in LGR5 expression among clusters.

However, Cluster 1 could be described as the one which corresponds to the samples that received surgery, especially when using the k-means method. We broadened on that and represented different biomarker expressions combinations. The aim is to determine if any biomarker was responsible for the classification of samples in cluster 1. Each dot corresponds to a sample, each cluster is colored different (C1 in blue, C2 in red) and the background is colored based on cluster distribution. As an example, in Figure R7B, a blue background was observed on the right side of the plot. This means that LGR5 expressions (X-axis) higher than 40% defined cluster 1 (blue). Although further analysis is required, this result suggests a possible influence of surgery on high LGR5 expressions, as cluster 1 was the one which includes the samples that received surgery.

We also evaluated whether there were differences in the percentage of biomarker expression between samples that were undergoing treatment at the time of sample collection and those that were not. The use of these biomarkers could be useful to evaluate the response to treatment, generally being lower when the response is good. For that purpose, we classified samples as True or False for a variable called *Treatment*. Supervised ML was used to determine whether the percentage of biomarkers could predict if the patient was undergoing treatment (True) or not (False) at the time of sample collection. SVM, kNN, Logistic Regression, Random Forest, Gradient Boosting and Tree models were employed. The scores for evaluating the models (including AUC, CA, F1, Pec, Recall and Spec), the confusion matrix for each model and the ROC curve, are shown in Figure R8. SVM and Logistic Regression were the ones that best predicted if the patient was undergoing treatment at the time of sample collection, with an AUC of 0.786 and 0.792, a sensitivity (Recall) of 0.784 and 0.712 and a specificity of 0.715 and 0.787, respectively. The confusion matrices also confirmed this result. Note a 67-83% success for both models in the confusion matrices (see the diagonal values, as higher values on the diagonal indicate better model accuracy and, therefore, greater predictive capacity). The ROC curves further confirmed the predictive ability, as the values are far from the diagonal reference curve (the closer to this line, the lower the predictive capacity. The nearer to the upper left part of the graph, the higher the predictive capacity).

Based on these results, we conclude that whether a patient is receiving treatment at the time of sample collection influences the percentage of expression of the biomarkers. That is why we considered the inclusion of this variable in the model of relapse (section 4.1.3.4), to check whether it improves or not.

A

Model	AUC	CA	F1	Prec	Recall	Spec
SVM	0.786	0.784	0.790	0.798	0.784	0.715
kNN	0.713	0.784	0.779	0.776	0.784	0.617
Logistic Regression	0.792	0.712	0.729	0.800	0.712	0.787
Random Forest	0.775	0.787	0.772	0.772	0.787	0.557
Gradient Boosting	0.781	0.802	0.790	0. 790	0.802	0.593
Tree	0.715	0.767	0.767	0.766	0.767	0.629

B

	SVM	Predicted False	Predicted True	N
Actual	False	82.3%	17.7%	294
Actual	True	32.4%	67.6%	105
	N	276	123	399

	kNN	Predicted False	Predicted True	N
Actual	False	87.8%	12.2%	294
Actual	True	47.6%	52.4%	105
	N	308	91	399

	Log Reg	Predicted False	Predicted True	N
Actual	False	67.0%	33.0%	294
Actual	True	17.1%	82.9%	105
	N	215	184	399

	RF	Predicted False	Predicted True	N
Actual	False	91.5%	8.5%	294
Actual	True	57.1%	42.9%	105
	N	329	70	399

	GB	Predicted False	Predicted True	N
Actual	False	91.8%	8.2%	294
Actual	True	52.4%	47.6%	105
	N	325	74	399

	Tree	Predicted False	Predicted True	N
Actual	False	84.4%	15.6%	294
Actual	True	44.8%	55.2%	105
	N	295	104	399

C

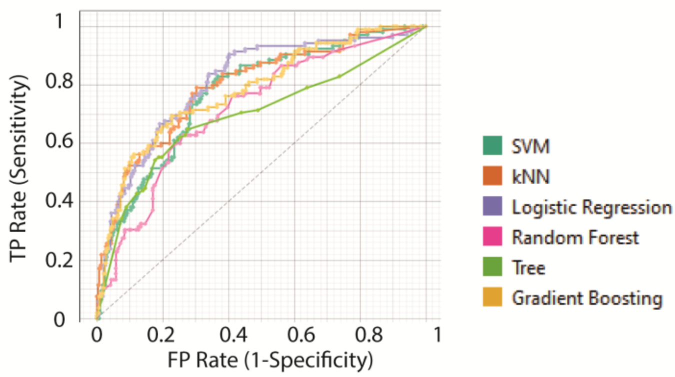

Figure R8. Correlation between biomarker expression and whether the patient was undergoing treatment at the time of sample collection analyzed by Supervised ML. SVM, kNN, Logistic Regression, Random Forest, Gradient Boosting and Tree models were used. The scores (A), the confusion matrix (B) and the ROC curve (C) for all these models are shown. AUC: Area under the curve, CA: classification accuracy, F1: F1 score, Prec: precision, Recall, Spec: specificity.

We also evaluated whether the clusters formed based on the biomarkers' expression (explained in the previous analysis regarding type of treatment) were different in terms of whether the patient was in treatment at time of sample collection.

The percentage of samples which correspond to each cluster for samples from patients undergoing treatment (True) and those not undergoing treatment at the time of sample collection (False) was represented in Figure R9. We found that the clusters did not distinguish samples based on treatment status, as samples from both categories were present in both clusters. Therefore, this analysis did not provide additional insights.

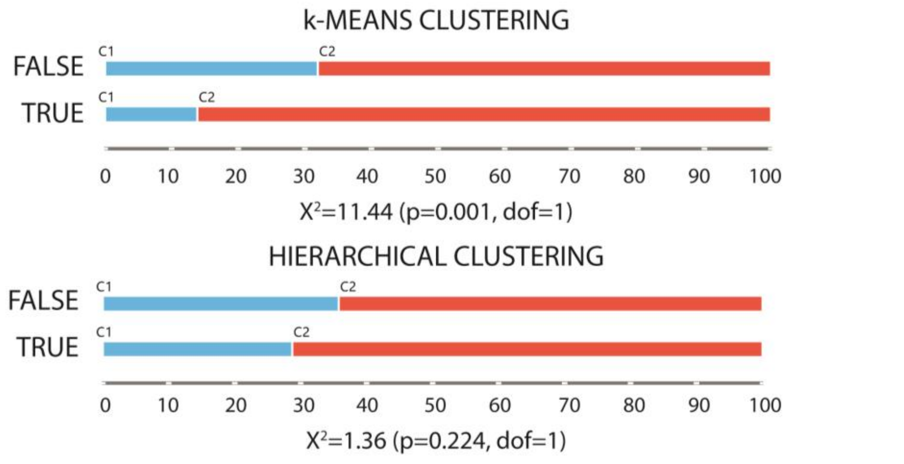

Figure R9. Correlation between biomarker expression and whether the patient was undergoing treatment at the time of sample collection analyzed by Unsupervised ML. Unsupervised ML was used to generate clusters depending on the expression of biomarkers. The percentage of samples which correspond to each cluster for each condition: samples from patients undergoing treatment (True) and those not undergoing treatment (False).

Finally, we calculated the difference in expression for each biomarker between each sample and the previous one. We also included a variable called *In_treatment_from_previous_sample* and checked whether any treatment was received for that patient between these samples. We assigned True if there were treatment in between and False if there were not. Then, Supervised ML was used to check whether the difference in percentage of biomarkers was able to predict if any treatment had been received between the sample and the previous one. The scores for evaluating the models (including AUC, CA, F1, Pec, Recall and Spec), the confusion matrix for each model and the ROC curve, are shown in Figure R10. Regarding the scores for evaluating the models, the sensitivity (recall) was high in some models, but the specificity was low. Moreover, the confusion matrices revealed a low predictive capacity for the True condition, as the percentage of true positives was low. The ROC curve further confirms the lack of predictive ability, as the values are close to the diagonal reference curve (the closer to this line, the lower the predictive capacity).

A

Model	AUC	CA	F1	Prec	Spec	Recall
SVM	0.494	0.782	0.787	0.791	0.782	0.350
kNN	0.604	0.833	0.811	0.797	0.833	0.274
Logistic Regression	0.558	0.620	0.676	0.772	0.620	0.436
Random Forest	0.611	0.755	0.77	0.788	0.755	0.373
Gradient Boosting	0.506	0.764	0.762	0.761	0.764	0.235
Tree	0.611	0.759	0.765	0.772	0.759	0.290

B

		Predicted		
	SVM	False	True	N
Actual	False	86.6%	13.4%	186
	True	73.3%	26.7%	30
	N	183	33	216

		Predicted		
	kNN	False	True	N
Actual	False	94.1%	5.9%	186
	True	83.3%	16.7%	30
	N	200	16	216

		Predicted		
	Log Reg	False	True	N
Actual	False	65.6%	34.4%	186
	True	60.0%	40.0%	30
	N	140	76	216

		Predicted		
	RF	False	True	N
Actual	False	88.7%	11.3%	186
	True	66.6%	33.3%	30
	N	185	31	216

		Predicted		
	GB	False	True	N
Actual	False	86.6%	13.4%	186
	True	86.7%	13.3%	30
	N	187	29	216

		Predicted		
	Tree	False	True	N
Actual	False	84.9%	15.1%	186
	True	80.0%	20.0%	30
	N	182	34	216

C

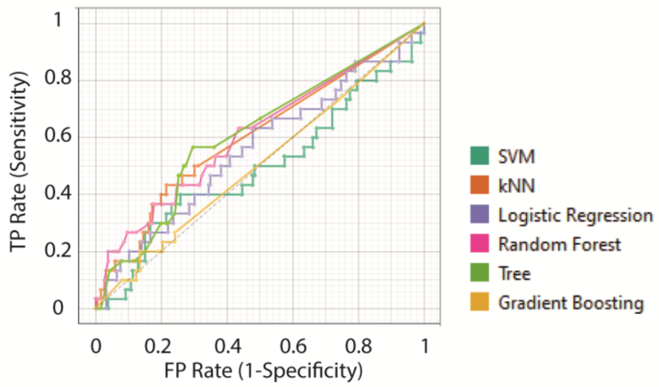

Figure R10. Correlation between the difference in biomarker expression and if any treatment was received between a sample and the previous one by Supervised ML. Supervised ML was used to predict the type of treatment with the biomarkers' expression. SVM, kNN, Logistic Regression, Random Forest, Gradient Boosting and Tree models were used. The scores (A), the confusion matrix (B) and the ROC curve (C) for all these models are shown. AUC: Area under the curve, CA: classification accuracy, F1: F1 score, Prec: precision, Recall, Spec: specificity.

Based on these results, we conclude that the difference in the set of biomarkers between a sample and the previous one does not correlate to treatment in that time.

We also tried to cluster the samples depending on the difference of biomarkers' expression between a sample and the previous one and check whether the clusters generated were different depending on whether any treatment was received between these samples (Figure R11).

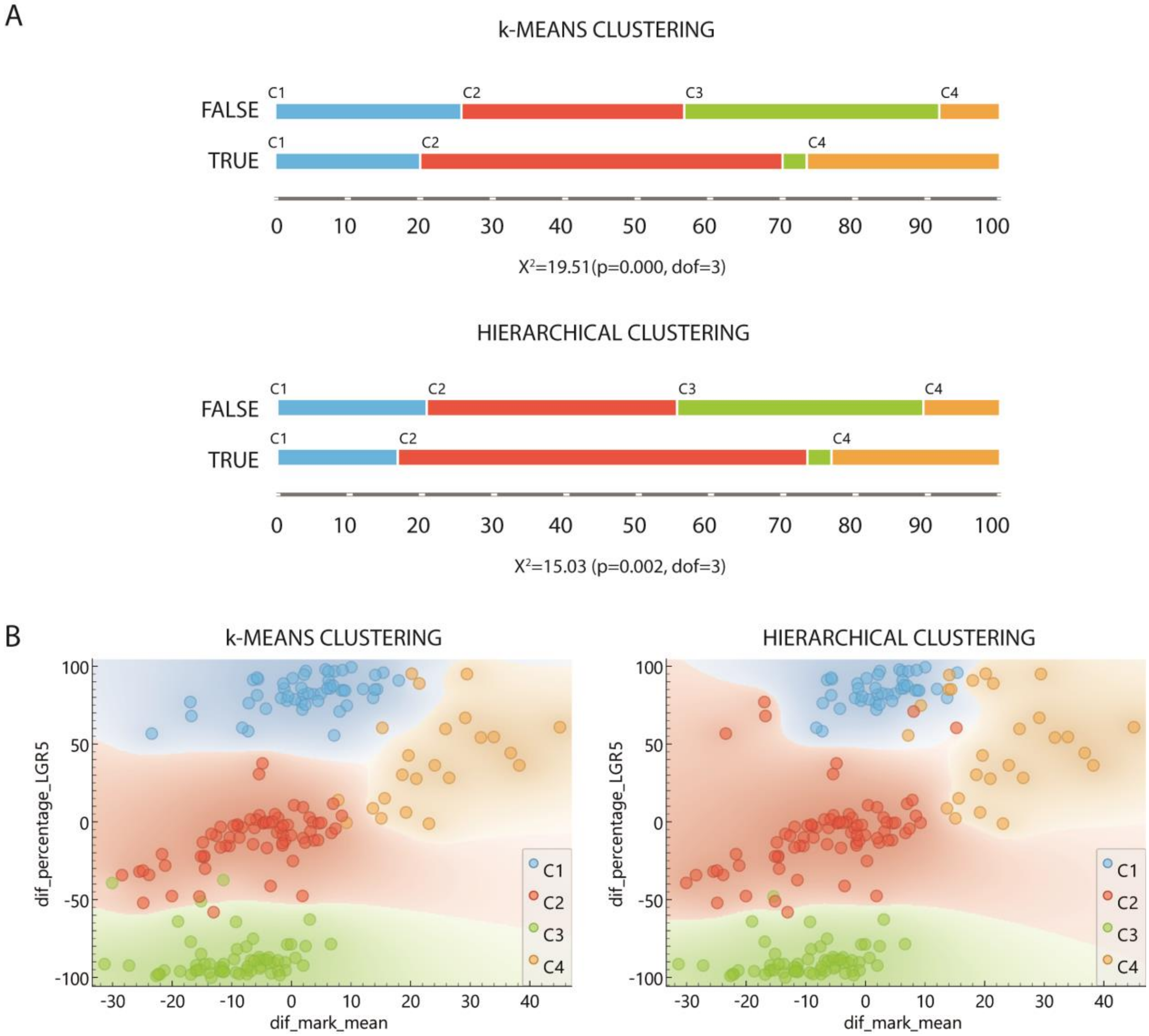

Figure R11. Correlation between the difference in biomarker expression and whereas treatment was received between a sample and the previous one by Unsupervised ML. Unsupervised ML was used to generate clusters depending on the difference of expression of biomarkers. (A) The percentage of samples which correspond to each cluster for samples that received a treatment between the sample and the previous one (True) or did not (False). (B) Difference in LGR5 expression versus the mean difference for all biomarkers.

Unsupervised ML was used for this purpose. K-means and Hierarchical clustering methods were employed, classifying the samples into four clusters. The percentage of samples which correspond to each cluster for samples that received a treatment between the sample and the previous one (True) or did not (False) was represented in Figure R11A. We found that the clusters did not separate samples depending on if any treatment was received between samples. Although there were samples for both groups in all the clusters, almost all the samples that did not receive treatment between samples were in cluster 3 in both clustering methods. We tried to analyze if any biomarker was responsible for that, and LGR5 was found to have an effect. A scatter plot showing the difference in percentage of LGR5 between a sample and the previous one was plotted versus the mean of the differences for all the biomarkers (Figure 11B). Each dot corresponds to a sample, each cluster is colored different (C1 in blue, C2 in red, C3 in green and C4 in yellow) and the background is colored based on cluster distribution. In Figure R11B, a green background was observed on the bottom side of the plot. This means that a decrease of more than 40% of LGR5 (Y-axis) defines the cluster 3 (green). The same result was obtained both by k-Means and Hierarchical clustering. Although further analysis is required, this result suggests a possible influence of the absence of treatment on a decrease of LGR5 expression.

Considering all the results obtained in this subsection, we decided to include the variable *status* (*treatment, follow-up* or *discharge*) as input for the model of relapses in the following section, as patients that were in treatment influenced on the biomarkers' expression.

4.1.3.4. Mathematical model for relapse prediction

Finally, a mathematical model was generated to predict relapse. A Python script was developed to obtain an input file with the variables of interest (expression of the different biomarkers and time from sample extraction to relapse). See more information about this Python script in the Materials and Methods section – 3.1.3. Then, Orange software was used to generate the model by applying different ML algorithms such as SVM, kNN, Logistic regression, Random Forest and Tree. We trained the model with the percentage of expression for the 13 biomarkers, and we also included the variable *status* (*treatment, follow-up* or *discharge*) as input as previous correlation analysis revealed a possible influence on the biomarkers' expression.

We used a stratified split to ensure that the proportion of each category remained similar. The purpose was to evaluate if any of these models predicts relapse and if so, which one predicts it better. We evaluated the models by different parameters (Figure R12), including AUC, CA, F1 score, Precision, Recall and Specificity. Logistic Regression, SVM, kNN Random Forest and Gradient Boosting models provided good results. However, from a clinical perspective, Recall and Specificity are the most relevant metrics, where Logistic Regression demonstrated better results. Although Recall was higher in some models, we considered both Recall and Specificity together, as it is not useful in this case to accurately predict negative results (patients that do not relapse) but fail to predict positives (patients that relapse). In fact, this last parameter (true positives, evaluated by Recall) is the most important in this project. Moreover, the model which includes the status of the patient at that moment improves the prediction. That is why we considered this one for the next steps.

A

Model	AUC	CA	F1	Prec	Spec	Recall
SVM	0.780	0.840	0.817	0.815	0.840	0.405
kNN	0.785	0.825	0.802	0.795	0.825	0.379
Logistic Regression	0.822	0.734	0.763	0.83	0.734	0.727
Random Forest	0.796	0.777	0.793	0.819	0.777	0.632
Gradient Boosting	0.729	0.807	0.806	0.806	0.807	0.513
Tree	0.518	0.789	0.787	0.785	0.789	0.452

B

Model	AUC	CA	F1	Prec	Spec	Recall
SVM	0.899	0.895	0.891	0.889	0.895	0.669
kNN	0.863	0.900	0.895	0.894	0.900	0.670
Logistic Regression	0.890	0.825	0.840	0.878	0.825	0.826
Random Forest	0.877	0.880	0.880	0.881	0.880	0.711
Gradient Boosting	0.840	0.872	0.867	0.864	0.872	0.606
Tree	0.751	0.872	0.872	0.871	0.872	0.675

Figure R12. Score of the different models based on ML to predict relapses. AUC, CA and Spec were the parameters we considered to check the best model. All these parameters range from 0 to 1. Different models (SVM, kNN, Logistic regression, Random Forest, Gradient Boosting and Tree) were tested. Two different inputs were considered: (A) 13 biomarkers' expression (B) 13 biomarkers' expression and the status of the patient at that time. The scores which evaluate the different models are shown. AUC: Area under the curve, CA: classification accuracy, F1: F1 score, Prec: precision, Recall, Spec: specificity.

Subsequently, we evaluated the results by a confusion matrix and a ROC curve (Figure R13). The predicted values are shown in the columns of the confusion matrix and the correct values are shown in the rows. The confusion matrix confirms that the best model was Logistic regression, with 17 % of false positives, 18 % of false negatives, 82 % of true negatives and 83 % of true positives. The rest of the models are better at predicting the false positive rate but less accurate at predicting the true positive rate. For this reason, we did not select those ones. The ROC curve showed the same result, the best model was Logistic Regression, as it is the closest to the top-left corner, representing 100% sensitivity and specificity.

A

	SVM	Predicted False	Predicted True	N
Actual	False	95.5%	4.5%	330
Actual	True	39.1%	60.9%	69
	N	342	57	399

	kNN	Predicted False	Predicted True	N
Actual	False	96.1%	3.9%	330
Actual	True	39.1%	60.9%	69
	N	344	55	399

	Log Reg	Predicted False	Predicted True	N
Actual	False	82.4%	17.6%	330
Actual	True	17.4%	82.6%	69
	N	284	115	399

	RF	Predicted False	Predicted True	N
Actual	False	92.1%	7.9%	330
Actual	True	33.3%	66.7%	69
	N	327	72	399

	GB	Predicted False	Predicted True	N
Actual	False	90.9%	9.1%	330
Actual	True	44.9%	55.1%	69
	N	331	68	399

	Tree	Predicted False	Predicted True	N
Actual	False	92.4%	7.6%	330
Actual	True	37.7%	62.3%	69
	N	331	68	399

B

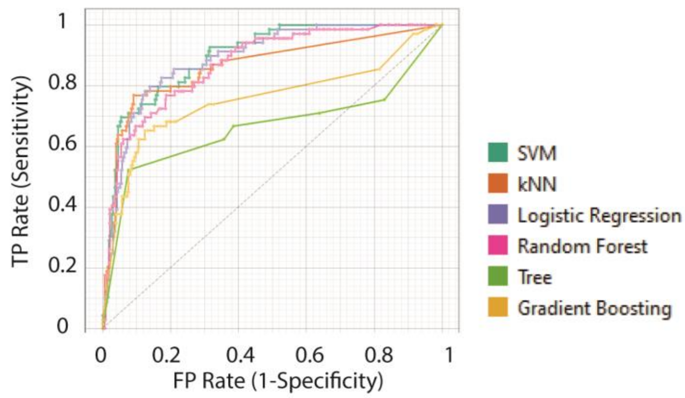

Figure R13. Evaluation of the different models based on ML to predict relapses. (A) Confusion matrix showing the false positive/negative and true positive/negative rates for each model. (B) ROC curve including all the models. The models used in this project were: SVM, kNN, Logistic regression, Random Forest. Gradient Boosting and Tree.

Although this model will be used with any new arriving samples, we simulated the model's capacity at predicting relapse, as we do not have new test samples at this moment. We randomly cropped the database into two subsets, one of them containing 25 samples and the other the rest. A Python script was also developed for that purpose. The bigger database was the one used to train the model and predict, and the smaller one to test it. The sampling and testing process was carried out ten times, and the average of each parameter was calculated. The parameters to evaluate the different models are shown in Figure R14A. Although slightly different results were obtained when compared to the one exposed in Figure R12, Logistic Regression is also the one with the best results. The worsening of the outcome is because a smaller number of samples was considered, as the samples were cropped in two lists (training and prediction), but the values still indicate a good model. The confusion matrix of the training data showed that 80.2% of samples were true negative and 83.4% were true positive (Figure 14B). Then, the model was tested with the database containing 25 random samples that were not used before in the training model. The confusion matrix calculated by the weight average of ten iterations is shown in Figure R14B. 72.7% of patients that will relapse were detected and 82.5% that will not relapse were also detected.

A

Model	AUC	CA	F1	Prec	Spec	Recall
SVM	0.887	0.898	0.893	0.893	0.898	0.669
kNN	0.867	0.892	0.890	0.889	0.885	0.711
Logistic Regression	**0.897**	**0.807**	**0.826**	**0.875**	**0.807**	**0.829**
Random Forest	0.883	0.878	0.881	0.886	0.878	0.736
Gradient Boosting	0.796	0.839	0.843	0.849	0.839	0.621
Tree	0.677	0.861	0.860	0.854	0.861	0.632

B

Training	False	True
False	80.2%	19.8%
True	16.6%	83.4%

Prediction	False	True
False	82.5%	17.5%
True	27.3%	72.7%

Figure R14. Evaluation of results from the generation of a new model after cropping the database into two new ones (training and prediction). (A) Confusion matrix for the training model (B) Confusion matrix for the test samples.

These results confirm that the model can predict relapses which is a promising result as relapse conditions the life of patients with CRC patients. We also tried to predict the time to relapse by regression models and clustering, but it was not possible due to the low number of samples with relapse (data not shown). More samples will be needed to use this prediction model in clinics.

Finally, we evaluated the influence of the biomarkers on the model for relapse prediction (Figure R15). We do not disclose the names of the biomarkers for confidentiality reasons, as we are currently in the process of patenting the relapse prediction method.

To understand this figure, we should consider that the greater the distance of the values from 0, the stronger the influence of the biomarker on the model. Conversely, the closer the values are to 0, the weaker the biomarker's influence on the model.

Moreover, the color scale, ranging from red to blue, represents the feature values (in this case, the percentage of expression of the biomarker). It is important to note that the figure is plotted with a focus on the true condition of the variable *relapse*. This means that if the red values are located on the right side, it indicates that higher feature values are associated with an increased likelihood of relapse.

Considering this analysis, the biomarkers 1 to 4 presented the greatest influence on the model, as their values were furthest from 0. Following them, biomarkers 5 to 8 also contributed significantly to the model. Meanwhile, biomarkers 9 to 11 displayed a moderate level of influence, whereas biomarkers 12 and 13 had a minimal impact, being positioned close to 0.

Regarding the feature values, higher expression levels of biomarkers 1, 5, 6, 8, 9 and 11 were associated with an increase likelihood of relapse whereas lower expression levels of biomarker 2, 3, 4, 7 and 10 contributed to a higher likelihood of relapse.

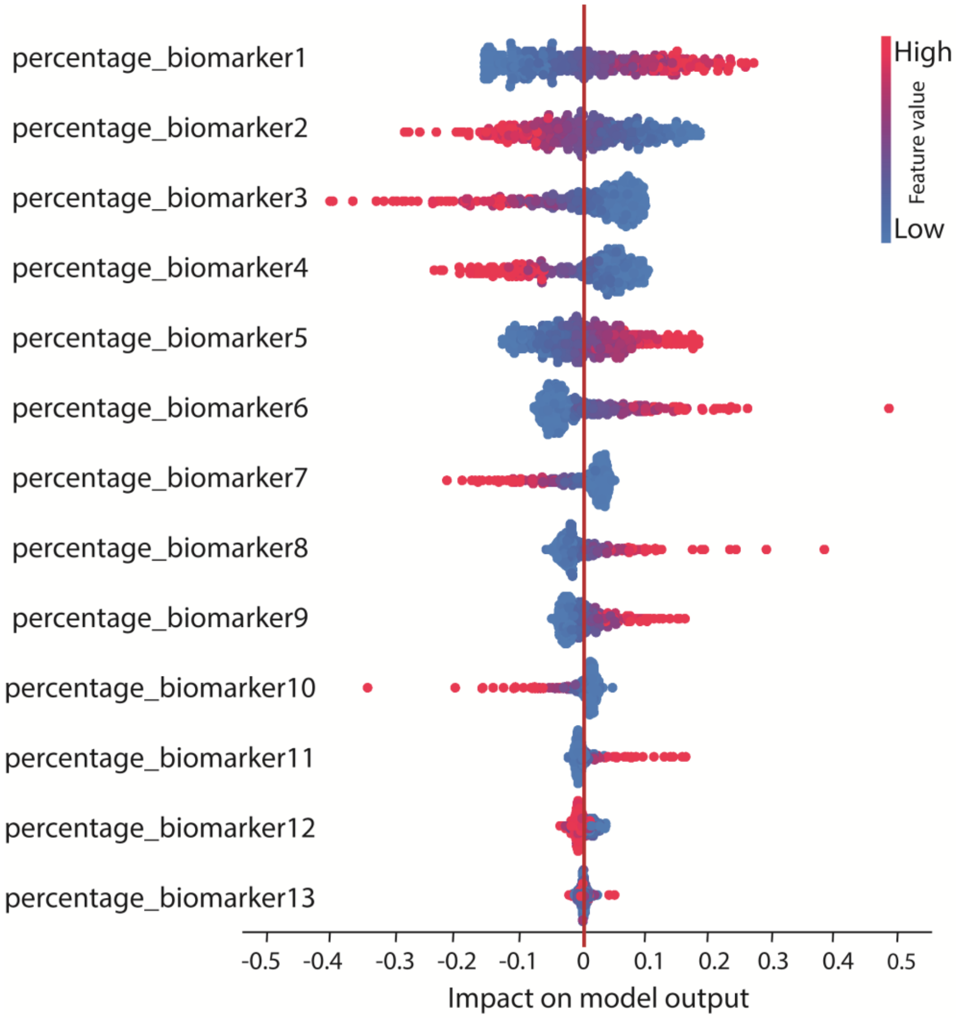

Figure R15. Influence of the biomarkers on the model for relapse prediction. The biomarkers are ordered according to their impact on the model output, with those positioned higher in the graph having greater influence. The farther the values are from 0, the stronger their impact on the model. The color scale which ranges from red to blue represents the feature values.

4.2. Resistance to chemotherapy

Once the mathematical model for relapse prediction was developed, our aim was to study the context of resistance. For that purpose, we calculated the percentage of patients that presented resistance for each treatment line: localized disease (LD), first line (1L), second line (2L) and third line (3L). Resistance was observed in 39.8% of the patients in LD. This percentage increased for the 1L and 2L to 67.10% and 60.8%, respectively. Then, the percentage of patients resistant to treatment decreased to 35.5% probably because it has been a short time since that diagnosis was made and most of these patients are now in treatment for this line.

Not only resistance but also toxicity was high in these patients. Most of them (from 50 to 70% approximately) presented toxicity to the treatment administered. Differences were not found between lines of treatment, but they were found depending on the chemotherapeutic. FOLFOX was the one which produced a higher proportion of patients with toxicity followed by capecitabine and FOLFIRI. All these results are shown in Figure R16.

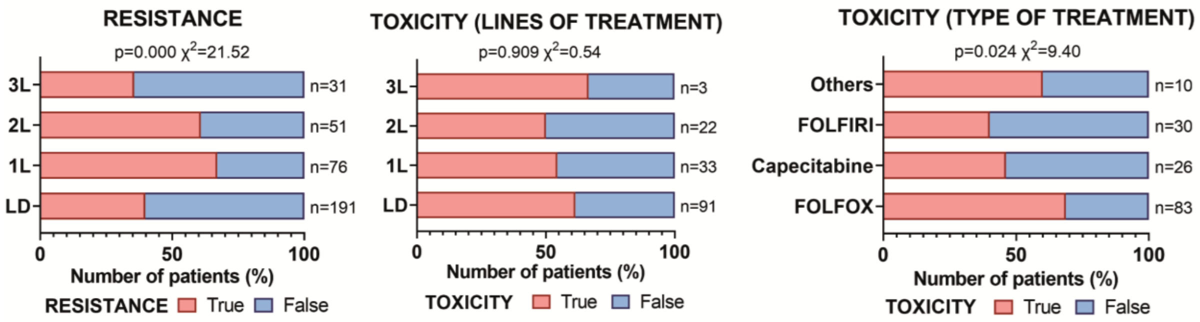

Figure R16. Resistance and toxicity analysis in CRC patients. The percentage of patients for each condition was calculated and compared by a Chi Square test (p-values are shown in the figure). Each variable was represented for patients that presented resistance/toxicity (True, in red) and did not (False, in blue). Different lines of treatment (LD: local disease, 1L: first line, 2L: second line, 3L: third line) and chemotherapies (FOLFOX, Capecitabine, FOLFIRI and others such as TAS102 and irinotecan) were evaluated. Resistance means patients who did not respond to treatment and progressed. Toxicity refers to patients who developed side effects due to treatment.

Given the increasing number of patients who develop resistance after chemotherapy, we aimed to generate patient-derived tumors to establish new patient-derived cultures useful for studying resistance to various chemotherapy agents. Additionally, we developed a novel growth rate-based method to measure chemotherapy resistance, addressing several limitations of the conventional IC_{50} method. Finally, we explored potential biomarker candidates for resistance detection. As part of a preliminary study, we identified TROP2, a biomarker used in the relapse prediction model, as potentially involved in resistance. Using the newly developed method, we investigated the correlation between TROP2 expression and resistance in five CRC cell lines. To further confirm its role in resistance, we examined whether chronic oxaliplatin treatment affected TROP2 expression and whether gene silencing influenced resistance.

4.2.1. Culture of patient-derived tumors

The purpose of this section was to grow tumors in animal models by injecting PBMCs isolated from blood of patients that may content the CSC population. Once the tumors grew, we would dissect them, disaggregate and establish a culture *in vitro*. These cultures will serve to study efficacy and resistance to treatments.

For that, we first tried to inject the higher number of PBMCs as possible. We selected some patients implied in the study that may have more probability to present CSCs in blood, that is stage IV patients or patients with a higher percentage of positive cells for the biomarkers studied in a previous sample. We used nude mice and injected the cells in a 1:1 proportion in Matrigel. We injected PBMCs from a total of 24 patients, but tumors did not grow in any of them (Figure R17A).

Next, we tried the same strategy but changed the animal model to chicken embryos (Figure R17B). Their immune system is not developed because of their condition of being embryos, allowing human cells to grow. Previous researchers have used this animal model for the same purpose [231], it consists of opening a window in the shell and injecting the cells over the CAM. If the injected cells have CSCs, a tumor is supposed to be grown on the site of injection. Moreover, if they have the capacity to migrate and invade, they will evade from the tumor and travel through the veins to the embryo. Metastasis could be analyzed in the organs of the embryos, usually in the liver and lungs as CRC usually metastasize there.

PBMCs were isolated and seeded onto the CAM using both the drop method and a sterile ring. The DLD1 cell line was used as a positive control. The windows were then sealed, and the embryos were returned to the incubator. Daily monitoring was performed to detect potential tumor growth. After seven days, the embryos were macroscopically examined to assess tumor development. No tumor growth was observed in any case, including the DLD1 positive control.

Although we have not yet achieved tumor growth, it is a tedious technique which requires a lot of experience. We have improved the method and have practice on that. Further attempts, incorporating varying cell line concentrations and in collaboration with experienced laboratories, will be essential to ultimately achieving the objective.

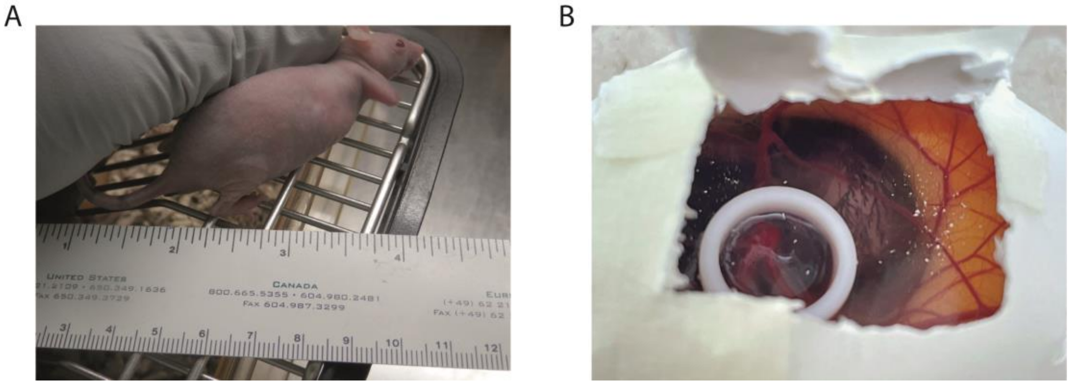

Figure R17. Representative image of animal models injected with PBMCs from CRC patients. The growth of the tumor was not achieved by injecting the highest possible number of PBMCs with Matrigel in a 1:1 proportion in nude mice (A) nor chicken embryos (B).

Taking into account these results, we considered profitable a placement in the laboratory of Dr. Frédéric Hollande, in the University of Melbourne, Australia. They are experts on culturing organoids and have developed organoids from tumor tissue from patients. Moreover, they have also managed to generate cultures from liquid biopsy [212]. In my placement in the laboratory of Frederic Hollande, he and his group taught me about these techniques. Unfortunately, I did not have access to the samples from my patients in there, so I carried out the experiments with my patient's samples when I came back to Spain.

In my placement, they advised me to incorporate the following improvements:

- Enrichment of CSCs by depleting the CD45 positive cells.
- Use the NSG model rather than the nude mice model as they have demonstrated to be a better model to generate Xenografts [232].
- Try to do tail vein injections to maintain the environment of CSCs, that are in suspension and travel through blood from one to another organ.
- Directly grow the enriched cells in culture plates, either in Matrigel domes or in suspension, using medium supplemented with growth factors.

To achieve these suggestions, we first depleted the CD45 positive cells corresponding to the immune cells present in blood. We incubated with CD45 antibody and then with a secondary anti-mouse antibody with magnetic beads to retain them in a MS Column. A MACS® magnetic separator was used to generate the magnetic field and CD45 negative and positive cells were collected in different tubes.

Then, they were processed separately for the following procedures:

- Confocal microscopy to check the magnetic separation and to detect some CSCs if possible.
- Injection in mice flank and vein tail from NSG mice.
- Culture directly with Matrigel and in suspension with M12 Medium.

For that purpose, we selected a total of five patients that were Stage IV at diagnosis as they will present CSCs in blood with a higher probability. After CD45 magnetic separation, cells were stained with Hoechst to stain the nuclei, CD45/CD68 conjugated to FITC to stain the immune cells and 8- and pan-keratin conjugated to PE to stain the cytoplasm of possible CSCs. Results are shown in Figure R18 where CD45 positive selected cells present a higher number of immune cells when compared to CD45 negative selected cells.

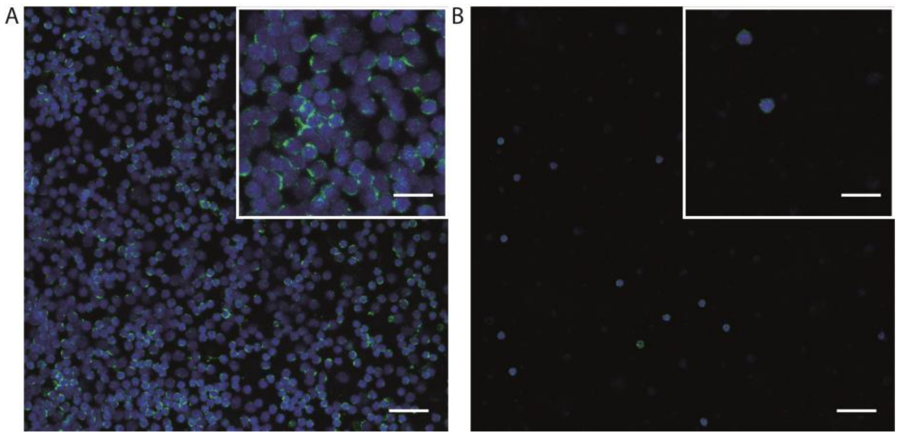

Figure R18. Immune-magnetic separation of positive and negative CD45 expressed cells. CD45 positive selected cells are shown on the left and CD45 negative selected cells are shown on the right. Depletion of CD45 positive cells allows to deplete most of the immune cells. The scale bar represents 100 μm for the full image and 25 μm for the inset.

Although we did not find keratin-stained cells compatible with a CSCs, we can conclude the experiment by saying that it is an effective way to enrich the population of CSCs by depleting the immune cells. If we deplete most of the immune cells, we now are more likely to inject the CSCs if considering the same number of cells in each injection. We first injected all the cells extracted from 4 ml blood tube, but another possibility is that the cells did not grow because there were such a large number of cells that they clumped, and the CSCs could not grow. The depletion of CD45 could also help in this aspect. That is why we considered it a positive result and included this depletion in the next experiments.

Next, we injected CD45 positive and negative isolated cells into NSG mice both in the flank and in the tail vein. The same selected patients were used for this procedure. Figure R19 shows the mice three months post-injection. In the case of flank injection, tumor growth was not observed in most cases (Figure R19 - upper left), but tumors developed in two mice (Figure R19 - lower left). For tail vein injection, no tumor growth was detected in any organs of the mice (Figure R19 - right).

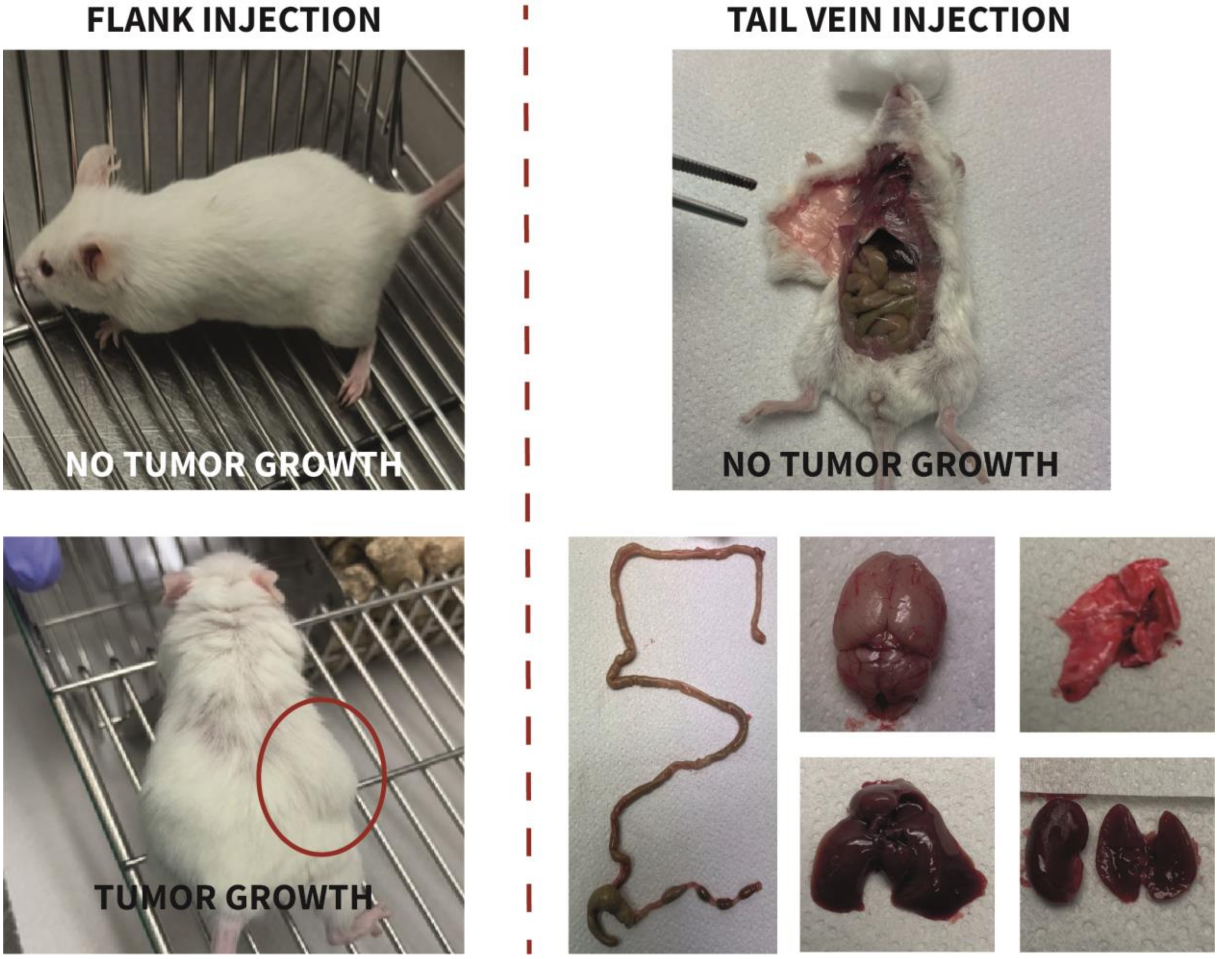

Figure R19. Injection of CD45 positive and negative PBMCs from CRC patients into NSG mice. Tumor growth was observed in two out of five mice following flank injection (left) whereas no tumor growth was observed in any mice after tail vein injection (right). A representative image is shown for each condition.

We dissected the tumors from the mice and cultured the cells both in Matrigel and in suspension. Figure R20 shows the growth of one-month post-seeding. While the cells remain viable, additional time is required to establish a stable culture.

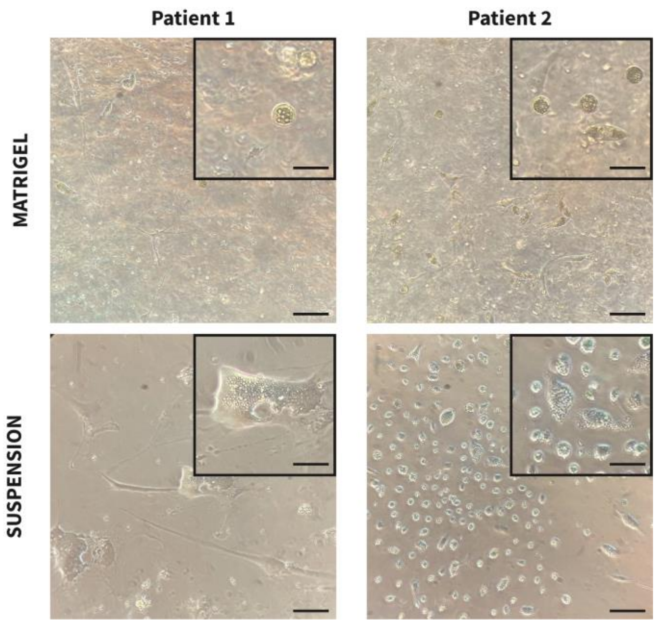

Figure R20. Growth of cells dissociated from patient-derived tumors. Tumors were dissected and seeded both in Matrigel and in suspension for two patients. A representative image is shown for each condition. The scale bar represents 100 µm for the full image and 25 µm for the inset.

Meanwhile, we also tried to establish cultures directly from CD45 positive and negative selected PBMCs in Matrigel domes and in suspension with M12 medium. Figure R21 shows the growth for suspension growth in a day, a week, three weeks and a month post-injection.

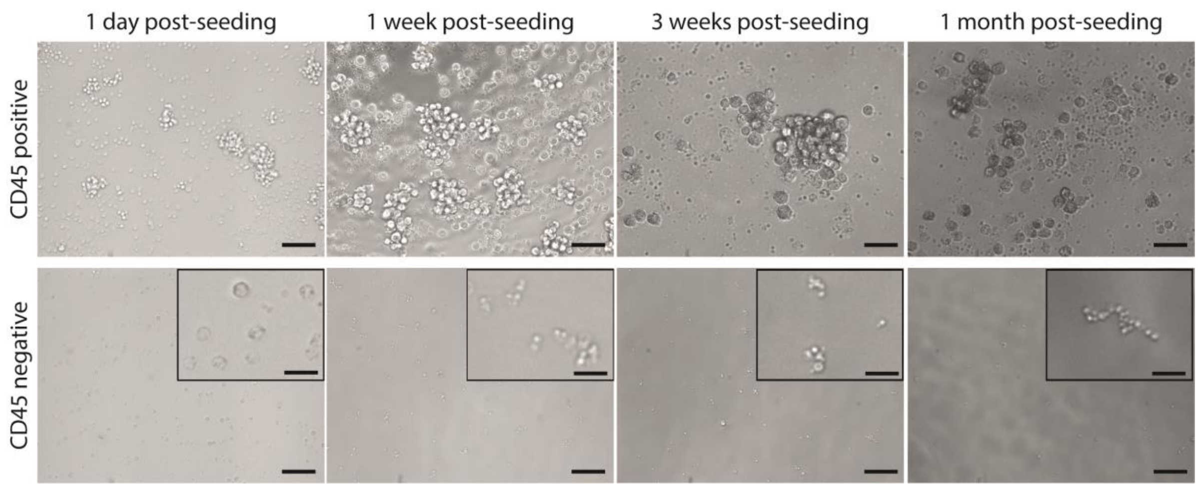

Figure R21. Growth of CD45 positive and negative PBMCs from CRC patients in suspension. Cells were isolated and seeded in M12 medium. They were maintained for a month and pictures were taken at different times. Some clumps were formed in the first weeks but then the cells died. We were not able to establish a suspension culture from CRC patients. A representative image of one patient is shown. The scale bar represents 100 µm for the full image and 25 µm for the inset.

Although some clumps were formed in the first week for CD45 positive cells, the cells finally died, and tumors were not established. In the case of CD45 negative cells, the number and size of the cells is too small, and they also finally died.

We also tried to generate patient-derived tumors in Matrigel. For that, we isolated the cells in the same way and seeded the cells in a 1:1 proportion in Matrigel. A dome of Matrigel was seeded in a 24-well plate and COM medium was added to cover it. The cells were maintained for a month. Although some clumps were formed in the first week, the cells then died so we were not able to establish the culture (Figure R22).

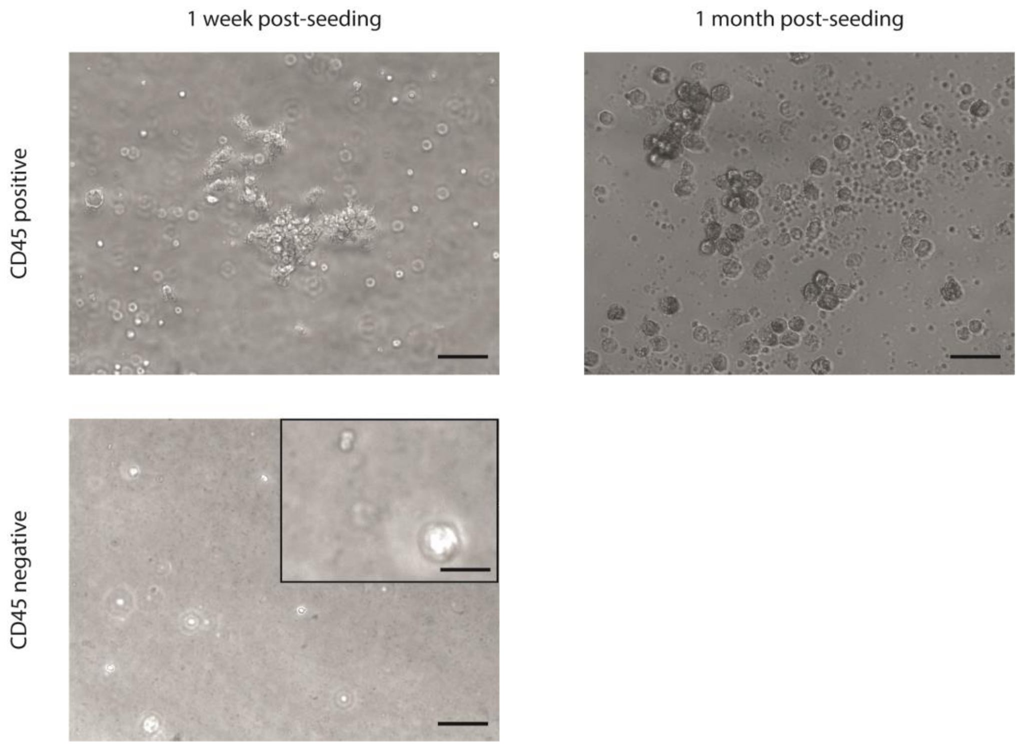

Figure R22. Growth of CD45 positive and negative PBMCs from CRC patients in Matrigel domes. Cells were isolated and seeded in a 1:1 proportion in Matrigel. They were maintained for a month and pictures were taken at different times. Some clumps were formed in the first week but then the cells died. We were not able to establish a culture in Matrigel domes from CRC patients. A representative image of one patient is shown. The scale bar represents 100 µm for the full image and 25 µm for the inset.

In conclusion, we improved the method and successfully achieved tumor growth for two patients. Since additional time is needed to establish the cultures for use in resistance experiments, we conducted several experiments on CRC cell lines, including the development of a novel method to measure resistance to chemotherapy.

4.2.2. New growth rate-based indices to measure resistance to chemotherapy

Cell viability assays have become instrumental tools for assessing the efficacy of anticancer agents, particularly in the early stages of treatment studies. Diverse methodologies have emerged, including MTT assays. Among these assays, the discovery of parameters to analyze and compare the efficacy and/or resistance to treatments is essential.

Our findings show (Figure R23), in five different CRC cell lines, that there are significant differences in the calculated IC_{50} value depending on the end-point at which the percentage of viable cells was calculated (24, 48 or 72 hours).

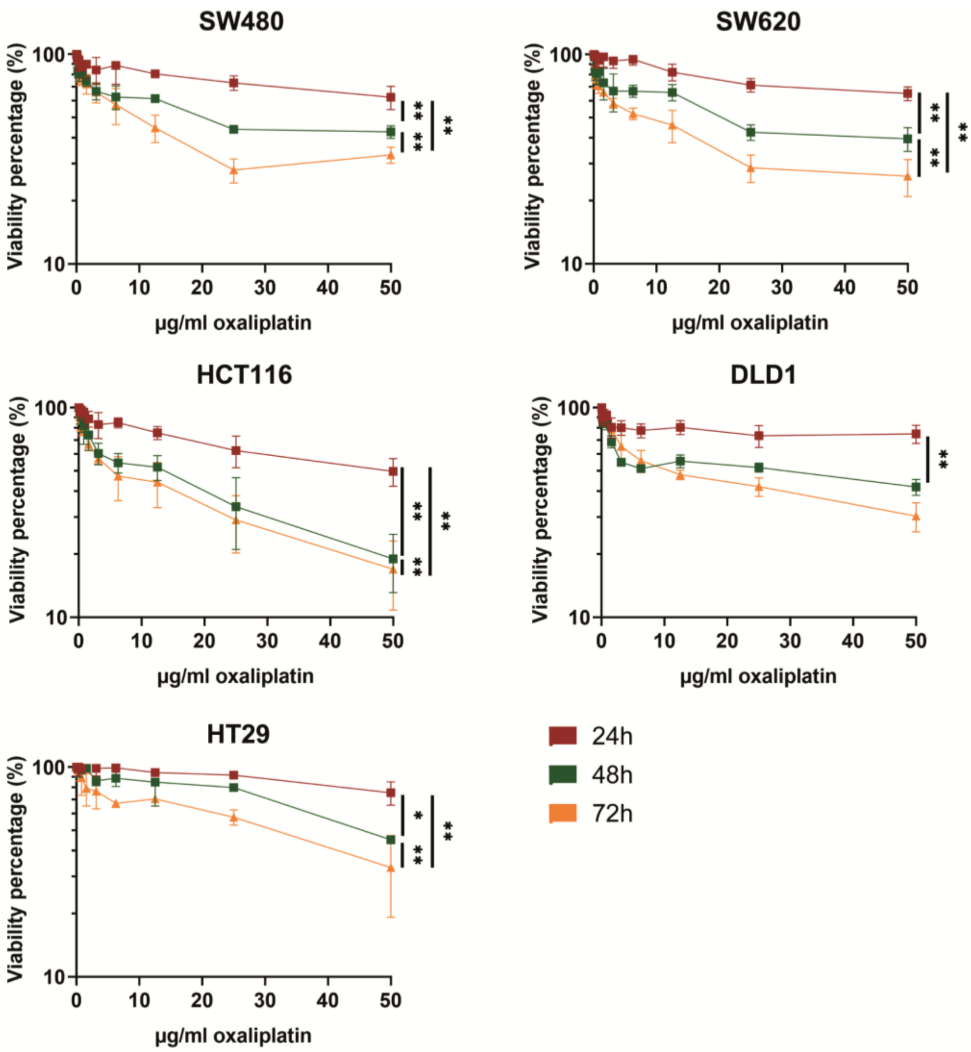

Figure R23. Response to chemotherapy is time dependent in different cell lines. (A) Viability was quantified in five cancer cell lines at three different endpoints (24, 48 and 72 hours). Differences between the curves were evaluated using the Wilcoxon test. The probability is indicated by * $p<0.05$, ** $p<0.01$ and *** $p<0.001$.

In addition, when comparing two different methods of IC_{50} calculation (GraphPad Prism [233] and IC_{50} Calculator [234]), the results for the same time-point and cell line also varied (Figure R24).

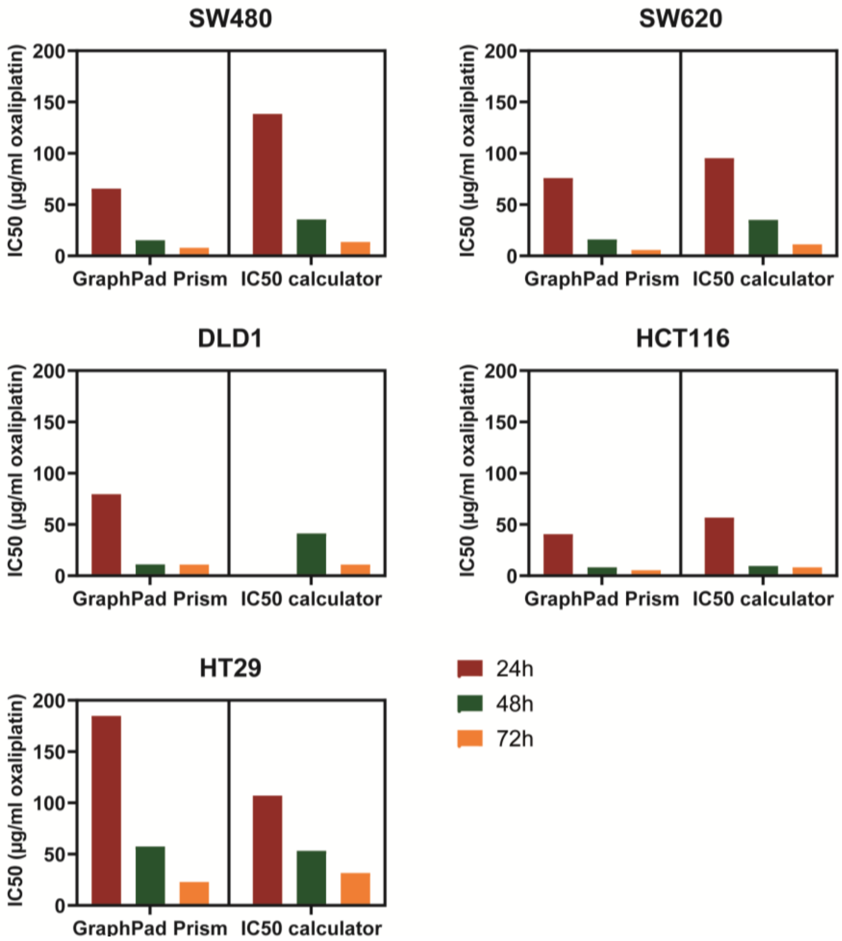

Figure R24. Response to chemotherapy is method dependent in different cell lines. IC_{50} value was calculated by GraphPad Prism and IC_{50} calculator programs in five cancer cell lines for three endpoints (24, 48 and 72 hours).

In this thesis we have addressed this issue by proposing a new time-independent method based on the calculation of the effective growth rate for a range of drug concentrations. We also propose the use of two new parameters to evaluate the response of cell lines to different treatments that are not time dependent. One of them is the IC_{r0} index, a value that represents the drug concentration at which the effective growth rate is zero. The other, the $IC_{rmed,}$ is the value that corresponds to the drug concentration that reduces the control population's growth rate by half.

To determine whether a cell colony is proliferating or declining, the population evolution was assessed. Here, we calculated the effective growth rate both in the control (without treatment) and in a range of drug concentrations.

First of all, the exponential growth condition had to be verified: control data are plotted on a semi-logarithmic scale to verify this exponential trend. The HCT116 cell line was used as an example in Figure R25A for this purpose; in this figure we see exponential growth in HCT116, as it depicts a linear trend when plotting on the semilog scale. This result ensures that external factors were not affecting the experiment, and the effect was due to the variable of interest. The experimental data are shown as orange dots and the fit for these data is shown as a red line. For each experimental condition, 1,000 bootstrap linear regressions were performed to obtain the growth rate distribution (each gray line corresponds to the linear fit of one bootstrap sample). The distribution of the bootstrap linear regression slopes is shown in Figure R25B, presenting a normal distribution of the data. Similar results were obtained at different drug concentrations (as an example, a concentration of 1.56 µg/ml of oxaliplatin is shown in Figure R25C and D). The same procedure was performed for the rest of the cell lines and drug concentrations.

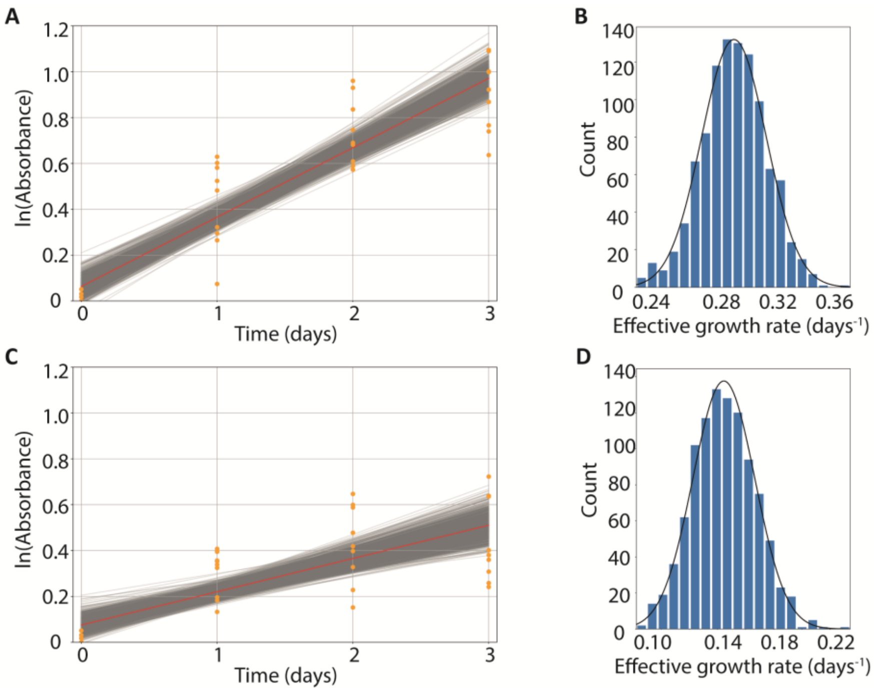

Figure R25. Growth rate calculation and distribution in HCT116 CRC cell line. (A) Growth rate calculation for HCT116 control condition. 1,000 linear regressions performed over the paired bootstrap samples of the experimental data (shown as orange dots). Each gray line corresponds to each bootstrap linear regression and the red line represents the mean of these fits. (B) Distribution of the bootstrap linear regression slopes in HCT116 control data. (C) Growth rate calculation for HCT116 with oxaliplatin at 1.56 µg/ml concentration (D) Distribution of the bootstrap linear regression slopes in HCT116 with oxaliplatin at 1.56 µg/ml concentration.

The effective growth rate for the HCT116 cell line under different oxaliplatin concentrations is shown in Table R2. The growth rate turns from positive to negative as the oxaliplatin concentration gets higher. The control growth rate of a cell line can be determined under drug-free conditions (C=0), and changes in the growth rate pattern observed with increasing drug concentrations can be used to characterize the treatment response.

Table R2. Effective growth rate and confidence intervals at 90% for HCT116 cell line under different oxaliplatin concentrations.

Concentration (μg/ml)	Effective growth rate (days^{-1})	Standard error
0	0.30	0.04
0.1	0.30	0.04
0.2	0.27	0.03
0.4	0.25	0.03
0.8	0.20	0.03
1.6	0.14	0.04
3.13	0.08	0.03
6.25	-0.01	0.04
12.5	-0.05	0.04
25	-0.21	0.06
50	-0.44	0.04

An exponential dose-dependent decrease of the growth rate was found in all the studied cell lines. Then, 1,000 Monte Carlo shifted exponential regressions considering the slope distribution obtained in the previous effective growth rate calculation were simulated. The shifted exponential fit for the five cell lines is shown in Figure R26, where each gray line is a shifted exponential regression fit. The dose-dependent reduction in growth rates in this model, as observed in all five of our cell lines, thus allows for a more robust method of study of *in vitro* response to drugs, as the use of bootstrapping methods means that values are the result of thousands of iterations of the fit to the curve. This larger number of iterations helps to reduce the error of the fitting parameters.

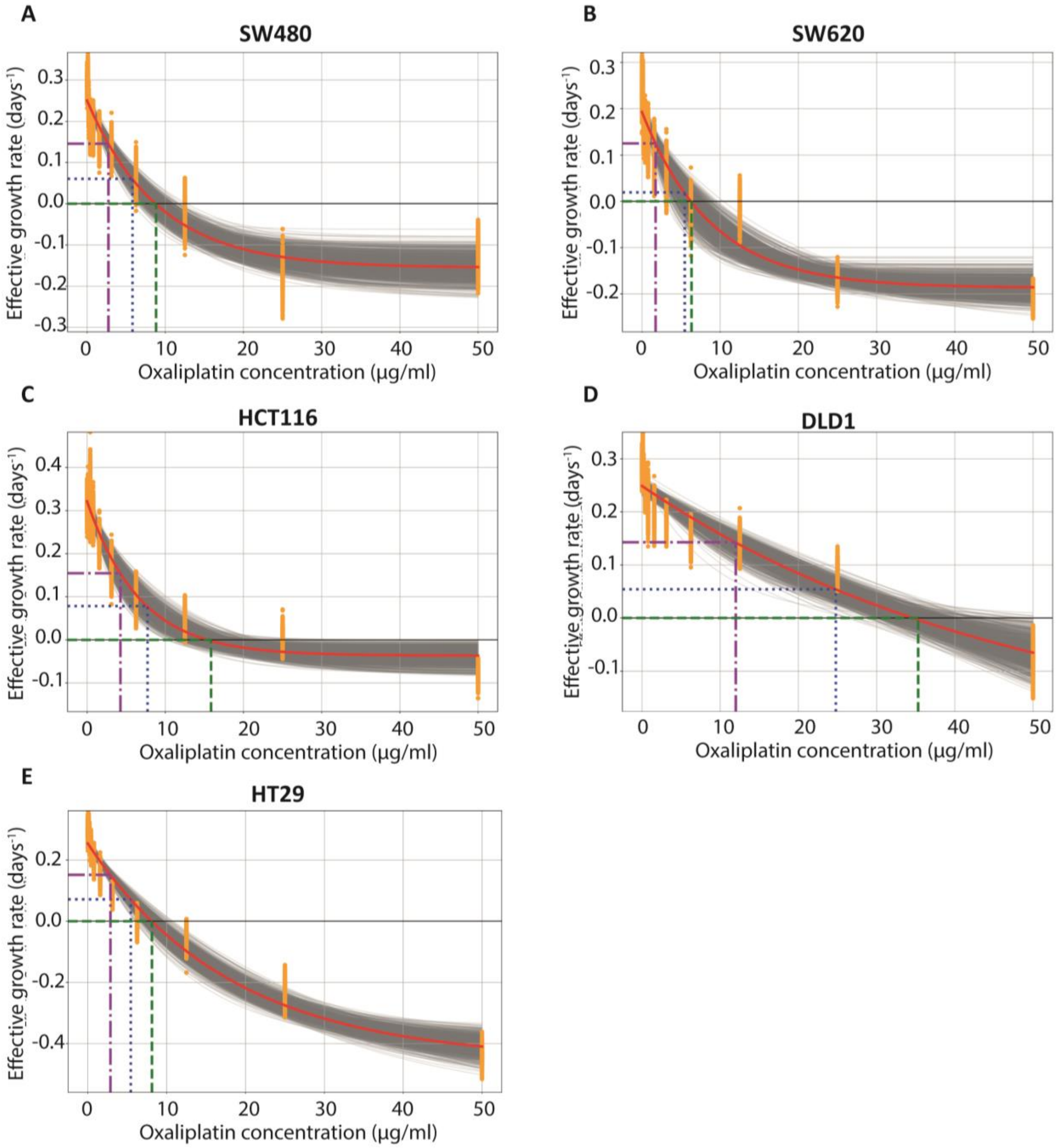

Figure R26. Effective growth rate decreases in an exponential dose-dependent manner. 1,000 Monte Carlo exponential regressions were performed from the slope distribution obtained in the previous effective growth rate calculation (shown as orange dots). Each gray line corresponds to each bootstrap exponential regression and the red line is the mean of these fits. The parameters to measure resistance are shown in different guide lines: IC_{50} (dotted blue), IC_{r0} (dashed green) and IC_{rmed} (dashdot purple). Five CRC cell lines were used: SW480 (A), SW620 (B), DLD1 (C), HCT116 (D) and HT29 (E).

This data was employed to calculate indices that may be useful to compare the response to treatment between cell lines or different culture conditions. IC_{50} value was calculated using the new method proposed in this thesis which uses the cell growth rate. As this parameter is the dose at which cell viability is 50%, we estimated the growth rate in which this assumption was met. Then, we calculated the concentration of oxaliplatin that corresponds to this growth rate for each of the 1,000 shifted exponential fit iterations. We finally obtained the mean and the confidence interval for all these iterations. The results for all the cell lines and a comparison with the values obtained by conventional methods are shown in Table R3. As the IC_{50} method provided values that differ from each other as well as from the value proposed in this PhD thesis, we conclude that it is not only time-dependent but also method-dependent.

Table R3. Parameters to measure cell lines' resistance to drugs. IC_{50} was calculated by three methods: GraphPad Prism, IC_{50} calculator and the new method described in this PhD thesis.

Cell line	IC_{50} GraphPad Prism (μg/ml)	IC_{50} Calculator (μg/ml)	IC_{50} New method (μg/ml)
SW480	7.89 [5.61 - 11.11]	13.52	5.8 ± 1.3
SW620	5.83 [3.90 - 8.71]	11.26	5.5 ± 2.1
DLD1	10.85 [8.57 - 13.76]	10.92	7.8 ± 2.2
HCT116	5.37 [3.95 – 7.29]	8.22	5.5 ± 1.0
HT29	22.88 [15.90 – 33.31]	31.54	25.5 ± 4.7

Moreover, two new parameters are proposed to study drug resistance in this thesis. IC_{r0} value is the dose at which the population is neither growing nor dying, so the growth rate is 0 (called r_0). On the other hand, the IC_{rmed} value corresponds to the dose at which the population is growing at half the rate of the control (called rmed).

These parameters are not time-dependent by definition as they are calculated from the growth rate. IC_{50}, IC_{r0} and IC_{rmed} values were obtained for each Monte Carlo exponential regression, resulting in the distribution shown in Figure R27 for all the cell lines.

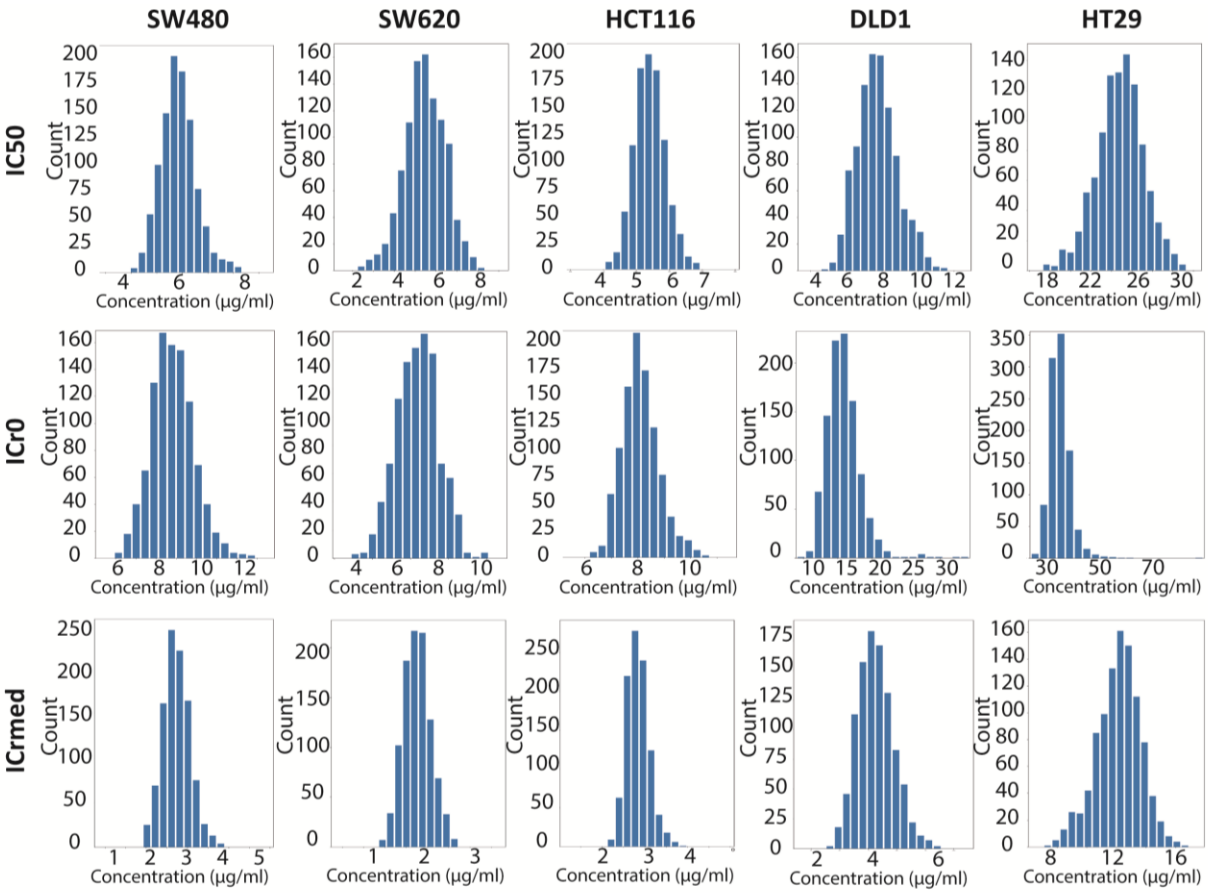

Figure R27. Distribution of the parameters to measure cell lines' resistance to drugs. IC_{50}, IC_{rmed} and IC_{r0} distributions are shown for the five CRC cell lines.

The means of these values and the 90% confidence intervals are shown in Table R4. IC_{r0} and IC_{rmed} classify the cell lines in terms of resistance, maintaining the same ranking pattern and yielding consistent results when compared to the conventional IC_{50} method. It is worth noting that although similar, we recommend the use of the IC_{rmed} value over the IC_{r0} value for classification, as the IC_{r0} can be difficult to define for highly resistant cell lines. Use of the IC_{rmed} allows for evaluation of resistant cell lines that might never reach IC_{50} or r_0 values in the experimental timeframe, and avoids the need to extrapolate this value, as would be done in conventional IC_{50} calculations. When using both IC_{r0} and IC_{rmed}, cell lines with higher values are classified as more resistant. This allows for direct comparison between cell lines or culture conditions, depending on the desired experimental outcome.

Table R4. Parameters to measure cell lines' resistance to drugs. IC_{50}, IC_{r0} and IC_{rmed} were calculated by the new method described in this article. Means and 90% confidence intervals are shown.

Cell line	IC_{50} (µg/ml)	IC_{r0} (µg/ml)	IC_{rmed} (µg/ml)
SW480	5.8 ± 1.3	8.9 ± 2.0	2.7 ± 0.6
SW620	5.5 ± 2.1	6.3 ± 2.4	1.8 ± 0.7
DLD1	7.8 ± 2.2	16.0 ± 5.0	4.3 ± 1.2
HCT116	5.5 ± 1.0	8.3 ± 1.5	2.9 ± 0.5
HT29	25.5 ± 4.7	37.0 ± 10.0	12.0 ± 3.0

4.2.3. Novel biomarker candidate for resistance detection

Once we have developed a robust method to evaluate resistance, we searched for novel biomarkers to detect resistance. For that purpose, we carried out a preliminary study by measuring the expression of the selected biomarkers in three CRC cell lines and compared to resistance. This preliminary assay allowed us to select TROP2 as a candidate biomarker for resistance detection, so we broadened on this. We increased the number of cell lines and evaluated the correlation between its expression at gene and protein levels and the resistance by the new IC_{rmed} method. We also carried out chronic treatment to evaluate differences in expression in resistant cells and silenced the gene to check whether silenced cells caused changes in resistance. All these experiments are detailed in the next subsections.

4.2.3.1. Preliminary study of correlation between biomarkers and resistance

To check whether any of the biomarkers of study were related to resistance, we measured the expression of each biomarker in three CRC cell lines: SW480, SW620 and DLD1 (Figure R28A).

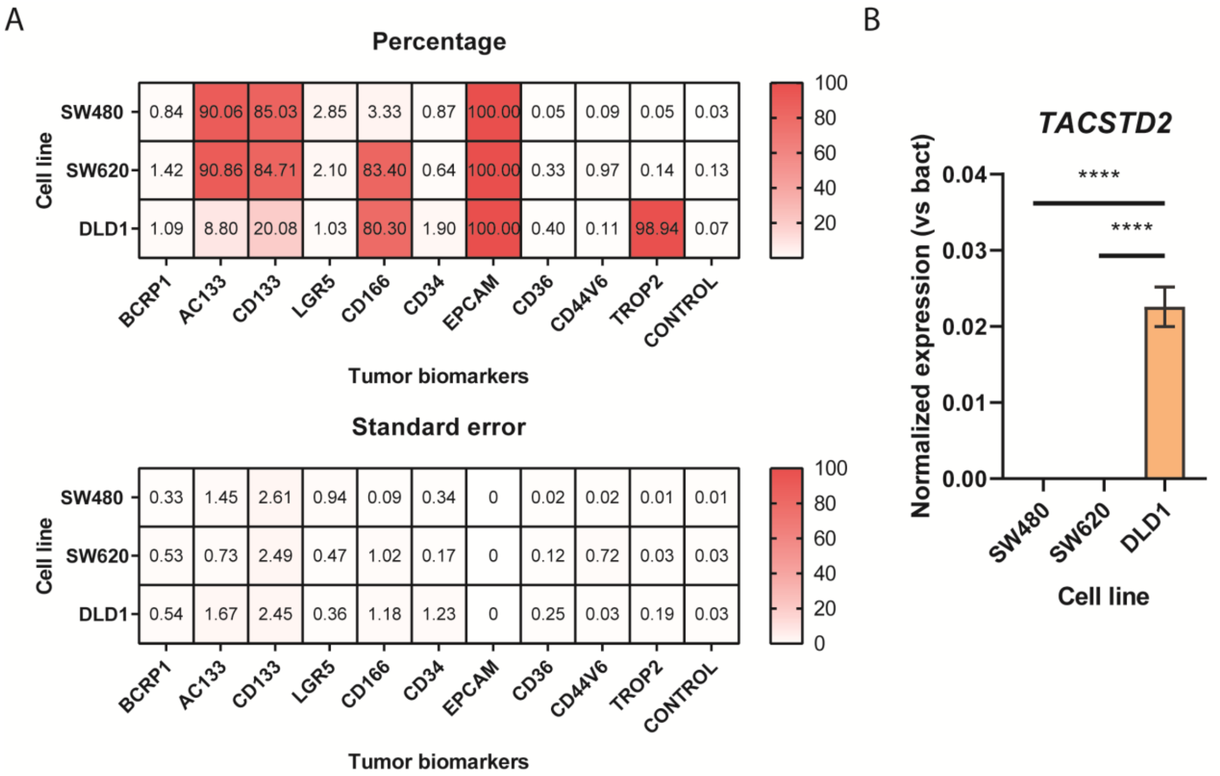

Figure R28. Expression of biomarkers in three CRC cell lines (SW480, SW620 and DLD1). (A) Flow cytometry was carried out for all the biomarkers in the three cell lines. The percentage and standard error are shown. N=3 (B) The expression of *TACSTD2* was measured by RT-qPCR. *β-ACTIN* gene was used to normalize the gene expression. N=3. Significant differences are indicated (ANOVA test, **** p < 0.0001).

DLD1 is characterized as being a more resistant cell line, so TROP2 seems to be a candidate for measuring resistance as its expression by flow cytometry in DLD1 was higher when compared to SW480 and SW620 (0.05% in SW480, 0.14% in SW620 and 98.94% in DLD1). Then, we evaluated its expression by RT-qPCR. The *TACSTD2* expression profile was similar to that observed in the flow cytometry assay, with no expression in the SW480 and SW620 cell lines and significantly different expression in DLD1, confirming the previous results (Figure R28B).

In order to evaluate the resistance of each of the cell lines to oxaliplatin, we exposed cells for two weeks to oxaliplatin and evaluated the behavior of each cell line. Figure R29A shows micrographs of each cell line and condition, where a decrease in the number of cells was observed comparing oxaliplatin-treated cell lines to control untreated cell lines. The number of cells was represented in Figure R29B for all the conditions. A significant decrease of the number of cells was found when comparing oxaliplatin treated cells to control non-treated cells in all the cell lines.

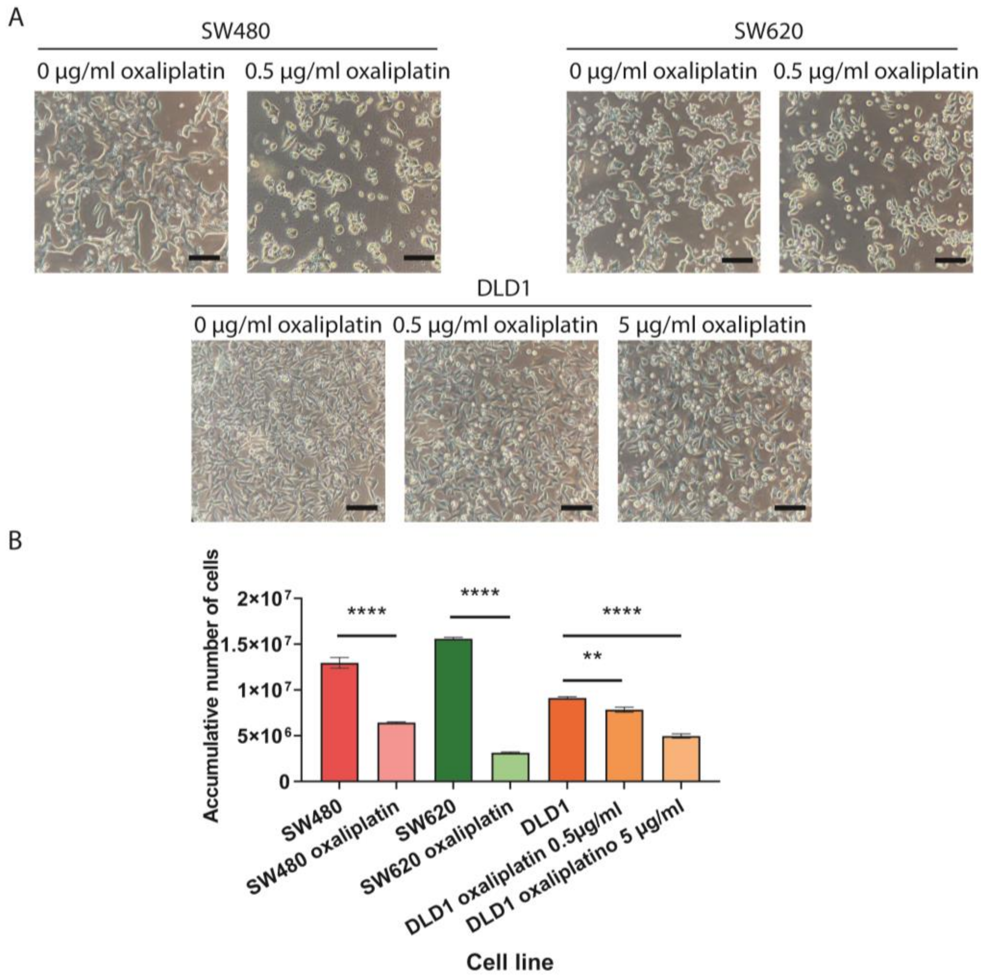

Figure R29. Representative images and cell count of non-treated and previously oxaliplatin treated cells in SW480, SW620 and DLD1 cell lines. Cells were treated for two weeks at a concentration of 0.5 µg/ml oxaliplatin. A concentration of 5 µg/ml was also used in the DLD1 cell line because of its known resistance to drugs (A) Micrographs of each cell line and condition (B) Number of cells comparing treated versus non-treated cells in SW480, SW620 and DLD1. Three measures were carried out. Significant differences are indicated (ANOVA test, ** $p<0.01$, **** $p< 0.0001$).

The number of cells for each passage was also represented (Figure R30). An exponential curve was obtained in controls, as expected for cell growth. The growth curve was flattened when cells were treated with 0.5 µg/ml oxaliplatin in the case of SW480 and SW620. Negative growth was even observed in SW620, confirming cell death. A specific behavior was seen in DLD1 cell line as a consequence of being a more resistant cell line, as this concentration of oxaliplatin (0.5 µg/ml) reduced the number of live cells, thus the growth rate, but was not enough to flatten the DLD1 cell line growth curve. For that purpose, a higher dose of oxaliplatin was employed (5 µg/ml oxaliplatin), obtaining similar results to those obtained in the other cell lines with the 0.5 µg/ml dose.

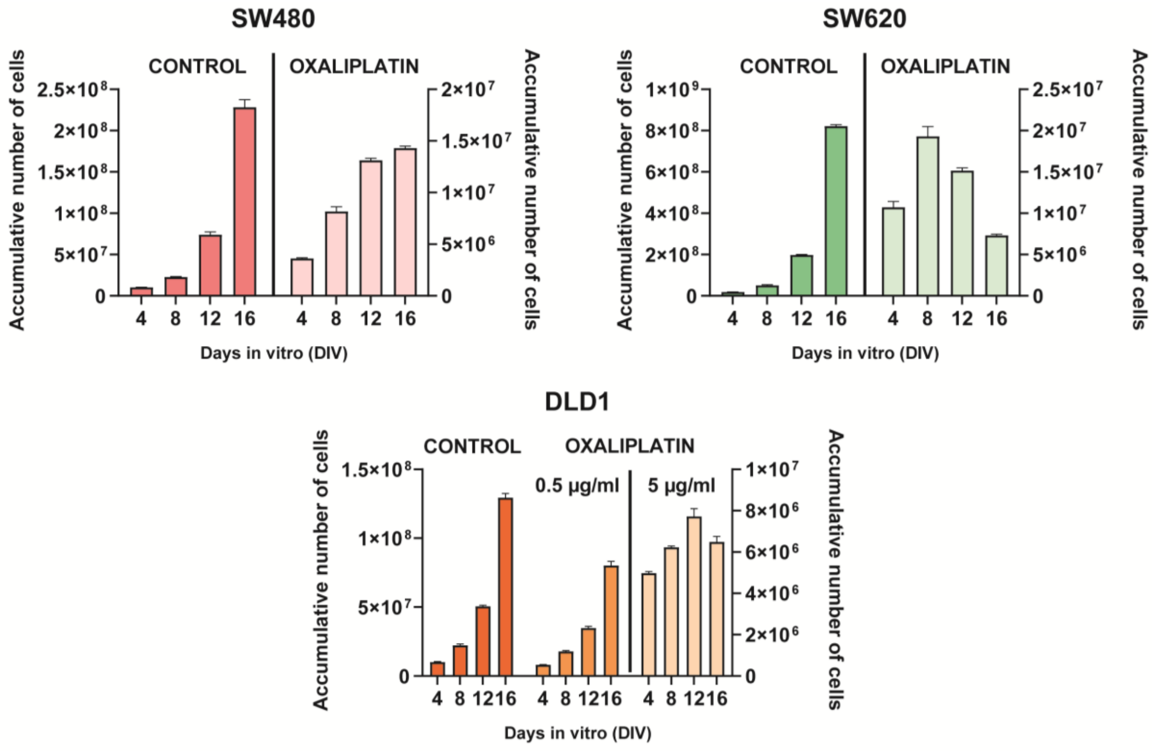

Figure R30. Growth pattern of non-treated and previously oxaliplatin treated cells in SW480, SW620 and DLD1 cell lines. Cells were treated for two weeks at a concentration of 0.5 µg/ml oxaliplatin. A concentration of 5 µg/ml was also used in the DLD1 cell line because of its known resistance to drugs. The accumulative number of cells for each passage and condition was plotted. Note a different scale in the left and right axis. The left axis corresponds to SW480 control, SW620 control, DLD1 control and DLD1 0.5 µg/ml oxaliplatin conditions The right axis corresponds to SW480 0.5 µg/ml oxaliplatin, SW620 0.5 µg/ml oxaliplatin and DLD1 5 µg/ml oxaliplatin conditions Three measures were carried out.

We carried out viability assays to evaluate if the cell lines were more resistant once treated with oxaliplatin. There were significant differences between non-treated and previously oxaliplatin treated cells in all cell lines. The cell viability percentages were significantly higher in previously oxaliplatin treated cells when compared to control non-treated cells on all the studied cell lines (Figure R31A), which means that they are more resistant to oxaliplatin. In order to discard the possibility of presenting general resistance, another chemotherapy drug was also evaluated. The results in Figure 31B show that non-treated and previously oxaliplatin treated cells presented no significant differences in relation to the response to irinotecan. All these results lead us to conclude that previously oxaliplatin treated cells are resistant to the same drug to which they were exposed (oxaliplatin) but not to others (irinotecan).

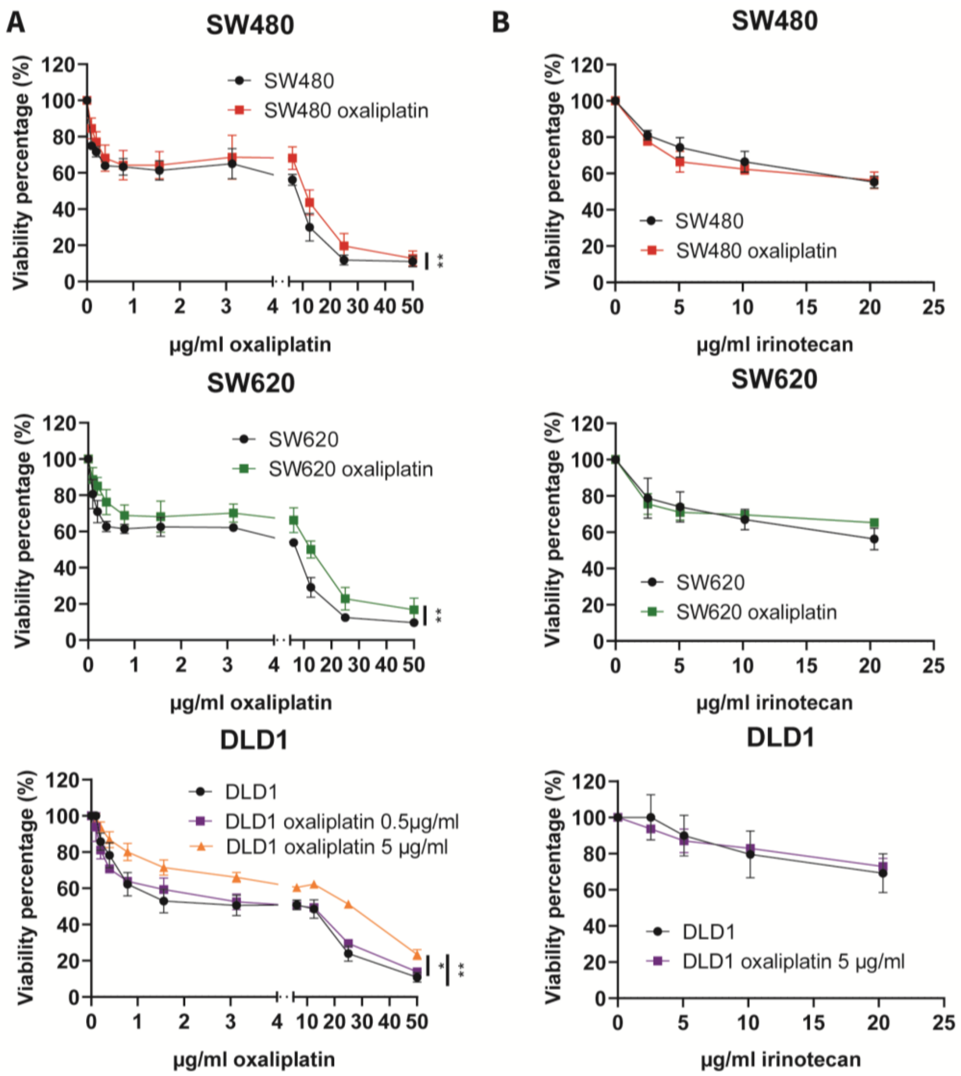

Figure R31. Evaluation of resistance to chemotherapy of SW480, SW620 and DLD1 cell lines after treatment with 0.5 or 5 µg/ml of oxaliplatin for two weeks. (A) resistance to oxaliplatin was evaluated in a range of 0 to 50 µg/ml (B) resistance to irinotecan was evaluated in a range of 0 to 20 µg/ml. Differences were evaluated by Wilcoxon test. The probability is indicated by * $p<0.05$ and ** $p<0.01$. N=3

We have already shown that TROP2 was the biomarker with higher expression in DLD1, with this line being the most resistant to oxaliplatin. Moreover, we have demonstrated that cells previously exposed to oxaliplatin were more resistant than non-exposed cells. Now we want to test whether the expression of *TACSTD2* has increased its expression or not in previously exposed cells to evaluate the role of TROP2 in the resistance to oxaliplatin. Figure R32 shows the normalized expression of *TACSTD2* versus *β-ACTIN* (reference gene). The SW480 and SW620 cell lines did not present *TACSTD2* expression in non-treated nor treated cells. However, significant differences were found between previously oxaliplatin treated DLD1 cells (at both doses: 0.5 and 5 µg/ml oxaliplatin) and non-treated cells.

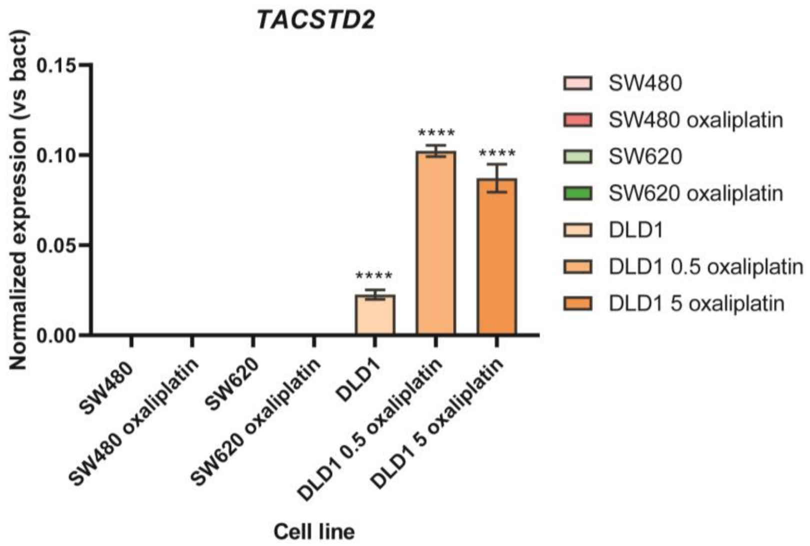

Figure R32. *TACSTD2* expression of SW480, SW620 and DLD1 previously oxaliplatin treated and non-treated cells with 0.5 or 5 µg/ml of oxaliplatin for two weeks. *β-ACTIN* was used as a housekeeping gene and data were normalized to it. One-Way ANOVA and Tukey´s multiple comparison test was carried out to evaluate differences between cell lines for each biomarker (**** p< 0.0001). N= 3.

We have shown a significantly higher expression of TROP2 (or *TACSTD2*) in the DLD1 cell line when compared to the SW480 and SW620 cell lines, being a more resistant to chemotherapy cell line. Moreover, we have demonstrated that previously oxaliplatin exposed cell lines are more resistant to oxaliplatin and, in the case of DLD1, present higher expression of *TACSTD2*. All these results together show a correlation between the expression of this protein and a drug-resistant phenotype, a promising result for predicting the response to oxaliplatin in patients. However, further experiments to confirm the role of TROP2 in resistance were carried out.

4.2.3.2. Correlation between gene/protein expression of TROP2 and resistance to chemotherapy

In the preliminary study, we discovered that the expression of TROP2 (*TACSTD2* gene) correlated with the resistance of the cell lines to oxaliplatin. We measured *TACSTD2* expression by RT-qPCR and protein expression by flow cytometry in three cell lines. In this section we broadened the scope of the study to include two additional cell lines, HCT116 and HT29. HCT116 is a chemotherapy sensitive cell line and HT29 is a chemotherapy resistant cell line. Moreover, we analyzed the resistance to chemotherapy by the new growth-rate based method implemented in this study.

Figure R33 shows that the expression of TROP2 correlated with the resistance of cell lines measured by the new growth rate parameter proposed in section 4.2.2 of this thesis, IC_{rmed}). This parameter was selected for analysis because HT29, a highly resistant cell line, does not reach r_{IC50} or r_0, highlighting the importance of discovering this new parameter. A higher value for $2^{-\Delta\Delta CT}$ or percentage of protein, indicates a greater expression of the gene and protein, respectively. Similarly, a higher IC_{rmed} value implies higher resistance. So, the most resistant cell line was the HT29 cell line, which was also the one that expressed the highest levels of TROP2 at both the protein and gene levels. Then, DLD1 followed by SW480 were the ones that presented an intermediate TROP2 expression which also correlated with the IC_{rmed}, being the next more resistant cell lines. Finally, HCT116 followed by SW620 were the ones that least expressed TROP2 and presented the lowest value for IC_{rmed}, though being more sensitive to oxaliplatin.

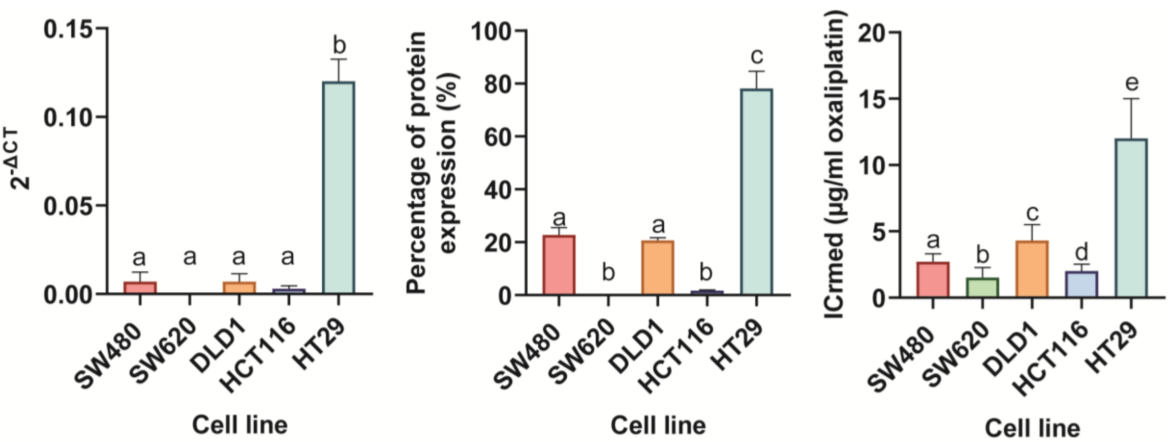

Figure R33. Correlation between TROP2 expression and resistance to oxaliplatin. The expression was measured both at the gene level (by RT-qPCR) and protein level (by flow cytometry). Resistance to oxaliplatin was measured with the previously described parameter IC_{rmed}. Means and standard deviation are represented in bar plots. ANOVA and Tukey´s multiple comparison test was carried out to evaluate differences between cell lines for each condition ($p<0.05$). N=3 for gene expression, N=5 for protein expression and IC_{rmed} was calculated with 1,000 iterations from three independent MTT experiments.

This broadened study confirms the preliminary findings, reinforcing TROP2 as a potential biomarker of chemotherapy resistance, as both the gene and protein expression correlate with the cell line´s resistance. Although promising results were obtained, further studies need to be performed to confirm that. That is why we checked whether chronic treatment to oxaliplatin changed the expression of TROP2 and whether silencing the gene changed resistance patterns.

4.2.3.3. Chronic treatment to oxaliplatin and TROP2 expression

To check whether this correlation could assign TROP2 as a good biomarker for resistance, we performed a chronic assay. Cells were treated with oxaliplatin at a concentration of 10 µg/ml for 2 months. A higher concentration of oxaliplatin and an extended assay duration were used in this experiment compared to those employed in the preliminary assay. SW620 and HCT116 cell lines died as they are the more sensitive cell lines, confirming the results obtained in Figure R29. A viability assay was carried out in the rest of the cell lines to check whether the treated cells have developed resistance or not. Figure R34 shows the IC_{rmed} values for SW480, DLD1 and HT29 cell lines. The treated cell lines were compared to a control (without treatment). Significant differences were found in SW480, DLD1 and HT29 cell lines, meaning that all of them developed resistance after the chronic treatment with oxaliplatin for 2 months.

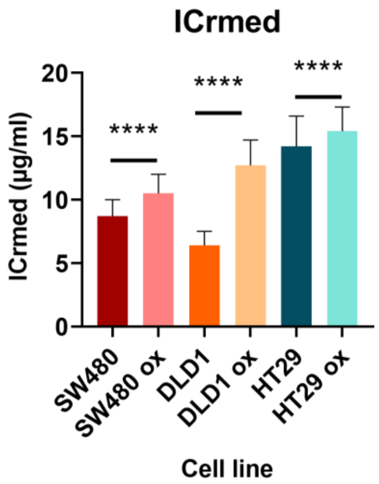

Figure R34. Evaluation of resistance to oxaliplatin of SW480, DLD1 and HT29 cell lines after treatment with 10 µg/ml of oxaliplatin for two months. A range of 0 to 50 µg/ml of oxaliplatin was used. Untreated cells were used as control. Differences were evaluated by the Student T test. The probability is indicated by **** $p<0.0001$. N=1000.

Furthermore, the expression of TROP2 was measured both in the gene and protein levels. The expression of *TACSTD2* was measured by RT-qPCR and the $2^{-\Delta\Delta CT}$ method was employed to analyze the data. Since this method compares the sample to a control and normalizes with a housekeeping gene, the treated cell lines were related to non-treated cell lines and normalized using *β-ACTIN*. That means that any expression which results in 1 (red line) means no changes in expression. Any expression above 1 means an increase in expression and any expression below it means a decrease in expression.

Figure R35A shows that the SW480 and DLD1 treated cells overexpressed *TACSTD2.* A higher increase was detected in the SW480 cell line (92% increase) when compared to DLD1 cell line (16% increase). This could be explained by the lower basal expression levels of SW480, which allows for a greater potential to increase expression (Figure R35B). We inverted the Y-axis to facilitate visualization, as higher C_T values correspond to lower expression levels. Although HT29 showed a decrease in expression of *TACSTD2,* the C_T analysis shows a higher basal expression of *TACSTD2* when compared to the rest of conditions, which means that it is a cell line that highly expresses *TACTSD2.* Further analysis is found in the Discussion section – 5.2.3.

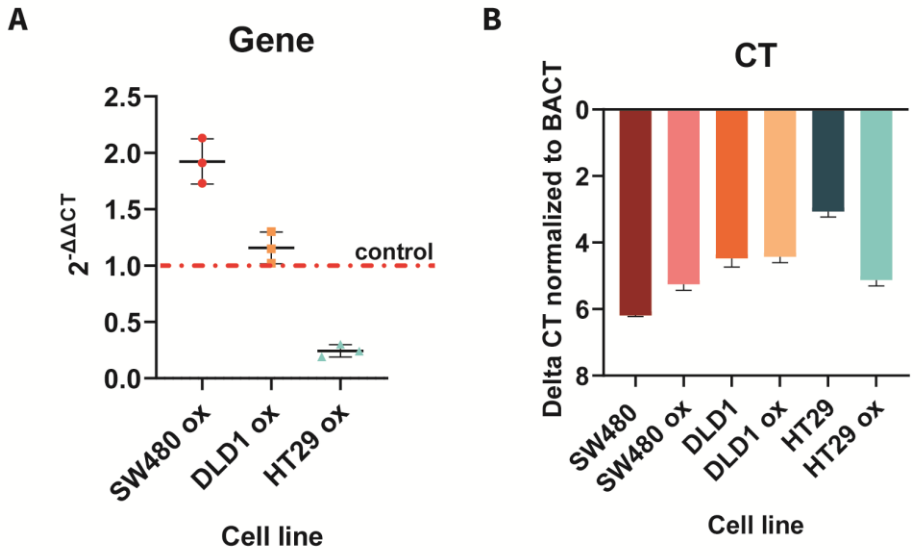

Figure R35. Gene expression of TACSTD2 in SW480, DLD1 and HT29 previously treated cells with 10 µg/ml of oxaliplatin for two months. RNA was extracted from cell lines and retrotranscribed prior to performing a qPCR. *β-ACTIN* was used as a housekeeping gene and each treated cell line was related to its control (untreated cell line). (A) Results were analyzed by the $2^{-\Delta\Delta CT}$ method. Any expression above 1 (red line) means an increase in expression in treated cells when compared to the control untreated cells. Any expression below 1 means a decrease in expression in treated cells when compared to the control untreated cells. (B) C_T values were also plotted (a higher value means less expression). N= 3.

We subsequently measured differences in protein expression, both total protein by confocal microscopy and membrane protein by flow cytometry. We found that TROP2 was overexpressed in terms of total protein but differences were not found at membrane level (Figure R36). In fact, HT29 showed a decrease in membrane protein expression that will be discussed in the next section.

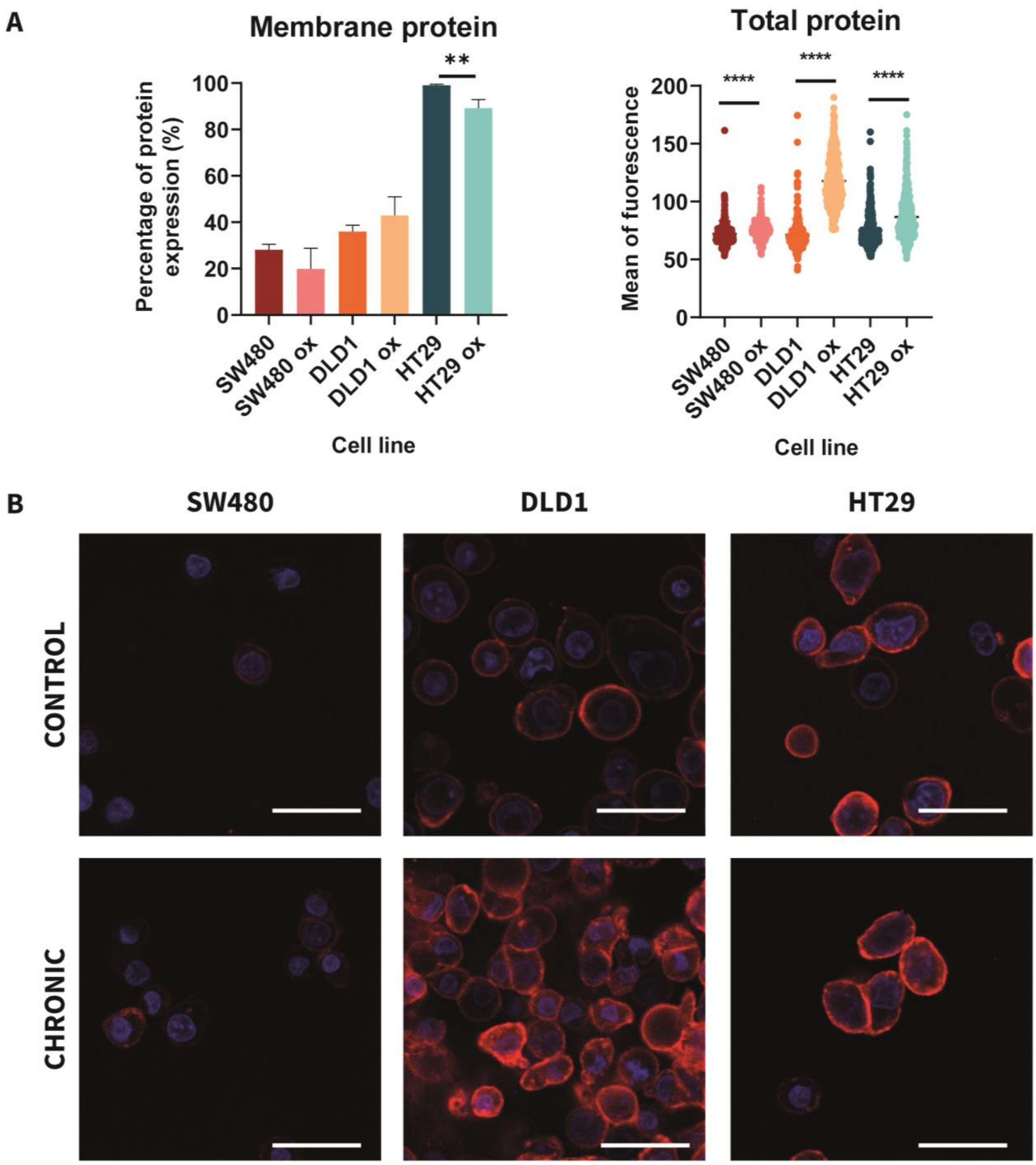

Figure R36. Protein expression of TROP2 in SW480, DLD1 and HT29 previously treated cells with 10 µg/ml of oxaliplatin for two months. (A) The protein was measured both membrane protein by flow cytometry (left) and total protein by confocal microscopy (right). (B) Representative confocal images for each condition. The scale bar represents 50 µm. Differences were evaluated by the Student T test. The probability is indicated by ** $p<0.01$ and **** $p<0.0001$. N=3.

4.2.3.4. *TACSTD2* silencing and resistance evaluation

Finally, we also carried out some silencing experiments. We transfected siRNA (RNA molecule that interferes with gene expression) for *TACSTD2* and siRNA negative (control that does not bind any sequence). We confirmed silencing measuring protein levels by flow cytometry. As shown in Figure R37, the expression was depleted at 24 hours post-transfection and maintained until 72 hours post-transfection.

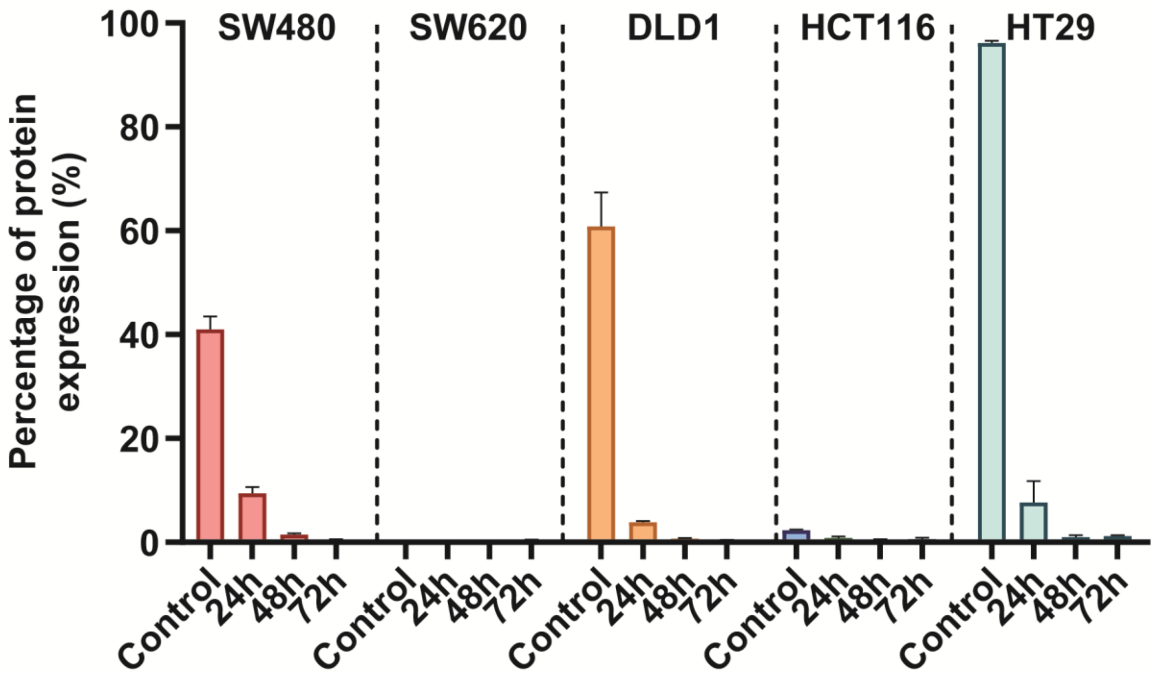

Figure R37. Protein expression of TROP2 in *TACSTD2* silenced cell lines. Protein expression was quantified by flow cytometry after 24, 48 and 72 h post-silencing. ANOVA and Dunnett´s multiple comparison test (comparing all conditions to control condition) was carried out and all of them were significant ($p<0.0001$). Significance is not shown in the graphic. N=3

Then, we conducted viability assays to check whether the *TACSTD2* silenced cells were more sensitive to oxaliplatin (Figure R38). If *TACSTD2* is involved in resistance, silencing the gene would render the cells more sensitive. In SW620 and HCT116 cell lines we did not appreciate differences, as expected because of their low/null expression of TROP2. Both the control and the silenced cells showed no significant differences. However, the results obtained for the rest of the cell lines were controversial, as the silenced cell lines were more resistant than the control ones.

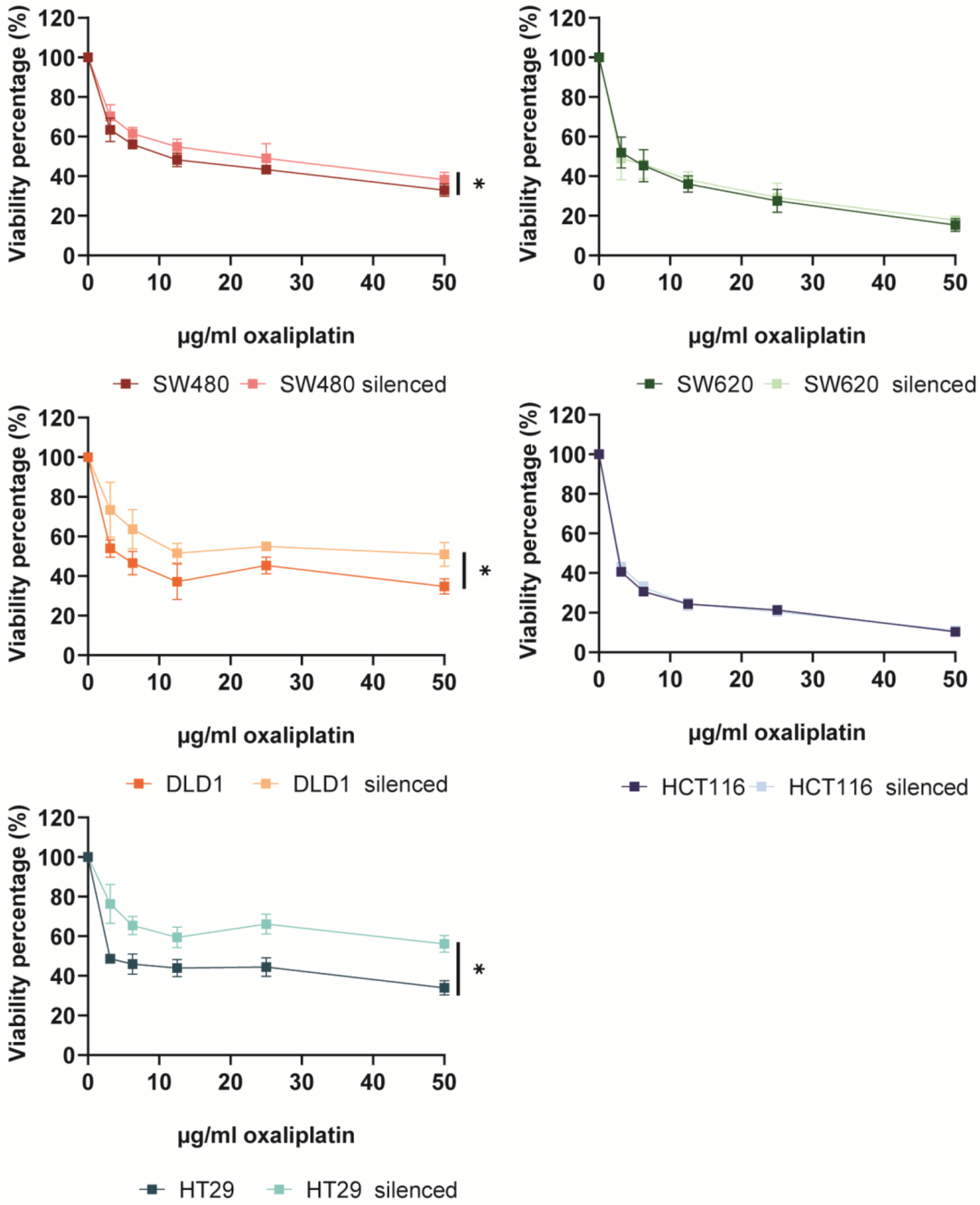

Figure R38. Evaluation of resistance to oxaliplatin of SW480, SW620, HCT116, DLD1 and HT29 cell lines after silencing of TACSTD2 (related to the control of each condition). A range of 0 to 50 µg/ml of oxaliplatin was used. Non-silenced cells were used as control. The viability was related to the concentration 0 of oxaliplatin for each condition. Differences were evaluated by Wilcoxon test. The probability is indicated by * $p<0.05$. N=3.

We explored this result in greater detail due to its controversial nature and found that the growth rate in control cells (without chemotherapy treatment) decreased in silenced cells when compared to non-silenced control cells (Table R5). It is important to note that the decrease correlated with TROP2 expression as the cell lines with higher initial expression of TROP2, were the ones with a higher decrease of the growth rate after silencing. This means that TROP2 is implied in proliferation and viability assays are limited to study resistance in these cases.

Table R5. Effective growth rate in silenced and non-silenced control cell lines. SW480, SW620, DLD1, HCT116 and HT29 cell lines were employed to calculate the effective growth rate by bootstrap method.

Cell line	Condition	Effective growth rate (days^{-1})
SW480	Control	0.38 ± 0.04
	Silenced	0.30 ± 0.06
SW620	Control	0.51 ± 0.08
	Silenced	0.53 ± 0.06
DLD1	Control	0.45 ± 0.06
	Silenced	0.32 ± 0.06
HCT116	Control	0.67 ± 0.06
	Silenced	0.58 ± 0.07
HT29	Control	0.52 ± 0.03
	Silenced	0.31 ± 0.07

To check whether this was the cause of the increase in viability in Figure R35, we related both the non-silenced control cells and the silenced cells to the non-silenced control cells without oxaliplatin. As shown in Figure R39, the effect was nullified, with no significant differences detected. A lower viability percentage was observed under control conditions in silenced cells compared to non-silenced control cells (e.g., 80% in DLD1). However, no significant differences were found at other oxaliplatin concentrations, suggesting that the observed effect was due to inhibited proliferation at 0 μg/ml of oxaliplatin.

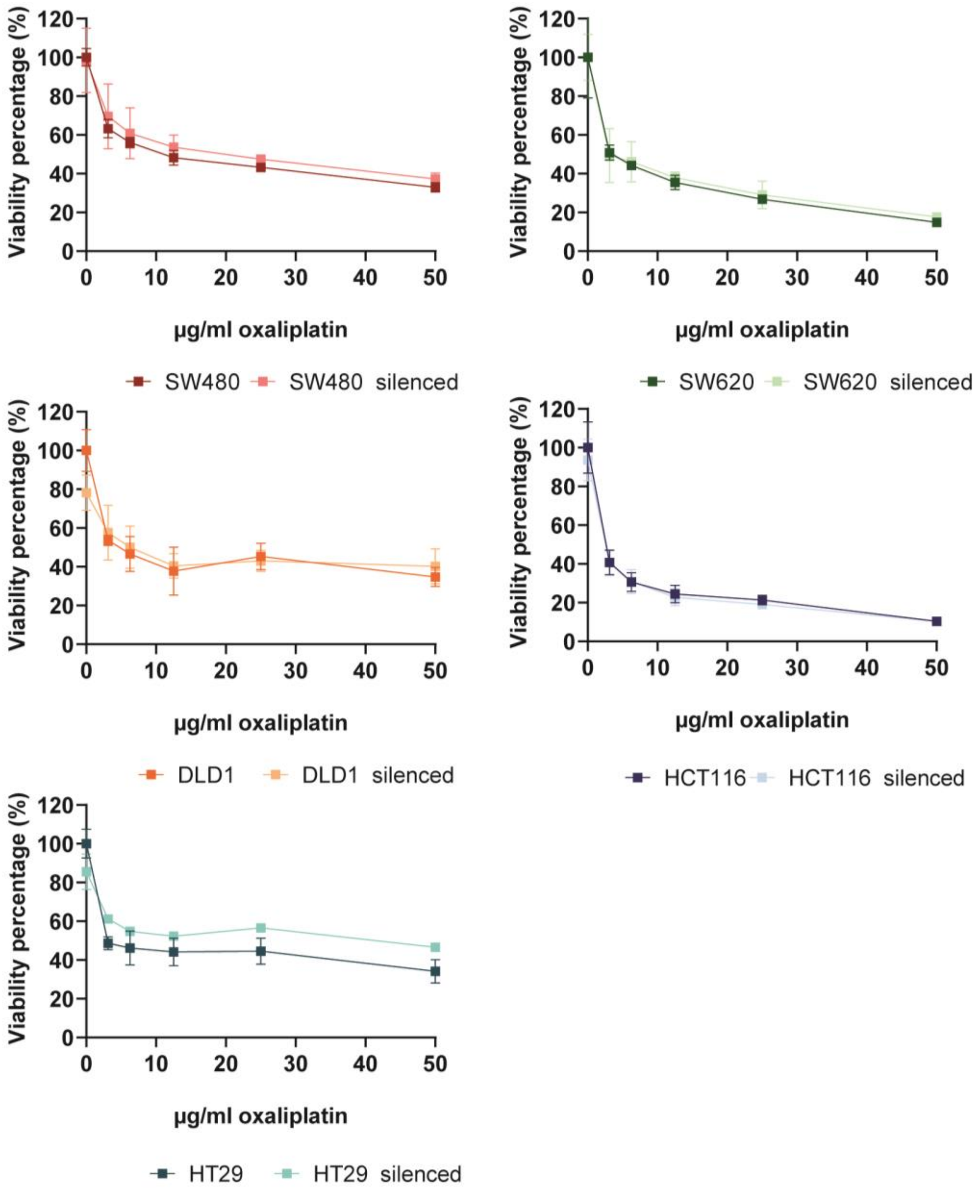

Figure R39. Evaluation of resistance to oxaliplatin of SW480, SW620, HCT116, DLD1 and HT29 cell lines after silencing TACSTD2 (all conditions related to the control). A range of 0 to 50 µg/ml of oxaliplatin was used. Non-silenced cells were used as control. The viability was related to the concentration 0 of oxaliplatin for the control of each cell line, both the treated and non-treated cell lines. Differences were evaluated by Wilcoxon test. The probability is indicated by * p<0.05, ** p<0.01 and *** p<0.001. N=3 except N=2 in HT29 cell line.

In conclusion, this PhD thesis presents the development of a mathematical model for prediction of relapses in CRC patients. Additionally, we focused on evaluating resistance to chemotherapy. First, we generated patient-derived tumors by injecting enriched PBMCs into NSG mice. We also developed a growth rate-based method to evaluate chemotherapy efficacy. Finally, we identified TROP2 as a potential biomarker for resistance detection.

5. Discussion

The purpose of this PhD thesis is to assess the early detection of relapses in CRC patients and the study of efficacy and resistance to chemotherapy, both being important issues in management of CRC. In this study we have focused on CRC due to its high rate of incidence, mortality and recurrence. While the primary tumor is effectively managed in most cases, approximately 20 to 40% of patients experience recurrence within the first two to three years post-treatment [235,236]. A significant number of patients experience multiple recurrences, with the mean time to recurrence estimated at 2 years [237,238]. On the one hand, this provided us with the opportunity to collect positive samples for relapses during the thesis project, making it viable. On the other hand, all these factors together highlight the importance of rigorous postoperative surveillance and effective strategies for early detection of recurrence, enabling timely treatment and improving patient prognosis.

5.1. Early detection of relapses

Nowadays, relapses are detected in clinical follow-up through diagnostic imaging by CT or PET and laboratory tests such as the measurement of the expression of biomarkers CEA and CA19.9 in blood. Patients undergo regular follow-up visits every 3-6 months for the first 2-3 years and every year after that. Imaging techniques are useful to detect tumors from 5-10 mm diameter but may miss smaller lesions.

Regarding laboratory tests, CEA and CA19.9 are biomarkers commonly used in clinics. CEA is an oncofetal glycoprotein expressed at high levels during fetal development and significantly reduced in adulthood. However, its levels can rise in pathological conditions such as CRC. Although it is not specific to CRC, elevated preoperative levels correlate with poor prognosis and advanced tumor stage [239]. It has also demonstrated to be useful for detection of recurrences [240]. Moreover, the diagnostic accuracy of CEA for detecting recurrence depends on initial CEA level [241]. However, CEA may also increase by other factors such as cigarette smoking, alcoholism or chronic inflammatory disease [242,243].

In relation to CA19.9, it is a tumor marker primarily associated with gastrointestinal cancers including CRC. Several studies have demonstrated its prognostic value in CRC [244,245]. Nevertheless, patients diagnosed with liver cirrhosis, endometriosis or bronchiectasis also show increased levels of CA19.9 [246].

The sensitivity of these two biomarkers is estimated to be 87% and 96%, respectively; however, their specificity is significantly lower, 36% and 23% [57]. This low specificity leads to a high rate of false negatives, meaning that while elevated biomarker levels can reliably predict poor prognosis, many patients with low biomarker levels may still develop cancer or experience relapse, remaining undetected.

To address these challenges, innovative technologies such as liquid biopsies and AI are emerging in early detection and monitoring of cancer. Liquid biopsies provide a non-invasive method to detect circulating tumor DNA, RNA, and other biomarkers, enabling real-time and dynamic insights into disease progression [247]. AI-powered algorithms further enhance diagnostic accuracy by analyzing vast and complex datasets to uncover patterns that might be missed by traditional diagnostic approaches [248]. By integrating complex data, these algorithms can provide more precise and personalized insights. Together, these technologies hold the potential to early detect relapses, minimize false negatives, and significantly improve patient outcomes.

5.1.1. Development of a new diagnostic method for early prediction of recurrence

In this PhD thesis, we developed OncoRec, a novel method for early relapse prediction. It consists of the measurement of 13 biomarker's expression in blood from patients previously diagnosed with CRC and the generation of a mathematical model by ML, a subfield of AI, which predicts relapses.

CSCs are responsible for relapses as they escape from the tumor and initiate new tumor formation, either at the original site or in distant locations. This occurs when CSCs intravasate into the bloodstream and migrate to distant organs, where they establish a new tumor. The detection of CSCs in blood from CRC patients is a promising strategy for early detection of recurrence but is challenging due to their phenotypic similarity to normal stem cells and the lack of a single marker [249,250]. Consequently, identifying CSCs often requires a combination of multiple biomarkers to accurately distinguish them from other cell populations. Multiple studies have already demonstrated that the combination of several biomarkers enhances specificity and accuracy, providing more reliable identification of CSCs [112,172,251,252].

In order to detect CSCs in the blood from previously diagnosed CRC patients, we searched for biomarkers in the literature. In contrast to CEA and CA19.9 biomarkers, which are the ones currently used in clinics, we selected biomarkers that were more specific to CRC and CSCs as described in the literature. So, they are implied in resistance, proliferation, migration, self-renewal and angiogenesis. We selected a total of 13 biomarkers from the literature and measured their expression in CRC cell lines. All the biomarkers were expressed in at least one of the cell lines, except for *CD34*. One possible explanation for the low expression of this biomarker is that CD34 is a marker for hematopoietic progenitor cells, typically expressed in suspension cells rather than in the adherent cell lines we used [253]. Regarding the comparison of *CD44* and its variant 6, *CD44v6*, a higher expression of *CD44* was observed in all the cell lines, which is consistent with the fact that CD44 is a more general biomarker with expression in non-tumoral cells. On the contrary, a more specific role in metastasis and invasion has been described for CD44V6 [254]. It is worth noting that in most cases, the expression was different among cell lines. This can be explained as they present different genetic backgrounds leading to the same variability as is seen in patients [255,256].

Once the expression of biomarkers was studied in CRC cell lines, we moved on to human samples. We measured the expression of them both in patients and control (healthy people) samples. The expression in controls was null or low for all the biomarkers except for CD44 that was highly expressed, as happened in cell lines because of its expression in other non-tumoral cell types. It was first described in 1980 as a surface molecule on T-lymphocytes, cortical thymocytes and granulocytes [257]. Regarding the rest of the biomarkers, the results showed that the expression was increased in some patients maintaining low its expression in control samples. This was the first encouraging result, as the biomarker levels demonstrated potential not only for distinguishing patient samples from controls but also for differentiating between patient samples themselves. Most patient samples presented expression rates similar to the control, but some of them expressed higher levels of expression of the biomarkers. This finding highlights the importance of investigating whether these differences correlate with clinicopathological characteristics as they may serve as indicators for relapse.

Before addressing this by the generation of the mathematical model for relapse prediction, we evaluated whether there were differences between patients who relapsed and those who did not, considering several variables. Regarding sex, BMI, smoking conditions and family background, they were not correlated with a higher risk to relapse in our population of study.

BMI is of especial interest in CRC as an elevated BMI has been related to poorer prognosis and higher mortality [258]. This may be explained by the increased difficulty of surgical treatment. Open surgery is often the preferred technique, as laparoscopy can be challenging due to mesenteric thickening [259]. Additionally, longer operative times and greater blood loss are common, contributing to more frequent side effects in these patients [260].

Moreover, a higher likelihood of recurrence has also been demonstrated in CRC patients with a high BMI [261]. However, our population of study did not show significant differences in relapse rates based on BMI, which may be attributed to the fact that a single BMI value was recorded from the medical records. Weight data can change throughout the course of the disease, as patients often experience weight loss due to chemotherapy or disease progression. The absence of significant differences in our analysis is likely due to BMI not being consistently measured at the same timepoint for all patients. To more accurately analyze the correlation, it would be essential to record weight both at the time of diagnosis and during the progression of the disease, rather than relying on a single value.

We also evaluated whether our population, consisting of CRC patients, presented a higher BMI when compared to the general population. For that purpose, we categorized the patients in obese, overweight, normal weight and underweight and compared the percentage of people in each category to the one for the general population (Table D1). We did not find differences when studying all the categories together (Chi Square test 4.7, p-value 0.2) as overweight and underweight proportions were conserved. However, the percentage of obese people increased, while the percentage of people with normal weight decreased when comparing our population of study to the general population (Chi Square test 4.7, p-value 0.03). Thus, obesity is more prevalent among CRC patients in our population of study compared to the general population, as expected, given that high BMI has been associated with an increased risk of developing CRC [262]. Although the precise mechanism is not fully understood, some factors such as the presence of reactive oxygen radicals, increased in obesity, contribute to cancer development [263].

Table D1. Comparison of BMI variable in the population of study and the normal population. Data obtained from the 2020 European Healthy Survey in Spain.

BMI category	Population of study (%)	General population (%)
Obesity (BMI>30 kg/m^2)	26.5	18.7
Overweight (25<BMI<30 kg/m^2)	37.0	37.1
Normal weight (18.5<BMI<25 kg/m^2)	32.8	40.5
Underweight (BMI<18.5 kg/m^2)	3.7	3.7

In relation to the smoking condition, although a slight tendency to higher relapse rates was shown in active smokers, followed by exposed (patients who smoked in the past and are no longer smoking) and never smokers, we did not find significant differences between groups. It is probably because the active group only includes 16 patients. This can be explained as patients with cancer diagnoses are more likely to quit smoking, as it has been demonstrated that smoke cessation can improve survival rates [264]. Therefore, a larger sample size may be required to get reliable and robust conclusions. Furthermore, we have identified an opportunity to enhance the analysis by distinguishing patients who quit smoking more than 10 years before diagnosis from those with shorter cessation periods. Patients who have been smoke-free for over a decade could be classified as never smokers, while those who quit less than 10 years ago might be grouped as active smokers or categorized into intermediate groups. This distinction is supported by evidence showing that the benefits of smoking cessation in reducing the risk of cancer become markedly evident after 10 years [265]. However, these data are not registered in the clinical record as such, so we will need to ask each individual patient in future appointments.

Although *BMI, sex and smoking conditions* did not show a correlation with relapse in our study, correlations were found between relapse and tumor-related factors such as stage, localization and differentiation grade. Advanced primary CRC (stage IV) is significantly associated with higher rates of both locoregional and distant recurrence, which is consistent with the results obtained in our population of study [266].

Moreover, relapses are more common in patients with rectal cancer than in those with colon cancer [267], a trend also observed in our study population. It is worth noting that the transverse colon showed a higher probability of relapse as well. However, the sample size for this group was very small (only 7 patients). A larger cohort will be required to confirm this.

In addition, poorly differentiated tumors also presented a higher probability of relapse [268], which is in accordance with the results obtained for our population of study. Nevertheless, moderately differentiated tumors exhibited a lower relapsing rate compared to well differentiated tumors. This finding can be attributed to the fact that 74.3% of the moderately differentiated tumors in our study were from patients classified as N0M0 stage, which may influence the observed relapse proportion.

Furthermore, we also analyzed the influence of the clinicopathological characteristics on the expression of biomarkers. For that purpose, we took advantage of ML capacities. Among the key advantages, ML presents the capacity to process large-scale and multidimensional data, enabling the identification of patterns and correlations between multiple biomarkers that are otherwise imperceptible to human analysis. Furthermore, ML offers a faster and more efficient approach, substantially reducing errors and enhancing reliability of CSC detection [269]. As biomarkers' expression varies between samples, we focused on clinicopathological characteristics that may also differ, such as treatment and relapse.

In relation to treatment, we evaluated whether the type of treatment (*Sx alone, Sx + other treatment or other treatment without Sx*) influenced the percentage of expression of biomarkers. We also evaluated if the biomarkers' expression varies if the patient has received any treatment from the previous sample or not. In both cases, correlation was not observed. However, we found that treatment in the moment of sample extraction could influence the biomarkers' expression as previously described in the literature for other biomarkers [270,271]. This is why we included the variable *status* (*treatment, follow-up, discharge*) in the prediction model for relapses as an input together with biomarker's expression and compare if it improves the model when compared to the biomarker's expression alone as an input.

ML technology represents a significant milestone in the development of advanced algorithms for detecting CSCs, thus, to predict relapses. We used several ML models for prediction, such as SVM, kNN, Logistic Regression, Random Forest and Decision Tree, obtaining better results when the *status* variable (*treatment, follow-up, discharge*) was used in combination with the biomarker's expression. Regarding the different models, Logistic Regression exhibited the highest sensitivity and specificity collectively. The rest of the models presented a slightly higher sensitivity but significantly lower specificity, which means the negatives are highly predicted, but positives were only truly predicted in 60-70% of the cases. During the project, we were focused on improving both parameters altogether. In fact, if we had to choose, the true positive rate is the most important one for us as it predicts the patients that will relapse. This is the reason why we selected the Logistic Regression model, obtaining a sensitivity of 82.5% and a specificity of 82.6%.

These findings demonstrate that the model can effectively predict relapses, which is a significant advancement since relapses greatly impact the lives of patients with CRC. Early prediction of relapse could influence doctors' decisions regarding patient treatment, enabling timely intervention to address the tumor before it becomes detectable through imaging tests, where its size is typically more substantial. This early detection may also reduce the dosage and/or number of chemotherapy cycles required, thereby minimizing the side effects experienced by patients [272]. Furthermore, this approach could lower healthcare costs associated with chemotherapeutic agents, as patients with a low predicted likelihood of relapse would avoid unnecessary preventive treatments.

In summary, although each biomarker exerts a different influence on the model, we advocate for the combined use of multiple biomarkers, as no single marker is sufficient on its own. The predictive model demonstrates strong potential in identifying complex patterns associated with relapse, offering a valuable tool for anticipating patient outcomes. Its ability to provide early insights could significantly enhance clinical decision-making, enabling more personalized and timely interventions. This promising approach sets the stage for a comparison with other predictive methods in the following section.

5.1.2. Comparison with an FDA-approved method and prior research results

Table D2 shows a comparison of sensitivity and specificity of our new proposed method (OncoRec) compared to the method approved by the FDA (CellSearch®) and the prior research results including 4 biomarkers and a ROC curve analysis of the data (CMT). Note that the last methods analyzed 44 samples and the one proposed here analyzed a total of 389 new samples.

The results show that CellSearch® method presents the highest specificity but fails in the sensitivity, which means that it will truly detect the patients that will not relapse but fails to detect the patients that will relapse. This method uses the expression of only one biomarker, EPCAM, together with a classification of CD45−/cytokeratin+/DAPI+ cells, describing an epithelial phenotype and not a tumor-specific phenotype. The mesenchymal tumor cells remain undetected by EPCAM-based enrichment methods [273].

CellSearch® is only validated by the FDA for metastatic patients, and we tested here whether it could be used for early detection in an heterogeneous population, including different stages. That is why this method only detected the negative samples for relapse, as the absence of expression determined correctly a good prognosis. However, it failed in the positive samples, probably because the use of a biomarker is not enough for detecting CSCs in early stages, and this method is only useful for advanced stages.

For that reason, a previous thesis carried out in the laboratory [213] tried to solve this problem and proposed a new method (CMT). This method included four biomarkers, and a ROC analysis was carried out to distinguish patients that will relapse and those that will not. Contrary to the results from the previous method, this one improved the sensitivity, that is, that it could truly detect the positive samples (relapse) in 96.8% of cases. However, it did not have the capacity to detect the false samples. This is probably due to the low number of negative samples used in this preliminary study (only 13 negative samples).

Then, we aimed to improve the method by incorporating the measurement of expression of 13 biomarkers and a more complex analysis of the data with ML. We found a sensitivity of 82.6% and a specificity of 82.4%, which improves both parameters simultaneously.

Moreover, we achieved an NPV of 96.1%, indicating that nearly all predicted negative samples were truly negative. This is particularly important, as the method can support clinical decision-making by helping identify patients who are unlikely to relapse and may not require unnecessary treatment, thereby minimizing secondary adverse effects and reducing healthcare costs.

Table D2. Comparison of sensitivity and specificity of different methods. CellSearch® is the FDA approved method, CMT corresponds to the preliminary study carried out in a previous PhD project and OncoRec is the method developed in this PhD project.

Method	Sensitivity (%)	Specificity (%)	NPV (%)
CellSearch®	48.4	92.3	42.9
CMT	96.8	46.2	85.7
OncoRec	**82.6**	**82.4**	**96.1**

We also cropped our data into two databases, one for training and the other for prediction, to perform internal validation of the model. The results showed that 72.7% of patients who will relapse were correctly identified and 82.5% of patients who will not relapse were also correctly identified. These are promising results although a slight decrease was observed due to the reduced number of samples.

So, we have developed a new method, OncoRec, which predicts relapses with a high sensitivity and specificity.

5.1.3. Other methods for early prediction of relapses

Several studies have been conducted in recent years to develop new methods for the early prediction of relapses. Table D3 presents a comparative analysis of studies focused on this objective in CRC. Most of these studies rely on detecting cell-free circulating tumor DNA (ctDNA), while one study explores the use of ML algorithms to analyze clinical characteristics of patients.

The European Society of Medical Oncology (ESMO) provided recommendations on the use of ctDNA assays for cancer patients in a special article published in 2022 [274]. The article pointed out that while multiple studies report excellent specificity values (≥90%), sensitivity values are often low (≤50%). In contrast, the studies included in Table D3 demonstrated high values for sensitivity and specificity, highlighting its potential as a predictive tool for relapses.

Regarding the use of ML algorithms for relapse prediction, the study showed in Table D3 achieved a specificity of 70%. This result may be due to the use of clinical characteristics, which may limit the model's predictive power.

Table D3. Comparative analysis of methods for early prediction of relapses in CRC.

Population of study	Number of patients	Technique	Sensitivity (%)	Specificity (%)	Reference
Stage II CRC	178	ctDNA	90.2	78.6	[275]
Stages I-III CRC	94	ctDNA	88.7	82.4	[276]
Stages II-III CRC	102	ctDNA	82.6	94.1	[277]
Stages I-IV CRC	309	ctDNA	64.6-75	94.8-94.9	[278]
Colon	396	ML	70	-	[279]

Moreover, the studies based on detecting ctDNA emphasize the importance of serial sampling (using multiple consecutive samples), as relying on a single sample did not yield satisfactory results. Note that Table D3 shows the best results for each study. As an example, Tarazona et al [276] obtained a 47.1% sensitivity in the first individual sample and 88.7% when considering serial sampling. Single-sample approaches often result in a longer interval between sampling and relapse detection. In contrast, multiple sampling at different time points increases the likelihood of capturing samples closer to the relapse, allowing for earlier and more accurate detection. This suggests that ctDNA-based methods are more effective for predicting relapses within shorter timeframes (Figure D1).

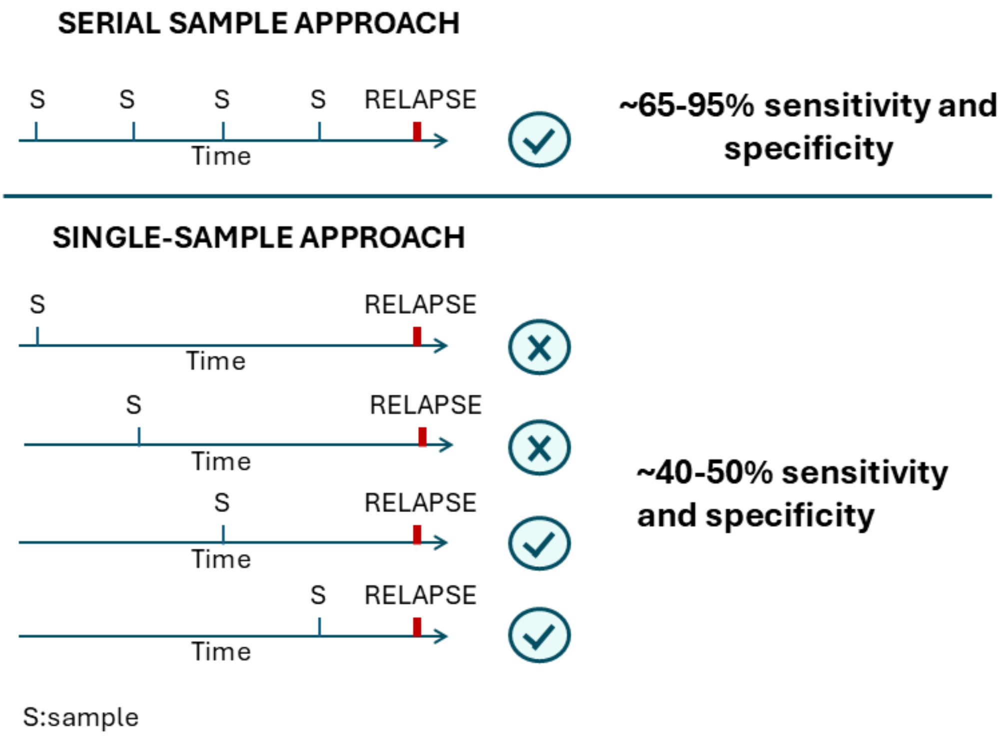

Figure D1. Schematic representation of serial-sample versus single-sample approaches. In the serial-sample approach, multiple samples from the same patient, collected at different time points, are analyzed together, yielding more robust results. In contrast, the single-sample approach, which relies on a single sample per patient, is more likely to detect cases closer to relapse while potentially missing those with a longer time to recurrence.

When comparing with the new proposed method in this thesis (OncoRec), we improved the values obtained in the ML method because we used a combination of biomarkers to detect relapses rather than clinical characteristics. Moreover, we obtained similar values to the ones for ctDNA detection but solved the limitations of the method, as we were able to detect relapses from 1 day to 17 months. Although the follow-up of the patient is recommended, the model could predict an isolated sample.

In summary, CTCs and ctDNA have shown to provide complementary information about prognosis and treatment benefit, since they contribute differentially as they refer to distinct cellular events.

5.1.4. Next steps towards clinical translation

The next steps for the project involve an external validation of the model, which requires testing with independent data. We are currently gathering additional samples from the centers collaborating on the project and will proceed with this step once enough samples are available. We also plan to refine the model by incorporating additional CRC samples and expanding its scope to include other cancer types, enhancing its cost-efficiency. Furthermore, the entire process will be automated to improve scalability and reproducibility.

Additionally, we will publish the data to gain validation within the scientific community and present our findings at conferences to increase visibility. We aim to develop a user-friendly tool, making the model more accessible.

Finally, we will design a prospective clinical trial to evaluate the model's accuracy in predicting relapses and validate its efficacy. Once validated, we intend to certify the model with regulatory agencies and work toward implementing it as a diagnostic tool in the hospitals.

5.2. Resistance to chemotherapy

The main treatment for patients with CRC is surgery followed by adjuvant and/or neoadjuvant chemotherapy depending on the stage at diagnosis. Although surgery is especially effective in this type of cancer because of the anatomical accessibility of the colon, chemotherapy is often needed. Chemotherapy triggers secondary effects such as nausea, diarrhea, neutropenia or hand-foot syndrome, causing toxicity. Moreover, cancer cells adapt to survive and proliferate despite chemotherapy, reducing treatment efficacy and causing resistance. CRC patients undergo multiple recurrences, so the correct use of chemotherapy to manage resistance to treatment is especially important as several lines of treatment may be necessary along the life of the patient. This reveals the necessity of personalized medicine and evaluation of the best treatment for each patient. The study of resistance is, therefore, essential for increasing the patients' quality of life.

5.2.1. Culture of patient-derived tumors

In this PhD thesis we have developed tumors derived from patients to study the efficacy of different treatments for each patient and contribute to this issue. Patient-derived xenografts (PDX) are preclinical models in which tumor cells obtained from a patient are injected into immunodeficient mice.

When CTCs are injected, the models are known as CDX (CTC-derived xenograft). The establishment of PDX from tumors is a technique that started in 1970s but is still used nowadays [280]. However, the establishment of CDX is challenging due to the lower success rate, which is influenced by the limited number of CTCs available, the high associated costs and the extended time required for engraftment and tumor development [281].

In this PhD project we tried to generate CDX from selected patients. We selected stage IV patients or patients with a higher percentage of positive cells for the biomarkers in order to increase the probability of success. Furthermore, we used two different immunodeficient strains: nude and NSG mice. A higher success rate has been demonstrated in NSG mice when injecting CSCs [282]. This is the reason why we moved to this strain after the first attempts, achieving tumor growth in two out of five patients. Moreover, alternative models have emerged in the last years to solve the relatively low efficacy, high costs and large time. One alternative model is the use of chicken embryos by injecting the cells over the CAM. The first generation of CDX in this model was published in 2022 [283,284] and is currently used [223,224]. We also tried to implement this strategy for our patients, although we did not achieve tumor growth. This technique requires a high level of technical expertise and meticulous optimization of various parameters, including incubation time prior to injection and the number of cells to be injected. While we have conducted numerous experiments and significantly improved the methodology, further attempts are necessary to achieve consistent results. Additionally, we are considering a future internship to gain comprehensive training and mastery of this highly complex technique.

In order to increase the probability of success in the generation of patient-derived tumors, we depleted immune cells (CD45+) to enrich PBMCs with CSCs. Confocal images confirmed the depletion although some immune cells with lower expression of CD45 were present in CD45 negative selected cells. CD45 is present in different isoforms depending on the cell type [285,286]. Moreover, there are different CD45 expressions in granulocytes, lymphocytes and monocytes [287]. Hematopoietic stem cells are also characterized by expressing less CD45 than mature stem cells [288]. So, we probably have some low-CD45 expressed granulocytes and immune precursors in the CD45 negative selected cells together with the CD45 negative cells that would be the CSCs. We used these CD45-depleted cells to inject NSG mice, successfully achieving tumor growth for two out of five patients.

The tumors were dissected and seeded to establish stable cultures, which will be valuable for studying resistance in patient-derived models. Finally, we tried to directly culture the cells derived from the patient both in M12 medium (suspension culture) and in Matrigel domes (PDO, patient-derived organoids). All these strategies were implemented after a placement in the laboratory of Dr. Frédéric Hollande in Melbourne, Australia.

They are used to generate PDO from biopsies and I had the opportunity to train in this technique in his laboratory. Subsequently, we attempted to implement this strategy in our laboratory using liquid biopsy. However, we have not yet succeeded in generating PDOs. Further attempts will be made. Regarding suspension culture, they successfully achieved it from liquid biopsy samples of their patients and showed the procedure to me [289]. Although we have not yet been able to replicate this in Spain, additional efforts are required as this technique is highly challenging and has low efficacy.

Multiple strategies were implemented, successfully generating tumors in two patients. The tumors have been dissected and are currently being cultured in Matrigel and suspension cultures, awaiting further growth. We consider that the procedure has been optimized, and further attempts will be conducted with other patients. Our efforts are now focused on the enrichment of CTCs by using different biomarkers and the optimization of the process in the *in-ovo* animal model. Moreover, CTCs can be utilized for analyses that are challenging with ctDNA, such as comprehensive molecular characterization, including whole genome sequencing, epigenomic, proteomic, and metabolomic studies.

5.2.2. New growth rate-based indices to measure resistance

Since additional time is needed to establish the cultures for use in resistance experiments, we considered using human CRC cell lines to study resistance. We first developed a new method to measure efficacy of treatments. The most widely used index to define the effects of a compound on cell viability is the IC_{50}. An important advantage of the index is its easy calculation, simply determined as the concentration at which the ratio between the number of viable treated cells and the number of viable control (untreated) cells is 50%. However, the IC_{50} as a measure of anti-cancer efficacy has been shown to have multiple weaknesses [290,291].

Among them we highlight the time-dependent nature of this index, where variations of relatively few hours to the end-point of an assay will result in a different IC_{50} value [292–295]. This is a result of the comparison of two populations (sample and control), which normally evolve over time with different growth rates [296,297].

To this end we have endeavored to find biologically meaningful parameters of cell proliferation to indicate the effects of chemotherapeutic compounds in a time-independent manner.

Cells' exponential growth can be represented as a differential equation where, rather than defining the size of a population at a given time, it is indicated as the rate at which the population increases in a particular amount of time (the growth rate) [296]. When considering short time frames, i.e., when growing is not affected by space, nutrients or any other constraints, this rate (the effective growth rate) is constant, although considerations such as density dependence mean that this value is altered across longer time frames [297]. The proposed mathematical model, after confirming the condition of exponential growth, can determine the effective growth rate of cells when exposed to a range of drug concentrations in a time-independent manner. The dose-dependent reduction in growth rates in this model allows for a more robust method of study of *in vitro* response to drugs, as the use of bootstrapping methods means that values are the result of thousands of iterations of the fit to the curve. This larger number of iterations helps to reduce the error of the fitting parameters. This method not only enables the calculation of specific indices to evaluate resistance but also provides a clear understanding of the drug effect by directly monitoring how the population evolves under a given concentration.

In order to remedy the limitations of the IC_{50} index and establish a time-independent standardized parameter for study of drug potency, we propose two new indices: IC_{r0} and IC_{rmed}, both derived from the effective growth rate. IC_{r0} is defined as the value that indicates the compound concentration at which there is an effective growth rate of zero, or the dose at which the treated cell population is neither growing nor dying. Drug concentrations above this threshold will thus result in a reduction in the population size. For its part, the other parameter that we propose, IC_{rmed}, is the value that corresponds to the drug concentration that reduces the control population's growth rate by half, or the dose at which the treated population grows at half the rate of the control untreated cells.

In conclusion, the IC_{50} index as conventionally calculated, has been used in a variety of applications in the field of oncology, including characterization of novel treatments, drug dosage calculations, and description of synergetic treatments, among others [298,299]. However, the lack of reproducibility and significant drop-off of viable treatments when moving to clinical trials emphasizes the need for new parameters, with biologically relevant interpretation and reduced error [293,300]. Our mathematical model allows for more precise comparison between cell lines, reduces the impact of negative culture conditions on the assay, and increases reproducibility by decreasing the error.

In addition, the use of effective growth rate based estimates (such as IC_{r0} and IC_{rmed}), in contrast to end-point assay based parameters, reduces the values of drug concentrations needed for an equivalent effect. Moreover, this approach yields more reliable results as it verifies an exponential growth of control non-treated cells, ensuring that the observed effect is attributable to the drug and not influenced by external factors. If any issues are detected, we are able to repeat the appropriate measures or identify data that should be excluded. In contrast, alternative methods rely solely on the input data without verifying whether the observed effect is genuinely drug-related. This more conservative value could in turn: i) reduce the amount of the compound needed for treatment and reduce costs, and ii) reduce the risk of harmful effects on surrounding healthy cells, when testing compounds *in vivo*.

Measuring the efficacy of treatments can help reduce the duration or doses of ineffective therapies, allowing for quicker transitions to more suitable alternatives. This approach not only optimizes patient outcomes but also minimizes exposure to unnecessary side effects. This is especially critical in CRC patients, where timely intervention can significantly influence prognosis and quality of life.

5.2.3. Novel biomarker candidates for resistance detection

In addition to evaluating treatment efficacy, there is a pressing need to identify new biomarkers that can detect resistance early, enabling the development of personalized therapeutic strategies. Early detection of resistance is crucial for preventing the emergence of long-term resistance, optimizing treatment strategies, and improving overall patient care in CRC.

In this context, TROP2 has shown to be a candidate for predicting resistance in CRC, as in the preliminary study we found an increase in the expression of TROP2 in oxaliplatin treated cells which present higher resistance than control non-treated cells.

TROP2 is a transmembrane glycoprotein that transduces its signal by means of calcium upon binding to PIP2 in its intracellular domain promoting cell proliferation. It belongs to the same family as EPCAM, sharing 49% sequence identity and 67% similarity in amino acids, and performs similar functions. Moreover, the intracellular domain is hydrolyzed by the action of TACE protein inducing the up-regulation of the expression of cyclin D1 and c-myc, thereby controlling self-renewal and proliferation.

TROP2 expression is associated with epithelial-mesenchymal transition and invasiveness in CRC by means of FAK/Src/PAK4 phosphorylation. It is also involved in multiple downstream signaling pathways, including MAPK (ERK1/2), JAK2/STAT3, PI3K/Akt and JNK [301]. Furthermore, TROP2 expression is induced early in intestinal tumorigenesis after inactivation of the APC tumor suppressor gene, and this TROP2 expression is associated with upregulation of genes involved in epithelial-mesenchymal transition, cell migration, invasiveness, and extracellular matrix remodeling. TROP2 promoter is activated by two key transcriptional regulators: β-catenin (part of the Wnt signaling pathway) and YAP (part of the Hippo signaling pathway), suggesting that aberrant activation of both the Wnt and Hippo pathways can lead to increased TROP2 expression in CRC [302].

Although TROP2 is quite frequently expressed in CRC and is associated with an aggressive histopathological phenotype, intratumoral heterogeneity has been observed [303]. Some authors have described TROP2 as a predictor of poor patient survival and have related it to the chance of disease recurrence and liver metastasis in colon cancer [17]. However, the relation between TROP2 expression and drug resistance is little studied. Several reports have suggested that TROP2 expression regulates tumor cell resistance to therapeutic drugs, including tamoxifen and trastuzumab, among others [304], which could be explained by its effect regulating the Notch1 signaling pathway in some cells. Similarly, Guerra et al, 2016 show how the expression of TROP2 generates cell survival by the activation of AKT [305,306].

Moreover, Sun et al, 2021 state that TROP2 allows for remodeling of the tumor microenvironment, leading to resistance to different drugs [307]. Finally, antibody-drug conjugate therapies have been developed such as Sacituzumab govitecan, an anti-TROP2 antibody conjugated to SN38 [308].

In this PhD thesis we have broadened the implication of TROP2 in resistance. For that purpose, we first demonstrated that TROP2 expression correlates with resistance measured by the new proposed parameter IC_{rmed} in all the five CRC cell lines both at gene and protein levels.

We also performed chronic treatment with oxaliplatin in all the five CRC cell lines to obtain resistant cell lines. SW620 and HCT116 died as they are more sensitive cell lines and did not survive to chronic treatment. However, we obtained resistant cell lines for SW480, DLD1 and HT29 after treatment with oxaliplatin for 2 months. We measured TROP2 expressions both at gene and protein levels and the results show that SW480 and DLD1 increased its expression at the gene and total protein levels, but this increase was not found in membrane.

Our hypothesis for SW480 and DLD1 cell lines is that an increase of the gene expression leads to the presence of the protein in the cytoplasm leading to the activation of resistance-related signaling pathways (Figure D2).

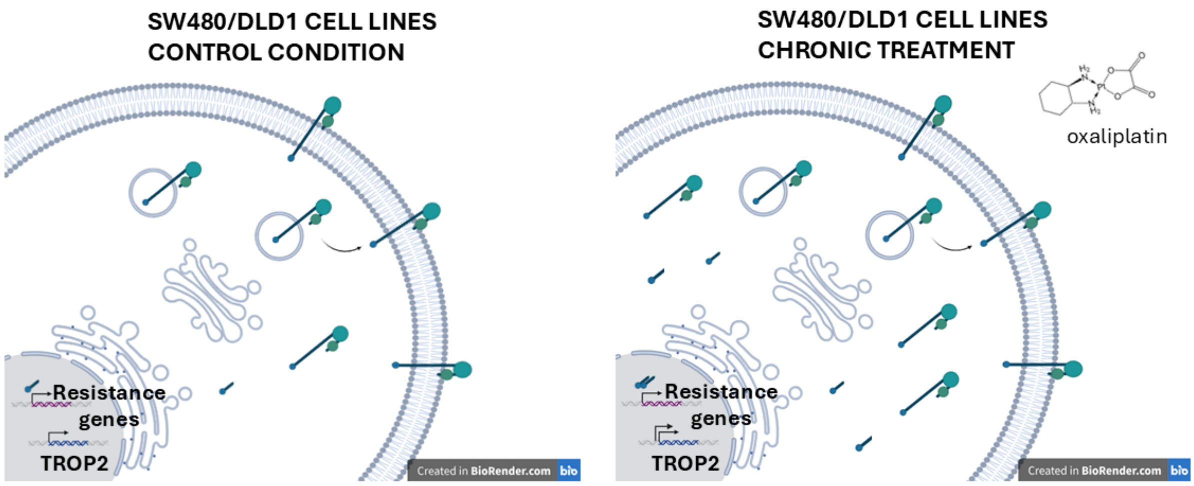

Figure D2. Hypothesis of the implication of TROP2 in resistance in SW480 and DLD1 cell lines. TROP2 is overexpressed in chronically treated cell lines compared to control untreated cells. This overexpression results in protein accumulation in the cytoplasm, leading to the activation of resistance-related genes.

In contrast, the expression of TROP2 in HT29 was increased in terms of total protein levels but decreased both in gene and membrane protein levels. As the expression was shown to be modified inside the cell but not in the membrane, it suggests that the resistance response occurs within the cytoplasm and may be due to a regulation of the expression as this cell line highly expresses TROP2 compared to SW480 and DLD1.

Although experiments are needed to confirm that, we hypothesize that the protein present in the membrane is hydrolyzed by TACE in HT29, leading to a downregulation of *TACSTD2* and an activation of resistance-related signaling pathways in the cytoplasm (Figure D3).

This hypothesis is consistent with the fact that a low expression of TROP2 in the membrane of patient samples, rather than a high expression, contributes to an increased likelihood of relapse.

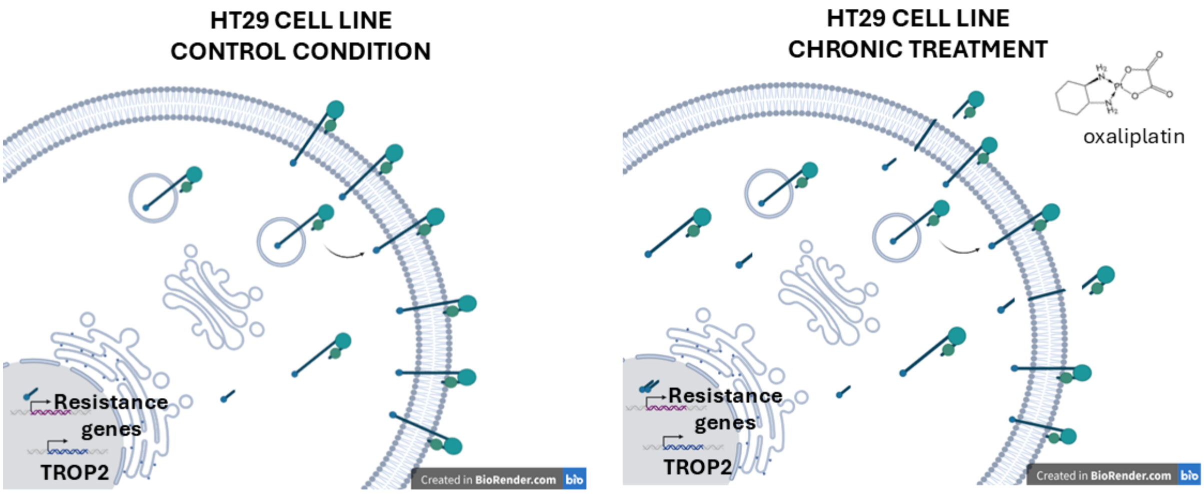

Figure D3. Hypothesis of the implication of TROP2 in resistance in HT29 cell line. HT29 is a cell line that highly expresses TROP2 on the membrane under control conditions. However, during chronic treatment with oxaliplatin, TROP2 is cleaved from the membrane, leading to its accumulation in the cytoplasm and the subsequent activation of resistance-related genes.

Several potential interactions between TROP2 and common signaling pathways involved in chemotherapy resistance in CRC have been described, including the AKT, Notch, and IGF2R-AKT pathways. TROP2 appears to interact with these pathways to promote cell survival, proliferation, and drug resistance [309,310].

The possible molecular mechanisms are:

1. Activation of protein kinase C, leading to the release of calcium from the endoplasmic reticulum and the subsequent activation of pro-survival signaling pathways like MAPK, NF-κB, and Raf.

2. Upregulation of the multidrug resistance protein MRP1 (not only ABCG2) through the activation of the Notch signaling pathway.

3. Binding of TROP2 to IGF2R, which activates the IGF2-IGF1R-AKT signaling pathway and increases cell survival and resistance to certain chemotherapeutic agents.

Furthermore, we checked whether differences in the genetic background of the cell lines could influence the response to chemotherapy and we found that SW480 and DLD1 present mutations in KRAS and P53 while HT29 only present mutations in P53 [311]. Further experiments regarding these signaling pathways will be carried out to study the molecular mechanism of resistance underlying TROP2.

Although additional molecular experiments are needed to confirm the relation between TROP2 and drug resistance, our work together with those mentioned above is an important contribution to the description of the new role for TROP2 protein in drug resistance.

In conclusion, this PhD thesis represents a meaningful advancement in the development of diagnostic tools for predicting relapses in CRC patients. By introducing an innovative predictive method based on ML, we have made progress in identifying key factors that could anticipate relapses. While the findings presented here are promising, especially in improving early intervention and treatment customization, it is important to highlight that significant work remains. Extensive and robust studies will be required to confirm its clinical relevance across a higher number of patients. To achieve this, we are preparing the documentation required by the National Agency for Medicines and Medical Devices (AEMPS) to apply for a clinical trial, categorized as a low-intervention clinical trial [312].

Additionally, we have also developed a mathematical model to evaluate the efficacy of chemotherapy. This method provides a more robust and time-independent method when compared to the standard IC_{50} method, as it is based on the growth rate measurement.

It is used for evaluating resistance and, in this PhD thesis, contributed to identifying TROP2 as a potential biomarker for detecting chemotherapy resistance, thereby paving the way for more personalized and effective therapeutic strategies.

Future experiments will focus on identifying new resistance biomarkers and validating the relapse prediction model, aiming to develop a predictive report on relapse risk and treatment efficacy. These advancements, once fully validated, have the potential to transform CRC management by reducing healthcare costs and enhancing patient outcomes. By enabling earlier and more targeted interventions, this approach could significantly improve the quality of life for CRC patients and reduce the impact of ineffective treatments.

6. Conclusions

1. The expression of the biomarkers BCRP1, CD133, AC133, CD34, CD44, CD44V6, EPCAM TROP2, CD166, LGR5, DCLK1, CD36 and RAGE varies among cell lines due to their heterogeneous origin.
2. The cells extracted from the blood of previously diagnosed CRC patients exhibit higher expression levels of the studied biomarkers compared to control (healthy) samples, except for CD44.
3. In our population of study, factors such as sex, BMI, family background, and smoking status do not appear to influence relapse. In contrast, relapse is associated with the stage, tumor localization and differentiation grade at time of diagnosis. Patients with advanced disease, distal location and poorly differentiated tumors are the most probable to relapse.
4. The expression of the studied biomarkers is influenced if the sample was collected during ongoing treatment. However, it is not affected by the type of treatment received or by whether treatment was administered between sample collections.
5. The mathematical model for predicting relapses in CRC patients achieved a sensitivity of 73 % and a specificity of 73 % when using the expression of 13 biomarkers as input. Incorporating the variable *status* (*treatment, follow-up,* or *discharge*) into the model enhanced its results, increasing sensitivity to 83 % and specificity to 82 %.
6. Logistic Regression was identified as the most effective ML model for predicting relapses.
7. CD44V6, CD36, EPCAM and LGR5 were the biomarkers that most influenced the model for predicting relapses in CRC patients.
8. The optimization of CDX requires enrichment of CTCs and the use of more efficient mouse strains such as NSG mice. We successfully achieved tumor growth for two out of five patients.
9. A novel procedure has been developed to evaluate the efficacy of chemotherapy based on growth rate measurement, providing a more robust and time-independent method when compared to the standard IC_{50} method.

10. The new proposed indices (IC_{rmed} and IC_{r0}) better evaluate the efficacy of treatments when compared to the conventional IC_{50} method, being time and method independent.
11. TROP2 is a novel candidate biomarker for resistance detection as its expression correlates with resistance both at gene and protein levels. Moreover, chronic treatment with oxaliplatin increased gene expression levels in SW480 and DLD1, and total protein levels increased in these cell lines as well as in HT29.

References

1. Surmik D, Słowiak-Morkovina J, Szczygielski T, Kamaszewski M, Kalita S, Teschner EM, et al. An insight into cancer palaeobiology: does the Mesozoic neoplasm support tissue organization field theory of tumorigenesis? BMC Ecol Evol. 2022;22: 1–13. doi:10.1186/s12862-022-02098-3

2. Breasted JH. The Edwin Smith Surgical. Chicago: University of Chicago Press. 1930.

3. Banov L, Ebbell B. The Papyrus Ebers: the greatest Egyptian medical document. Levin and Munksgaard. 1937.

4. Jones WHS. Hippocrates: Heracleitus on the universe. Harvard University Press. 1931; 521.

5. Karpozilos A, Pavlidis N. The treatment of cancer in Greek antiquity. Eur J Cancer. 2004;40: 2033–2040. doi:10.1016/j.ejca.2004.04.036

6. Izuo M. Medical history: Seishu Hanaoka and his success in breast cancer surgery under general anesthesia two hundred years ago. Breast Cancer. 2004;11: 319–324. doi:10.1007/BF02968037

7. HALSTED WS. the Results of Radical Operations for the Cure of Carcinoma of the Breast.*. Ann Surg. 1907;46: 1–19. doi:10.1097/00000658-190707000-00001

8. Hooke R. Micrographia. Royal Society of London. 1665.

9. di Lonardo A, Nasi S, Pulciani S. Cancer: We Should Not Forget The Past. J Cancer. 2015;6: 29. doi:10.7150/JCA.10336

10. Mould RF. The discovery of radium in 1898 by Maria Sklodowska-Curie (1867-1934) and Pierre Curie (1859-1906) with commentary on their life and times. British Journal of Radiology. 1998;71: 1229–1254. doi:10.1259/bjr.71.852.10318996

11. Dunn PM. Wilhelm Conrad Röentgen (1845-1923), the discovery of x rays and perinatal diagnosis. Arch Dis Child Fetal Neonatal Ed. 2001;84: 138–140. doi:10.1136/fn.84.2.f138

12. Bhattacharyya K. Godfrey Newbold Hounsfield (1919-2004): The man who revolutionized neuroimaging. Ann Indian Acad Neurol. 2016;19: 448–450. doi:10.4103/0972-2327.194414

13. LUIKART CB, GILBERT NS. Nitrogen mustard therapy. New Orleans Med Surg J. 1948;101: 32–38. doi:10.1136/bmj.2.4628.622

14. BATES B. First FDA-Approved Chemo Agent Turns 60. Internal Medicine News. 2009;42: 18. doi:10.1016/s1097-8690(09)70173-8

15. Devita VT, Serpick AA. Combination Chemotherapy in the Treatment of Advanced Hodgkin ' s Disease. Ann Intern Med. 2016;73: 881–896. doi:10.7326/0003-4819-73-6-881

16. Dalerba P, Dylla SJ, Park I-K, Liu R, Wang X, Cho RW, et al. Phenotypic characterization of human colorectal cancer stem cells. Proceedings of the National Academy of Sciences. 2007;104: 10158–10163. doi:10.1073/pnas.0703478104

17. Fang YJ, Lu ZH, Wang GQ, Pan ZZ, Zhou ZW, Yun JP, et al. Elevated expressions of MMP7, TROP2, and survivin are associated with survival, disease recurrence, and liver metastasis of colon cancer. Int J Colorectal Dis. 2009;24: 875. doi:10.1007/s00384-009-0725-z

18. Wu XS, Xi HQ, Chen L. Lgr5 is a potential marker of colorectal carcinoma stem cells that correlates with patient survival. World J Surg Oncol. 2012;10: 1–8. doi:10.1186/1477-7819-10-244

19. Chandrakesan P, Yao J, Qu D, May R, Weygant N, Ge Y, et al. Dclk1, a tumor stem cell marker, regulates pro-survival signaling and self-renewal of intestinal tumor cells. Mol Cancer. 2017;16: 30. doi:10.1186/s12943-017-0594-y

20. Grillo-Lopez A, White C, Dallaire B, Varns C, Shen C, Wei A, et al. Rituximab The First Monoclonal Antibody Approved for the Treatment of Lymphoma. Curr Pharm Biotechnol. 2005;1: 1–9. doi:10.2174/1389201003379059

21. Jeyakumar A, Younis T. Trastuzumab for HER2-positive metastatic breast cancer: Clinical and economic considerations. Clin Med Insights Oncol. 2012;6: 179–187. doi:10.4137/CMO.S6460

22. Lipson, Evan J. Drake CG. Ipilimumab: An Anti-CTLA-4 Antibody for Metastatic Melanoma Evan. Clinical Cancer Research. 2011;15: 6958–6962. doi:10.1158/1078-0432.CCR-11-1595

23. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. N Engl J Med. 2012;366: 2443. doi:10.1056/NEJMOA1200690

24. Singh AK, McGuirk JP. CAR T cells: continuation in a revolution of immunotherapy. Lancet Oncol. 2020;21: e168–e178. doi:10.1016/S1470-2045(19)30823-X

25. Ercilla-Rodríguez P, Sánchez-Díez M, Alegría-Aravena N, Quiroz-Troncoso J, Gavira-O'Neill CE, González-Martos R, et al. CAR-T lymphocyte-based cell therapies; mechanistic substantiation, applications and biosafety enhancement with suicide genes: new opportunities to melt side effects. Front Immunol. 2024;15. doi:10.3389/FIMMU.2024.1333150

26. Falzone L, Salomone S, Libra M. Evolution of cancer pharmacological treatments at the turn of the third millennium. Front Pharmacol. 2018;9. doi:10.3389/fphar.2018.01300

27. Chen M, Mao A, Xu M, Weng Q, Mao J, Ji J. CRISPR-Cas9 for cancer therapy: Opportunities and challenges. Cancer Lett. 2019;447: 48–55. doi:10.1016/j.canlet.2019.01.017

28. Saxena M, van der Burg SH, Melief CJM, Bhardwaj N. Therapeutic cancer vaccines. Nat Rev Cancer. 2021;21: 360–378. doi:10.1038/s41568-021-00346-0

29. Hickam DH, Shortliffe EH, Bischoff MB, Scott AC, Jacobs CD, Stanford ; The Treatment Advice of a Computer-Based Cancer Chemotherapy Protocol Advisor. Available: https://annals.org

30. Verlingue L, Boyer C, Olgiati L, Brutti Mairesse C, Morel D, Blay JY. Artificial intelligence in oncology: ensuring safe and effective integration of language models in clinical practice. The Lancet Regional Health - Europe. Elsevier Ltd; 2024. doi:10.1016/j.lanepe.2024.101064

31. Chen MM, Terzic A, Becker AS, Johnson JM, Wu CC, Wintermark M, et al. Artificial intelligence in oncologic imaging. Eur J Radiol Open. 2022;9. doi:10.1016/j.ejro.2022.100441

32. Froicu EM, Oniciuc OM, Afrăsânie VA, Marinca MV, Riondino S, Dumitrescu EA, et al. The Use of Artificial Intelligence in Predicting Chemotherapy-Induced Toxicities in Metastatic Colorectal Cancer: A Data-Driven Approach for Personalized Oncology. Diagnostics. 2024;14. doi:10.3390/diagnostics14182074

33. Edition T. International Classification of Diseases for Oncology, Third Edition. Definitions. 2020. doi:10.32388/5xg1qe

34. Rosen R, Sapra A. TNM Classification. StatPearls. 2023.

35. Ferlay, J; Ervik, M; Lam, F;Laversanne, M;Colombet, M; Mery, L; Piñeros, M; Znaor, A; Soerjomataram, I; Bray F. Global Cancer Observatory. International Agency for Research on Cancer. 2024.

36. Ferlay, J; Ervik, M; Lam, F;Laversanne, M;Colombet, M; Mery, L; Piñeros, M; Znaor, A; Soerjomataram, I; Bray F. Global Cancer Observatory. International Agency for Research on Cancer. 2024.

37. Balmain A. The critical roles of somatic mutations and environmental tumor-promoting agents in cancer risk. Nature Genetics. Nature Research; 2020. pp. 1139–1143. doi:10.1038/s41588-020-00727-5

38. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V., et al. Signatures of mutational processes in human cancer. Nature. 2013;500: 415–421. doi:10.1038/nature12477

39. Dariya B, Aliya S, Merchant N, Alam A, Nagaraju GP. Colorectal Cancer Biology, Diagnosis, and Therapeutic Approaches. Crit Rev Oncog. 2020;25: 71–94. doi:10.1615/CritRevOncog.2020035067

40. Brozyna A, Zbytek B, Granese J, Carlson JA, Ross J, Slominski A. Mechanism of UV-related carcinogenesis and its contribution to nevi/melanoma. Expert Rev Dermatol. 2007;2: 451–469. doi:10.1586/17469872.2.4.451

41. Gaglia MM, Munger K. More than just oncogenes: mechanisms of tumorigenesis by human viruses. Curr Opin Virol. 2018;32: 48–59. doi:10.1016/j.coviro.2018.09.003

42. Mesri EA, Feitelson MA, Munger K. Human viral oncogenesis: A cancer hallmarks analysis. Cell Host Microbe. 2014;15: 266–282. doi:10.1016/j.chom.2014.02.011

43. Phillips DH, Arlt VM. Genotoxicity: damage to DNA and its consequences. EXS. 2009;99: 87–110. doi:10.1007/978-3-7643-8336-7_4

44. Tao Pu LZC, Rana K, Singh G, Nakamura M, Yamamura T, Koay DSC, et al. Different factors are associated with conventional adenoma and serrated colorectal Neoplasia. Nagoya J Med Sci. 2020;82: 335–343. doi:10.18999/nagjms.82.2.335

45. Scappaticcio L, Maiorino MI, Bellastella G, Giugliano D, Esposito K. Insights into the relationships between diabetes, prediabetes, and cancer. Endocrine. Humana Press Inc.; 2017. pp. 231–239. doi:10.1007/s12020-016-1216-y

46. Sninsky JA, Shore BM, Lupu G V, Crockett SD, Crockett S. Risk Factors for Colorectal Polyps and Cancer Author names. Gastrointest Endosc Clin N Am. 2022;32(2): 195–213. doi:10.1016/j.giec.2021.12.008

47. Behrens G, Gredner T, Stock C, Leitzmann MF, Brenner H, Mons U. Cancers Due to Excess Weight, Low Physical Activity, and Unhealthy Diet. Dtsch Arzteb Int. 2018;115: 578–585. doi:10.3238/arz

48. Papier K, Bradbury KE, Balkwill A, Barnes I, Smith-Byrne K, Gunter MJ, et al. Diet-wide analyses for risk of colorectal cancer: prospective study of 12,251 incident cases among 542,778 women in the UK. Nat Commun. 2025;16: 375. doi:10.1038/s41467-024-55219-5

49. Li CH, Haider S, Boutros PC. Age influences on the molecular presentation of tumours. Nat Commun. 2022;13. doi:10.1038/s41467-021-27889-y

50. Malcomson FC, Wiggins C, Parra-Soto S, Ho FK, Celis-Morales C, Sharp L, et al. Adherence to the 2018 World Cancer Research Fund/American Institute for Cancer Research Cancer Prevention Recommendations and cancer risk: A systematic review and meta-analysis. Cancer. 2023;129: 2655–2670. doi:10.1002/cncr.34842

51. Young GP, Rabeneck L, Winawer SJ. The Global Paradigm Shift in Screening for Colorectal Cancer. Gastroenterology. 2019;156: 843-851.e2. doi:10.1053/j.gastro.2019.02.006

52. Duran-Sanchon S, Moreno L, Gómez-Matas J, Augé JM, Serra-Burriel M, Cuatrecasas M, et al. Fecal MicroRNA-Based Algorithm Increases Effectiveness of Fecal Immunochemical Test–Based Screening for Colorectal Cancer. Clinical Gastroenterology and Hepatology. 2021;19: 323-330.e1. doi:10.1016/j.cgh.2020.02.043

53. Park SH, Lee JH, Lee SS, Kim JC, Yu CS, Kim HC, et al. CT colonography for detection and characterisation of synchronous proximal colonic lesions in patients with stenosing colorectal cancer. Gut. 2012;61: 1716–1722. doi:10.1136/gutjnl-2011-301135

54. Kijima S, Sasaki T, Nagata K, Utano K, Lefor AT, Sugimot H. Preoperative evaluation of colorectal cancer using CT colonography, MRI, and PET/CT. World J Gastroenterol. 2014;20: 16964–16975. doi:10.3748/wjg.v20.i45.16964

55. Zhang Y, Wang Y, Zhang B, Li P, Zhao Y. Methods and biomarkers for early detection, prediction, and diagnosis of colorectal cancer. Biomedicine and Pharmacotherapy. 2023;163: 114786. doi:10.1016/j.biopha.2023.114786

56. Baraibar I, Ros J, Mulet N, Salvà F, Argilés G, Martini G, et al. Incorporating traditional and emerging biomarkers in the clinical management of metastatic colorectal cancer: an update. Expert Rev Mol Diagn. 2020;20: 653–664. doi:10.1080/14737159.2020.1782194

57. Zhong W, Yu Z, Zhan J, Yu T, Lin Y, Xia ZS, et al. Association of Serum Levels of CEA, CA199, CA125, CYFRA21-1 and CA72-4 and Disease Characteristics in Colorectal Cancer. Pathology and Oncology Research. 2015;21: 83–95. doi:10.1007/s12253-014-9791-9

58. Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. Annals of Oncology. 2016;27: 1386–1422. doi:10.1093/annonc/mdw235

59. Macqueen G, Santaguida P, Keshavarz H, Jaworska N, Levine M, Beyene J, et al. Systematic Review of Clinical Practice Guidelines for Failed Antidepressant Treatment Response in Major Depressive Disorder, Dysthymia, and Subthreshold Depression in Adults. Canadian Journal of Psychiatry. 2017;62: 11–23. doi:10.1177/0706743716664885

60. Argilés G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Annals of Oncology. 2020;31: 1291–1305. doi:10.1016/j.annonc.2020.06.022

61. Bonjer HJ, Deijen CL, Abis GA, Cuesta MA, van der Pas MHGM, de Lange-de Klerk ESM, et al. A Randomized Trial of Laparoscopic versus Open

Surgery for Rectal Cancer. New England Journal of Medicine. 2015;372: 1324–1332. doi:10.1056/nejmoa1414882

62. Nagtegaal ID, Marijnen CAM, Kranenbarg EK, Van De Velde CJH, Van Krieken JHJM. Circumferential margin involvement is still an important predictor of local recurrence in rectal carcinoma: Not one millimeter but two millimeters is the limit. American Journal of Surgical Pathology. 2002;26: 350–357. doi:10.1097/00000478-200203000-00009

63. Van De Velde CJH, Boelens PG, Borras JM, Coebergh JW, Cervantes A, Blomqvist L, et al. EURECCA colorectal: Multidisciplinary management: European consensus conference colon & rectum. Eur J Cancer. 2014;50: 1.e1-1.e34. doi:10.1016/j.ejca.2013.06.048

64. Vodenkova S, Buchler T, Cervena K, Veskrnova V, Vodicka P, Vymetalkova V. 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. Pharmacol Ther. 2020;206: 107447. doi:10.1016/j.pharmthera.2019.107447

65. Falcone A, Ricci S, Brunetti I, Pfanner E, Allegrine G, Barbara C, et al. Phase III trial of infusional fluorouracil, leucovorin, oxaliplatin, and irinotecan (FOLFOXIRI) compared with infusional fluorouracil, leucovorin, and irinotecan (FOLFIRI) as first-line treatment for metastatic colorectal cancer: The gruppo oncologico nord ovest. Journal of Clinical Oncology. 2007;25: 1670–1676. doi:10.1200/JCO.2006.09.0928

66. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. Journal of Clinical Oncology. 2008;26: 1626–1634. doi:10.1200/JCO.2007.14.7116

67. Douillard J-Y, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, et al. Panitumumab–FOLFOX4 Treatment and RAS Mutations in Colorectal Cancer . New England Journal of Medicine. 2013;369: 1023–1034. doi:10.1056/nejmoa1305275

68. Hainsworth J, Heim W, Berlin J, Baron A, Griffing S, Holmgren E, et al. Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for Metastatic Colorectal Cancer. 2004; 2335–2342.

69. Van Cutsem E, Tabernero J, Lakomy R, Prenen H, Prausová J, Macarulla T, et al. Addition of aflibercept to fluorouracil, leucovorin, and irinotecan

improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. Journal of Clinical Oncology. 2012;30: 3499–3506. doi:10.1200/JCO.2012.42.8201

70. Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): An international, multicentre, randomised, placebo-controlled, phase 3 trial. The Lancet. 2013;381: 303–312. doi:10.1016/S0140-6736(12)61900-X

71. Comprehensive Guide to Cell Health Assays in Flow Cytometry: A Multidimensional Exploration | AAT Bioquest. [cited 5 Sep 2024]. Available: https://www.aatbio.com/resources/assaywise/2023-12-1/comprehensive-guide-to-cell-health-assays-in-flow-cytometry-a-multidimensional-exploration

72. van Meerloo J, Kaspers GJL, Cloos J. Cell Sensitivity Assays: The MTT Assay. 2011. pp. 237–245. doi:10.1007/978-1-61779-080-5_20

73. Ali-Osman F, Maurer HR. Comparison of Cytostatic Sensitivities of L 1210 Cells and Human Stimulated Lymphocytes in Three Cell Proliferation Assays. Journal of Cancer Research Clinical On~logy. Springer-Verlag; 1980. doi:10.1007/BF00410786

74. Tobiiren D. Differential Survival of Cell Subpopulations in Spheroids After Treatment with Bleomycin. J Cancer Res Clin Oncol. 1981. doi:10.1007/BF00410671

75. Ng CE, Lnch WR. Comparison of the Densities of Clonogenic Cells From EMT6 Fibrosarcoma Monolayer Cultures, Multicell Spheroids, and Solid Tumors in Ficoll Density Gradients. 1978. doi:10.1093/jnci/60.5.1017

76. Benzekry S, Lamont C, Beheshti A, Tracz A, Ebos JML, Hlatky L, et al. Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth. PLoS Comput Biol. 2014;10. doi:10.1371/journal.pcbi.1003800

77. Karnes RJ, MacKintosh FR, Morrell CH, Rawson L, Sprenkle PC, Kattan MW, et al. Prostate-specific antigen trends predict the probability of prostate cancer in a very large U.S. Veterans affairs cohort. Front Oncol. 2018;8. doi:10.3389/fonc.2018.00296

78. Sebaugh JL. Guidelines for accurate EC50/IC50 estimation. Pharm Stat. 2011;10: 128–134. doi:10.1002/pst.426

79. Adile AA, Bakhshinyan D, Venugopal C, Singh SK. In Vitro Assays for Screening Small Molecules. Methods in Molecular Biology. Humana Press Inc.; 2019. pp. 189–196. doi:10.1007/978-1-4939-8805-1_16

80. Islam MT, Aktaruzzaman M, Saif A, Hasan AR, Sourov MMH, Sikdar B, et al. Identification of acetylcholinesterase inhibitors from traditional medicinal plants for Alzheimer's disease using in silico and machine learning approaches. RSC Adv. 2024;14: 34620–34636. doi:10.1039/d4ra05073h

81. Lyles RH, Poindexter C, Evans A, Brown M, Cooper CR. NONLINEAR MODEL-BASED ESTIMATES OF IC50 FOR STUDIES INVOLVING CONTINUOUS THERAPEUTIC DOSE-RESPONSE DATA. Contemp Clin Trials. 2008;29: 878. doi:10.1016/J.CCT.2008.05.009

82. Motulsky H, Christopoulos A. Fitting Models to Biological Data using Linear and Nonlinear Regression A practical guide to curve fitting Contents at a Glance. Available: www.graphpad.com.

83. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 4th edition. . In: New York: Garland Science. 2002.

84. Biology of Cultured Cells. 2010.

85. Sonnenschein C, Soto AM, Ma B. Theories of carconogenesis. 2009;18: 372–377. doi:10.1016/j.semcancer.2008.03.012.THEORIES

86. Frank NY, Schatton T, Frank MH. The therapeutic promise of the cancer stem cell concept. Journal of Clinical Investigation. 2010;120: 41–50. doi:10.1172/JCI41004

87. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature. 2001;414: 105–111. doi:10.1038/35102167

88. Qilin Xu and AD Le. Targeting Cancer Stem Cells in Oral Cancer. J Calif Dent Assoc. 2016;44: 112–120. doi:10.1080/19424396.2016.12220978

89. DeRisi Joseph, Penland Lolita BPO, Tyagi S, Kramer FR, Group NP, DeRisi Joseph, Penland Lolita BPO. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature publishing group. 1997;4: 303–308. doi:10.1038/nm0798-822

90. Vaiopoulos AG, Kostakis ID, Koutsilieris M, Papavassiliou AG. Concise review: Colorectal cancer stem cells. Stem Cells. 2012;30: 363–371. doi:10.1002/stem.1031

91. Pretzsch E, Bösch F, Neumann J, Ganschow P, Bazhin A, Guba M, et al. Mechanisms of Metastasis in Colorectal Cancer and Metastatic Organotropism: Hematogenous versus Peritoneal Spread. J Oncol. 2019;2019. doi:10.1155/2019/7407190

92. Kim DH, Xing T, Yang Z, Dudek R, Lu Q, Chen YH. Epithelial mesenchymal transition in embryonic development, tissue repair and cancer: A comprehensive overview. J Clin Med. 2018;7: 1–25. doi:10.3390/jcm7010001

93. Fanali C, Lucchetti D, Farina M, Corbi M, Cufino V, Cittadini A, et al. Cancer stem cells in colorectal cancer from pathogenesis to therapy: controversies and perspectives. World J Gastroenterol. 2014;20: 923–42. doi:10.3748/wjg.v20.i4.923

94. Huang EH, Heidt DG, Li CW, Simeone DM. Cancer stem cells: A new paradigm for understanding tumor progression and therapeutic resistance. Surgery. 2007;141: 415–419. doi:10.1016/j.surg.2006.12.015

95. Zeng Z, Fu M, Hu Y, Wei Y, Wei X, Luo M. Regulation and signaling pathways in cancer stem cells: implications for targeted therapy for cancer. Mol Cancer. 2023;22: 1–31. doi:10.1186/s12943-023-01877-w

96. Vall d´Hebron Instituto de Oncología. Primer paciente tratado en el mundo con un innovador fármaco que ataca las células madre tumorales. In: Vall d´Hebron. 2024.

97. Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC. Isolation and Functional Properties of Murine Hematopoietic Stem Cells that are Replicating In Vivo. 1996. doi:10.1084/jem.183.4.1797

98. Kondo T, Setoguchi T, Taga T. Persistence of a small subpopulation of cancer stem-like cells in the C6 glioma cell line. 2004. doi:10.1073/pnas.0307618100

99. Ding X wei, Wu J hua, Jiang C ping. ABCG2: A potential marker of stem cells and novel target in stem cell and cancer therapy. Life Sci. 2010;86: 631–637. doi:10.1016/j.lfs.2010.02.012

100. Westover D, Li F. New trends for overcoming ABCG2/BCRP-mediated resistance to cancer therapies. Journal of Experimental and Clinical Cancer Research. 2015;34: 1–9. doi:10.1186/s13046-015-0275-x

101. Tuy HD, Shiomi H, Mukaisho KI, Naka S, Shimizu T, Sonoda H, et al. ABCG2 expression in colorectal adenocarcinomas may predict resistance to irinotecan. Oncol Lett. 2016;12: 2752–2760. doi:10.3892/ol.2016.4937

102. Wang X, Xia B, Liang Y, Peng L, Wang Z, Zhuo J, et al. Membranous ABCG2 expression in colorectal cancer independently correlates with shortened patient survival. Cancer Biomarkers. 2013;13: 81–88. doi:10.3233/CBM-130344

103. Giampieri R, Scartozzi M, Loretelli C, Piva F, Mandolesi A, Lezoche G, et al. Cancer Stem Cell Gene Profile as Predictor of Relapse in High Risk Stage II and Stage III, Radically Resected Colon Cancer Patients. PLoS One. 2013;8: 1–6. doi:10.1371/journal.pone.0072843

104. Brien CAO, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. 2007;445: 106–110. doi:10.1038/nature05372

105. Ricci-vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, et al. Identification and expansion of human colon-cancer-initiating cells. 2007;445: 111–115. doi:10.1038/nature05384

106. Liu C, Li Y, Xing Y, Cao B, Yang F, Yang T, et al. The interaction between cancer stem cell marker CD133 and Src protein promotes Focal Adhesion Kinase (FAK) phosphorylation and cell migration. Journal of Biological Chemistry. 2016;291: 15540–15550. doi:10.1074/jbc.M115.712976

107. Sahlberg SH, Spiegelberg D, Glimelius B, Stenerlöw B, Nestor M. Evaluation of cancer stem cell markers CD133, CD44, CD24: Association with AKT isoforms and radiation resistance in colon cancer cells. PLoS One. 2014;9. doi:10.1371/journal.pone.0094621

108. Song S, Pei G, Du Y, Wu J, Ni X, Wang S, et al. Interaction between CD133 and PI3K-p85 promotes chemoresistance in gastric cancer cells. Am J Transl Res. 2018;10: 304–314.

109. Sipos F, Műzes G. Interconnection of CD133 Stem Cell Marker with Autophagy and Apoptosis in Colorectal Cancer. International Journal of

Molecular Sciences. Multidisciplinary Digital Publishing Institute (MDPI); 2024. doi:10.3390/ijms252011201

110. Ying X, Wu J, Meng X, Zuo Y, Xia Q, Chen J, et al. AC133 expression associated with poor prognosis in stage II colorectal cancer. Med Oncol. 2013;30: 356. doi:10.1007/s12032-012-0356-z

111. Wang B-B, Li Z-J, Zhang F-F, Hou H-T, Yu J-K, Li F. Clinical significance of stem cell marker CD133 expression in colorectal cancer. Histol Histopathol. 2016;31: 299–306. doi:10.14670/HH-11-676

112. Huang JL, Oshi M, Endo I, Takabe K. Clinical relevance of stem cell surface markers CD133 , CD24 , and CD44 in colorectal cancer. 2021;11: 5141–5154.

113. Yang SJ, Tseng SY, Wang CH, Young TH, Chen KC, Shieh MJ. Magnetic nanomedicine for CD133-expressing cancer therapy using locoregional hyperthermia combined with chemotherapy. Nanomedicine. 2020;15: 2543–2561. doi:10.2217/nnm-2020-0222

114. Yan S, Tang D, Hong Z, Wang J, Yao H, Lu L, et al. CD133 peptide-conjugated pyropheophorbide-a as a novel photosensitizer for targeted photodynamic therapy in colorectal cancer stem cells. Biomater Sci. 2021;9: 2020–2031. doi:10.1039/d0bm01874k

115. Ning ST, Lee SY, Wei MF, Peng CL, Lin SYF, Tsai MH, et al. Targeting Colorectal Cancer Stem-Like Cells with Anti-CD133 Antibody-Conjugated SN-38 Nanoparticles. ACS Appl Mater Interfaces. 2016;8. doi:10.1021/acsami.6b04403

116. Ma YL, Peng JY, Zhang P, Liu WJ, Huang L, Qin HL. Immunohistochemical analysis revealed CD34 and Ki67 protein expression as significant prognostic factors in colorectal cancer. Medical Oncology. 2010;27: 304–309. doi:10.1007/s12032-009-9210-3

117. Tampakis A, Weixler B, Rast S, Tampaki EC, Cremonesi E, Kancherla V, et al. Nestin and CD34 expression in colorectal cancer predicts improved overall survival. Acta Oncol (Madr). 2021;60: 727–734. doi:10.1080/0284186X.2021.1891280

118. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature. 1994;367: 645–648. doi:10.1038/367645a0

119. Kapoor S, Shenoy SP, Bose B. CD34 cells in somatic, regenerative and cancer stem cells: Developmental biology, cell therapy, and omics big data perspective. J Cell Biochem. 2020;121: 3058–3069. doi:10.1002/jcb.29571

120. Arruda LCM, Stikvoort A, Lambert M, Jin L, Rivera LS, Alves RMP, et al. A novel CD34-specific T-cell engager efficiently depletes acute myeloid leukemia and leukemic stem cells in vitro and in vivo. Haematologica. 2022;107: 1786–1795. doi:10.3324/haematol.2021.279486

121. Bhatavdekar JM, Patel DD, Chikhlikar PR, Trivedi TI, Gosalia NM, Ghosh N, et al. Overexpression of CD44: A useful independent predictor of prognosis in patients with colorectal carcinomas. Ann Surg Oncol. 1998;5: 495–501. doi:10.1007/BF02303641

122. Shang Z, Cai Q, Zhang M, Zhu S, Ma Y, Sun L, et al. A switch from CD44+ cell to EMT cell drives the metastasis of prostate cancer. Oncotarget. 2015;6: 1202–1216. doi:10.18632/oncotarget.2841

123. Miletti-González KE, Chen S, Muthukumaran N, Saglimbeni GN, Wu X, Yang J, et al. The CD44 receptor interacts with P-glycoprotein to promote cell migration and invasion in cancer. Cancer Res. 2005;65: 6660–6667. doi:10.1158/0008-5472.CAN-04-3478

124. Chen L, Fu C, Zhang Q, He C, Zhang F, Wei Q. The role of CD44 in pathological angiogenesis. FASEB Journal. 2020;34: 13125–13139. doi:10.1096/fj.202000380RR

125. Lee SY, Kim KA, Kim CH, Kim YJ, Lee JH, Kim HR. CD44-shRNA recombinant adenovirus inhibits cell proliferation, invasion, and migration, and promotes apoptosis in HCT116 colon cancer cells. Int J Oncol. 2017;50: 329–336. doi:10.3892/ijo.2016.3801

126. Ohata H, Ishiguro T, Aihara Y, Sato A, Sakai H, Sekine S, et al. Induction of the stem-like cell regulator CD44 by Rho kinase inhibition contributes to the maintenance of colon cancer-initiating cells. Cancer Res. 2012;72: 5101–5110. doi:10.1158/0008-5472.CAN-11-3812

127. Du L, Wang H, He L, Zhang J, Ni B, Wang X, et al. CD44 is of functional importance for colorectal cancer stem cells. Clinical Cancer Research. 2008;14: 6751–6760. doi:10.1158/1078-0432.CCR-08-1034

128. Sackstein R. The bone marrow is akin to skin: HCELL and the biology of hematopoietic stem cell homing. Journal of Investigative Dermatology. 2004;122: 1061–1069. doi:10.1111/j.0022-202X.2004.09301.x

129. Ma L, Dong L, Chang P. CD44v6 engages in colorectal cancer progression. Cell Death Dis. 2019;10. doi:10.1038/s41419-018-1265-7

130. Garouniatis A, Zizi-Sermpetzoglou A, Rizos S, Kostakis A, Nikiteas N, Papavassiliou AG. FAK, CD44v6, c-Met and EGFR in colorectal cancer parameters: Tumour progression, metastasis, patient survival and receptor crosstalk. Int J Colorectal Dis. 2013;28: 9–18. doi:10.1007/s00384-012-1520-9

131. Saito S, Okabe H, Watanabe M, Ishimoto T, Iwatsuki M, Baba Y, et al. CD44v6 expression is related to mesenchymal phenotype and poor prognosis in patients with colorectal cancer. Oncol Rep. 2013;29: 1570–1578. doi:10.3892/or.2013.2273

132. Todaro M, Gaggianesi M, Catalano V, Benfante A, Iovino F, Biffoni M, et al. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. Cell Stem Cell. 2014;14: 342–356. doi:10.1016/j.stem.2014.01.009

133. Andrade F, Rafael D, Vilar-Hernández M, Montero S, Martínez-Trucharte F, Seras-Franzoso J, et al. Polymeric micelles targeted against CD44v6 receptor increase niclosamide efficacy against colorectal cancer stem cells and reduce circulating tumor cells in vivo. Journal of Controlled Release. 2021;331: 198–212. doi:10.1016/j.jconrel.2021.01.022

134. Raftery MJ, Franzén AS, Radecke C, Boulifa A, Schönrich G, Stintzing S, et al. Next Generation CD44v6-Specific CAR-NK Cells Effective against Triple Negative Breast Cancer. Int J Mol Sci. 2023;24. doi:10.3390/ijms24109038

135. Lourenço BN, Pereira RF, Barrias CC, Fischbach C, Oliveira C, Granja PL. Engineering modular half-antibody conjugated nanoparticles for targeting CD44v6-expressing cancer cells. Nanomaterials. 2021;11: 1–16. doi:10.3390/nano11020295

136. Wu C, Mannan P, Lu M, Udey MC. Epithelial Cell Adhesion Molecule (EpCAM) Regulates Claudin Dynamics and Tight Junctions * □. 2013;288: 12253–12268. doi:10.1074/jbc.M113.457499

137. Kalantari E, Taheri T, Fata S, Abolhasani M, Mehrazma M, Madjd Z, et al. Significant co-expression of putative cancer stem cell markers, EpCAM and CD166, correlates with tumor stage and invasive behavior in colorectal cancer. World J Surg Oncol. 2022;20: 1–11. doi:10.1186/s12957-021-02469-y

138. Lugli A, Iezzi G, Hostettler I, Muraro MG, Mele V, Tornillo L, et al. Prognostic impact of the expression of putative cancer stem cell markers CD133, CD166, CD44s, EpCAM, and ALDH1 in colorectal cancer. Br J Cancer. 2010;103: 382–390. doi:10.1038/sj.bjc.6605762

139. Wu CJ, Feng X, Lu M, Morimura S, Udey MC. Matriptase-mediated cleavage of EpCAM destabilizes claudins and dysregulates intestinal epithelial homeostasis. Journal of Clinical Investigation. 2017;127: 623–634. doi:10.1172/JCI88428

140. Maetzel D, Denzel S, Mack B, Canis M, Went P, Benk M, et al. Nuclear signalling by tumour-associated antigen EpCAM. Nat Cell Biol. 2009;11: 162–171. doi:10.1038/ncb1824

141. Dalerba P, Dylla SJ, Park I-K, Liu R, Wang X, Cho RW, et al. Phenotypic characterization of human colorectal cancer stem cells. Proceedings of the National Academy of Sciences. 2007;104: 10158–10163. doi:10.1073/pnas.0703478104

142. Li D, Guo X, Yang K, Yang Y, Zhou W, Huang Y, et al. EpCAM-targeting CAR-T cell immunotherapy is safe and efficacious for epithelial tumors. Sci Adv. 2023;9: 1–17. doi:10.1126/SCIADV.ADG9721

143. Eyvazi S, Farajnia S, Dastmalchi S, Kanipour F, Zarredar H, Bandehpour M. Antibody Based EpCAM Targeted Therapy of Cancer, Review and Update. Current Cancer Drug Targets. 2018. pp. 857–868. doi:http://dx.doi.org/10.2174/1568009618666180102102311

144. Pavšič M. Trop2 forms a stable dimer with significant structural differences within the membrane-distal region as compared to epcam. Int J Mol Sci. 2021;22. doi:10.3390/ijms221910640

145. Lin JC, Wu YY, Wu JY, Lin TC, Wu CT, Chang YL, et al. TROP2 is epigenetically inactivated and modulates IGF-1R signalling in lung adenocarcinoma. EMBO Mol Med. 2012;4: 472–485. doi:10.1002/emmm.201200222

146. Trerotola M, Cantanelli P, Guerra E, Tripaldi R, Aloisi AL, Bonasera V, et al. Upregulation of Trop-2 quantitatively stimulates human cancer growth. Oncogene. 2013;32: 222–233. doi:10.1038/onc.2012.36

147. Zhao W, Jia L, kuai X, Tang Q, Huang X, Yang T, et al. The role and molecular mechanism of Trop2 induced epithelial-mesenchymal transition through mediated β-catenin in gastric cancer. Cancer Med. 2019;8: 1135–1147. doi:10.1002/cam4.1934

148. Mori Y, Akita K, Ojima K, Iwamoto S, Yamashita T, Morii E, et al. Trophoblast cell surface antigen 2 (Trop-2) phosphorylation by protein kinase C / (PKC/) enhances cell motility. Journal of Biological Chemistry. 2019;294: 11513–11524. doi:10.1074/jbc.RA119.008084

149. Kuai X, Jia L, Yang T, Huang X, Zhao W, Zhang M, et al. TROP2 promotes multidrug resistance by regulating Notch1 signaling pathway in gastric cancer cells. Medical Science Monitor. 2020;26: 1–9. doi:10.12659/MSM.919566

150. Švec J, Šťastná M, Janečková L, Hrčkulák D, Vojtěchová M, Onhajzer J, et al. TROP2 Represents a Negative Prognostic Factor in Colorectal Adenocarcinoma and Its Expression Is Associated with Features of Epithelial–Mesenchymal Transition and Invasiveness. Cancers (Basel). 2022;14. doi:10.3390/cancers14174137

151. Li X, Teng S, Zhang Y, Zhang W, Zhang X, Xu K, et al. TROP2 promotes proliferation, migration and metastasis of gallbladder cancer cells by regulating PI3K/AKT pathway and inducing EMT. Oncotarget. 2017;8: 47052–47063. doi:10.18632/oncotarget.16789

152. Ambrogi F, Fornili M, Boracchi P, Trerotola M, Relli V, Simeone P, et al. Trop-2 is a determinant of breast cancer survival. PLoS One. 2014;9. doi:10.1371/journal.pone.0096993

153. Stoyanova T, Goldstein AS, Cai H, Drake JM, Huang J, Witte ON. Regulated proteolysis of Trop2 drives epithelial hyperplasia and stem cell self-renewal via β-catenin signaling. Genes Dev. 2012;26: 2271–2285. doi:10.1101/gad.196451.112

154. Wanger TM, Dewitt S, Collins A, Maitland NJ, Poghosyan Z, Knäuper V. Differential regulation of TROP2 release by PKC isoforms through vesicles

and ADAM17. Cell Signal. 2015;27: 1325–1335. doi:10.1016/j.cellsig.2015.03.017

155. Goldenberg DM, Sharkey RM. Antibody-drug conjugates targeting TROP-2 and incorporating SN-38: A case study of anti-TROP-2 sacituzumab govitecan. MAbs. 2019;11: 987–995. doi:10.1080/19420862.2019.1632115

156. Goldenberg DM, Cardillo TM, Govindan S V., Rossi EA, Sharkey RM. Trop-2 is a novel target for solid cancer therapy with sacituzumab govitecan (IMMU-132), an antibody-drug conjugate (ADC) (Oncotarget (2015) 6 (22496-22512) DOI: 10.18632/oncotarget.4318). Oncotarget. 2020;11: 942. doi:10.18632/oncotarget.27512

157. Son S, Shin S, Rao NV, Um W, Jeon J, Ko H, et al. Anti-Trop2 antibody-conjugated bioreducible nanoparticles for targeted triple negative breast cancer therapy. Int J Biol Macromol. 2018;110: 406–415. doi:10.1016/j.ijbiomac.2017.10.113

158. Chang CH, Goldenberg DM. Enhancing the antitumor potency of T cells redirected by bispecific antibodies. Oncoscience. 2017;4: 120–121. doi:10.18632/oncoscience.366

159. Horst D, Kriegl L, Engel J, Kirchner T, Jung A. Prognostic significance of the cancer stem cell markers CD133, CD44, and CD166 in colorectal cancer. Cancer Invest. 2009;27: 844–850. doi:10.1080/07357900902744502

160. Manhas J, Bhattacharya A, Agrawal SK, Gupta B, Das P, Deo SVS, et al. Characterization of cancer stem cells from different grades of human colorectal cancer. Tumor Biology. 2016;37: 14069–14081. doi:10.1007/s13277-016-5232-6

161. Bellach J, Dietel M, Kristiansen G. ALCAM/CD166 is overexpressed in colorectal carcinoma and correlates with shortened patient survival. 2004; 1160–1164. doi:10.1136/jcp.2004.016238

162. 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. Journal of Surgical Research. 2012;177: e15–e20. doi:10.1016/j.jss.2012.02.013

163. Victoria M. Ho, Ji-Ann Lee and KCM, Dias, Vera Junn, Eunsung Mouradian MM. Elevated ALCAM shedding in colorectal cancer correlates with poor

patient outcome. Bone. 2012;23: 1–7. doi:10.1158/0008-5472.CAN-12-2052.Elevated

164. He S, Li S, Guo J, Zeng X, Liang D, Zhu Y, et al. CD166-specific CAR-T cells potently target colorectal cancer cells. Transl Oncol. 2023;27. doi:10.1016/j.tranon.2022.101575

165. Boni V, Fidler MJ, Arkenau HT, Spira A, Meric-Bernstam F, Uboha N, et al. Praluzatamab Ravtansine, a CD166-Targeting Antibody–Drug Conjugate, in Patients with Advanced Solid Tumors: An Open-Label Phase I/II Trial. Clinical Cancer Research. 2022;28: 2020–2029. doi:10.1158/1078-0432.CCR-21-3656

166. Lin Y, Wu T, Yao Q, Zi S, Cui L, Yang M, et al. LGR5 promotes the proliferation of colorectal cancer cells via the Wnt/β-catenin signaling pathway. Oncol Lett. 2015;9: 2859–2863. doi:10.3892/ol.2015.3144

167. 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: 511–524. doi:10.1016/j.stem.2011.02.020

168. 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. Int J Colorectal Dis. 2013;28: 1535–1546. doi:10.1007/s00384-013-1721-x

169. Chen X, Wei B, Han X, Zheng Z, Huang J, Liu J, et al. LGR5 is required for the maintenance of spheroid-derived colon cancer stem cells. Int J Mol Med. 2014;34: 35–42. doi:10.3892/ijmm.2014.1752

170. Hirsch D, Barker N, Mcneil N, Hu Y, Camps J, Mckinnon K, et al. LGR5 positivity defines stem-like cells in colorectal cancer. Carcinogenesis. 2014;35: 849–858. doi:10.1093/carcin/bgt377

171. Cancer HC. Expression Patterns for TETs , LGR5 and BMI1 in Cancer Stem-like Cells Isolated from Human Colon Cancer. 2019;11.

172. 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. Cellular Physiology and Biochemistry. 2018;46: 860–872. doi:10.1159/000488743

173. De Sousa E Melo F, Kurtova A V., Harnoss JM, Kljavin N, Hoeck JD, Hung J, et al. A distinct role for Lgr5 + stem cells in primary and metastatic colon cancer. Nature. 2017;543: 676–680. doi:10.1038/nature21713

174. Shimokawa M, Ohta Y, Nishikori S, Matano M, Takano A, Fujii M, et al. Visualization and targeting of LGR5 + human colon cancer stem cells. Nature. 2017;545: 187–192. doi:10.1038/nature22081

175. Fumagalli A, Oost KC, Kester L, Morgner J, Bornes L, Bruens L, et al. Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. Cell Stem Cell. 2020;26: 569-578.e7. doi:10.1016/j.stem.2020.02.008

176. Cepero A, Jiménez-Carretero M, Jabalera Y, Gago L, Luque C, Cabeza L, et al. LGR5 as a Therapeutic Target of Antibody-Functionalized Biomimetic Magnetoliposomes for Colon Cancer Therapy. Int J Nanomedicine. 2024;19: 1843–1865. doi:10.2147/IJN.S440881

177. Chandrakesan P, Weygant N, May R, Qu D, Chinthalapally HR, Sureban SM, et al. Dclk1 facilitates intestinal tumor growth via enhancing pluripotency and epithelial mesenchymal transition. Oncotarget. 2014;5: 9269–9280. doi:10.18632/oncotarget.2393

178. DCLK1 as part of EMT feedback-loop promotes colorectal cancer cell proliferation, invasion, and 5-Fu resistance. 2018;29: 2018. doi:10.1093/annonc/mdy431

179. Weygant N, Qu D, Berry WL, May R, Chandrakesan P, Owen DB, et al. Small molecule kinase inhibitor LRRK2-IN-1 demonstrates potent activity against colorectal and pancreatic cancer through inhibition of doublecortin-like kinase 1. Mol Cancer. 2014;13: 1–14. doi:10.1186/1476-4598-13-103

180. Ji D, Zhan T, Li M, Yao Y, Jia J, Yi H, et al. Enhancement of Sensitivity to Chemo/Radiation Therapy by Using miR-15b against DCLK1 in Colorectal Cancer. Stem Cell Reports. 2018;11: 1506–1522. doi:10.1016/j.stemcr.2018.10.015

181. Nakanishi Y, Seno H, Fukuoka A, Ueo T, Yamaga Y, Maruno T, et al. Dclk1 distinguishes between tumor and normal stem cells in the intestine. Nat Genet. 2013;45: 98–103. doi:10.1038/ng.24813

182. Vedeld HM, Skotheim RI, Lothe RA, Lind GE. The recently suggested intestinal cancer stem cell marker DCLK1 is an epigenetic biomarker for colorectal cancer. 2014;9: 346–350. doi:/10.4161/epi.27582

183. Sureban SM, May R, Weygant N, Qu D, Chandrakesan P, Bannerman-Menson E, et al. XMD8-92 inhibits pancreatic tumor xenograft growth via a DCLK1-dependent mechanism. Cancer Lett. 2014;351: 151–161. doi:10.1016/j.canlet.2014.05.011

184. Ferguson FM, Nabet B, Raghavan S, Liu Y, Leggett AL, Kuljanin M, et al. Discovery of a selective inhibitor of doublecortin like kinase 1. Nat Chem Biol. 2020;16: 635–643. doi:10.1038/s41589-020-0506-0

185. Sureban SM, May R, Mondalek FG, Qu D, Ponnurangam S, Pantazis P, et al. Nanoparticle-based delivery of siDCAMKL-1 increases microRNA-144 and inhibits colorectal cancer tumor growth via a Notch-1 dependent mechanism. J Nanobiotechnology. 2011;9: 1–13. doi:10.1186/1477-3155-9-40

186. Qu D, Weygant N, Yao J, Chandrakesan P, Berry WL, May R, et al. Overexpression of DCLK1-AL Increases Tumor Cell Invasion, Drug Resistance, and KRAS Activation and Can Be Targeted to Inhibit Tumorigenesis in Pancreatic Cancer. J Oncol. 2019;2019. doi:10.1155/2019/6402925

187. Sureban SM, Berahovich R, Zhou H, Xu S, Wu L, Ding K, et al. a Novel Treatment Strategy against Human Colorectal Cancers. : 1–17. doi:10.3390/cancers12010054

188. Gong J, Lin Y, Zhang H, Liu C, Cheng Z, Yang X, et al. Reprogramming of lipid metabolism in cancer-associated fibroblasts potentiates migration of colorectal cancer cells. Cell Death Dis. 2020;11. doi:10.1038/s41419-020-2434-z

189. Wang J, Xi MJ, Lu Q, Xia BH, Liu YZ, Yang JL. Single-cell and bulk transcriptomics identifies a tumor-specific CD36+ cancer-associated fibroblast subpopulation in colorectal cancer. Cancer Commun. 2023; 495–498. doi:10.1002/cac2.12506

190. Hale JS, Otvos B, Sinyuk M, Alvarado AG, Hitomi M, Stoltz K, et al. Cancer stem cell-specific scavenger receptor 36 drives glioblastoma progression. Stem Cells. 2014;32: 1746–1758. doi:10.1002/stem.1716

191. Drury J, Rychahou PG, He D, Jafari N, Wang C, Lee EY, et al. Inhibition of Fatty Acid Synthase Upregulates Expression of CD36 to Sustain Proliferation of Colorectal Cancer Cells. Front Oncol. 2020;10. doi:10.3389/fonc.2020.01185

192. Pascual G, Avgustinova A, Mejetta S, Martín M, Castellanos A, Attolini CSO, et al. Targeting metastasis-initiating cells through the fatty acid receptor CD36. Nature. 2017;541: 41–45. doi:10.1038/nature20791

193. Drury J, Rychahou PG, Kelson CO, Geisen ME, Wu Y, He D, et al. Upregulation of CD36, a Fatty Acid Translocase, Promotes Colorectal Cancer Metastasis by Increasing MMP28 and Decreasing E-Cadherin Expression. Cancers (Basel). 2022;14. doi:10.3390/cancers14010252

194. Fang Y, Shen ZY, Zhan YZ, Feng XC, Chen KL, Li YS, et al. CD36 inhibits β-catenin/c-myc-mediated glycolysis through ubiquitination of GPC4 to repress colorectal tumorigenesis. Nat Commun. 2019;10: 1–16. doi:10.1038/s41467-019-11662-3

195. Mahalingam D, Harb W, Patnaik A, Bullock A, Watnick RS, Vincent MY, et al. First-in-human phase I dose escalation trial of the first-in-class tumor microenvironment modulator VT1021 in advanced solid tumors. Communications Medicine. 2024;4. doi:10.1038/s43856-024-00433-x

196. Qian F, Sun BL, Zhang WY, Ke J, Zhu J. Gly82Ser polymorphism of the receptor for advanced glycation end-product (RAGE) potential high risk in patients with colorectal cancer. Tumor Biology. 2014;35: 3171–3175. doi:10.1007/s13277-013-1414-7

197. Kong SY, Takeuchi M, Hyogo H, McKeown-Eyssen G, Yamagishi SI, Chayama K, et al. The association between glyceraldehyde-derived advanced glycation end-products and colorectal cancer risk. Cancer Epidemiology Biomarkers and Prevention. 2015;24: 1855–1863. doi:10.1158/1055-9965.EPI-15-0422

198. Dahlmann M, Okhrimenko A, Marcinkowski P, Osterland M, Herrmann P, Smith J, et al. RAGE mediates S100A4-induced cell motility via MAPK/ERK and hypoxia signaling and is a prognostic biomarker for human colorectal cancer metastasis. Oncotarget. 2014;5: 3220–3233. doi:10.18632/oncotarget.1908

199. Huang CY, Chiang SF, Chen WTL, Ke TW, Chen TW, You YS, et al. HMGB1 promotes ERK-mediated mitochondrial Drp1 phosphorylation for chemoresistance through RAGE in colorectal cancer. Cell Death Dis. 2018;9. doi:10.1038/s41419-018-1019-6

200. Liang H, Zhong Y, Zhou S, Peng L. Knockdown of RAGE expression inhibits colorectal cancer cell invasion and suppresses angiogenesis in vitro and in vivo. Cancer Lett. 2011;313: 91–98. doi:10.1016/j.canlet.2011.08.028

201. Fuentes MK, Nigavekar SS, Arumugam T, Logsdon CD, Schmidt AM, Park JC, et al. RAGE activation by S100P in colon cancer stimulates growth, migration, and cell signaling pathways. Dis Colon Rectum. 2007;50: 1230–1240. doi:10.1007/s10350-006-0850-5

202. Liang H. Advanced glycation end products induce proliferation, invasion and epithelial-mesenchymal transition of human SW480 colon cancer cells through the PI3K/AKT signaling pathway. Oncol Lett. 2020;19: 3215–3222. doi:10.3892/ol.2020.11413

203. Taguchi A, Blood DC, Del Toro G, Canet A, Lee DC, Qu W, et al. Blockade of RAGE-amphoterin signalling suppresses tumour growth and metastases. Nature. 2000;405: 354–360. doi:10.1038/35012626

204. Kuniyasu H, Chihara Y, Takahashi T. Co-expression of receptor for advanced glycation end products and the ligand amphoterin associates closely with metastasis of colorectal cancer. Oncol Rep. 2003;10: 445–448. doi:10.3892/or.10.2.445

205. Zheng J, Zhu W, He F, Li Z, Cai N, Wang H. An Aptamer-Based Antagonist against the Receptor for Advanced Glycation End-Products (RAGE) Blocks Development of Colorectal Cancer. 2021;2021. doi:https://doi.org/10.1155/2021/9958051 Research

206. Huang J, Wang K, Xu J, Huang J, Zhang T. Prognostic significance of circulating tumor cells in non-small-cell lung cancer patients: A meta-analysis. PLoS One. 2013;8. doi:10.1371/journal.pone.0078070

207. De Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, et al. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. Clinical Cancer Research. 2008;14: 6302–6309. doi:10.1158/1078-0432.CCR-08-0872

208. Martini V, Timme-Bronsert S, Fichtner-Feigl S, Hoeppner J, Kulemann B. Circulating tumor cells in pancreatic cancer: Current perspectives. Cancers. MDPI AG; 2019. doi:10.3390/cancers11111659

209. Cohen SJ, Punt CJA, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, et al. Relationship of circulating tumor cells to tumor response, progression-

free survival, and overall survival in patients with metastatic colorectal cancer. Journal of Clinical Oncology. 2008;26: 3213–3221. doi:10.1200/JCO.2007.15.8923

210. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, et al. Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients With Nonmalignant Diseases. Clinical Cancer Research. 2004;10: 6897–904. doi:10.1158/1078-0432.CCR-04-0378

211. Bos MK, Kraan J, Starmans MPA, Helmijr JCA, Verschoor N, De Jonge MJA, et al. Comprehensive characterization of circulating tumor cells and cell-free DNA in patients with metastatic melanoma. Mol Oncol. 2024. doi:10.1002/1878-0261.13650

212. Lawrence R, Watters M, Davies CR, Pantel K, Lu YJ. Circulating tumour cells for early detection of clinically relevant cancer. Nat Rev Clin Oncol. 2023;20: 487–500. doi:10.1038/s41571-023-00781-y

213. Honrubia Gómez P. Localización de la célula madre tumoral circulante en sangre de pacientes con adenocarcinoma colorrectal mediante marcadores de superficie y regulación de su capacidad de autorrenovación. Universidad Castilla La Mancha. 2016.

214. Ugolkov Y, Nikitich A, Leon C, Helmlinger G, Peskov K, Sokolov V, et al. Mathematical modeling in autoimmune diseases: from theory to clinical application. Frontiers in Immunology. Frontiers Media SA; 2024. doi:10.3389/fimmu.2024.1371620

215. Jenner AL, Aogo RA, Davis CL, Smith AM, Craig M. Leveraging Computational Modeling to Understand Infectious Diseases. Current Pathobiology Reports. Springer; 2020. pp. 149–161. doi:10.1007/s40139-020-00213-x

216. Robben M, Hajighasemi A, Sadegh Nasr M, Prakesh Veerla J, Alsup AM, Rout B, et al. The State of Applying Artificial Intelligence to Tissue Imaging for Cancer Research and Early Detection. 2023. doi:10.12688/f1000research.139210.1

217. Iglesias G, Talavera E, Garcìa JT, Díaz-Álvarez A, Gracía-Remesal M. Artificial Intelligence Model for Tumoral Clinical Decision Support Systems. 2023. doi:10.1016/j.cmpb.2024.108228

218. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science. Springer; 2021. doi:10.1007/s42979-021-00592-x

219. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. Sci Rep. 2024;14. doi:10.1038/s41598-024-56706-x

220. Ahmed D, Eide PW, Eilertsen IA, Danielsen SA, Eknæs M, Hektoen M, et al. Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis. 2013;2. doi:10.1038/oncsis.2013.35

221. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. Nucleic Acids Res. 2012;40. doi:10.1093/nar/gks596

222. Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, et al. Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research. 2013;14: 2349–2353.

223. Mesas C, Chico MA, Doello K, Lara P, Moreno J, Melguizo C, et al. Experimental Tumor Induction and Evaluation of Its Treatment in the Chicken Embryo Chorioallantoic Membrane Model: A Systematic Review. International Journal of Molecular Sciences. Multidisciplinary Digital Publishing Institute (MDPI); 2024. doi:10.3390/ijms25020837

224. Tsimpaki T, Bechrakis NE, Seitz B, Kraemer MM, Liu H, Dalbah S, et al. Chick Chorioallantoic Membrane as a Patient-Derived Xenograft Model for Uveal Melanoma: Imaging Modalities for Growth and Vascular Evaluation. Cancers (Basel). 2023;15. doi:10.3390/cancers15051436

225. Efron Bradley, Tibshirani Robert. An introduction to the bootstrap. Chapman & Hall/CRC; 1998.

226. Cheng RCH. Chapter 14 Resampling Methods. Handbooks in Operations Research and Management Science. 2006;13: 415–453. doi:10.1016/S0927-0507(06)13014-5

227. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17: 261–272. doi:10.1038/S41592-019-0686-2

228. pandas development team T. pandas-dev/pandas: Pandas. Zenodo; 2024. doi:10.5281/zenodo.10957263

229. Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng. 2007;9: 90–95. doi:10.1109/MCSE.2007.55

230. Efron B, Tibshirani R. The bootstrap method for assesing statistical accuracy. Behaviormetrika. 1985.

231. Cecilia Subauste M, Kupriyanova TA, Conn EM, Ardi VC, Quigley JP, Deryugina EI. Evaluation of metastatic and angiogenic potentials of human colon carcinoma cells in chick embryo model systems. Clin Exp Metastasis. 2009;26: 1033–1047. doi:10.1007/s10585-009-9293-4

232. Cho SY. Patient-derived xenografts as compatible models for precision oncology. Laboratory Animal Research. BioMed Central Ltd; 2020. doi:10.1186/s42826-020-00045-1

233. How to determine an IC50 - FAQ 1859 - GraphPad. [cited 10 Sep 2024]. Available: https://www.graphpad.com/support/faq/how-to-determine-an-icsub50sub/

234. IC50 Calculator | AAT Bioquest. [cited 10 Sep 2024]. Available: https://www.aatbio.com/tools/ic50-calculator

235. Osterman E, Glimelius B. Recurrence risk after up-to-date colon cancer staging, surgery, and pathology: Analysis of the entire Swedish population. Dis Colon Rectum. 2018;61: 1016–1025. doi:10.1097/DCR.0000000000001158

236. Bouvier AM, Jooste V, Lillini R, Marcos-Gragera R, Katalinic A, Giorgi Rossi P, et al. Differences in survival and recurrence of colorectal cancer by stage across population-based European registries. Int J Cancer. 2024;155: 807–815. doi:10.1002/ijc.34944

237. Platell CFE. Changing patterns of recurrence after treatment for colorectal cancer. Int J Colorectal Dis. 2007;22: 1223–1231. doi:10.1007/s00384-007-0306-y

238. Rasola C, Laurent-Puig P, André T, Falcoz A, Lepage C, Aparicio T, et al. Time to recurrence and its relation to survival after recurrence in patients resected for stage III colon cancer. Eur J Cancer. 2023;194. doi:10.1016/j.ejca.2023.113321

239. Becerra AZ, Probst CP, Tejani MA, Aquina CT, González MG, Hensley BJ, et al. Evaluating the Prognostic Role of Elevated Preoperative

Carcinoembryonic Antigen Levels in Colon Cancer Patients: Results from the National Cancer Database. Ann Surg Oncol. 2016;23: 1554–1561. doi:10.1245/s10434-015-5014-1

240. Nicholson BD, Shinkins B, Pathiraja I, Roberts NW, James TJ, Mallett S, et al. Blood CEA levels for detecting recurrent colorectal cancer. Cochrane Database of Systematic Reviews. John Wiley and Sons Ltd; 2015. doi:10.1002/14651858.CD011134.pub2

241. Lee TH, Kim JS, Baek SJ, Kwak JM, Kim J. Diagnostic Accuracy of Carcinoembryonic Antigen (CEA) in Detecting Colorectal Cancer Recurrence Depending on Its Preoperative Level. Journal of Gastrointestinal Surgery. 2023;27: 1694–1701. doi:10.1007/s11605-023-05761-2

242. Huang CS, Chen CY, Huang LK, Wang WS, Yang SH. Prognostic value of postoperative serum carcinoembryonic antigen levels in colorectal cancer patients who smoke. PLoS One. 2020;15. doi:10.1371/journal.pone.0233687

243. Keim LM, Praus AW, Campbell WS, Samson KK, Mohr AM, Alcohol BL. Alcohol Consumption is associated with Increased CEA Levels in Alcohol Consumption is associated with Increased CEA Levels in Male Patients with Stage IV Colorectal Cancer-A Single-Institution Male Patients with Stage IV Colorectal Cancer-A Single-Institution Retrospective Analysis Retrospective Analysis Recommended Citation Recommended Citation. Graduate Medical Education Graduate Medical Education Research Journal Research Journal. 2021. doi:10.32873/unmc.dc.gmerj.3.1.001

244. Lin PC, Lin JK, Lin CC, Wang HS, Yang SH, Jiang JK, et al. Carbohydrate antigen 19-9 is a valuable prognostic factor in colorectal cancer patients with normal levels of carcinoembryonic antigen and may help predict lung metastasis. Int J Colorectal Dis. 2012;27: 1333–1338. doi:10.1007/s00384-012-1447-1

245. Li Z, Zhu H, Pang X, Mao Y, Yi X, Li C, et al. Preoperative serum CA19-9 should be routinely measured in the colorectal patients with preoperative normal serum CEA: a multicenter retrospective cohort study. BMC Cancer. 2022;22. doi:10.1186/s12885-022-10051-2

246. Kim S, Park BK, Seo JH, Choi J, Choi JW, Lee CK, et al. Carbohydrate antigen 19-9 elevation without evidence of malignant or pancreatobiliary diseases. Sci Rep. 2020;10. doi:10.1038/s41598-020-65720-8

247. Zhang Q, Zhang X, Xie P, Zhang W. Liquid biopsy: An arsenal for tumour screening and early diagnosis. Cancer Treatment Reviews. W.B. Saunders Ltd; 2024. doi:10.1016/j.ctrv.2024.102774

248. da Silva HEC, Santos GNM, Leite AF, Mesquita CRM, de Souza Figueiredo PT, Stefani CM, et al. The use of artificial intelligence tools in cancer detection compared to the traditional diagnostic imaging methods: An overview of the systematic reviews. PLoS One. 2023;18. doi:10.1371/journal.pone.0292063

249. Munro MJ, Wickremesekera SK, Peng L, Tan ST, Itinteang T. Cancer stem cells in colorectal cancer: a review. J Clin Pathol. 2018;71: 110–116. doi:10.1136/JCLINPATH-2017-204739

250. Soleimani A, Saeedi N, Al-Asady AM, Nazari E, Hanaie R, Khazaei M, et al. Colorectal Cancer Stem Cell Biomarkers: Biological Traits and Prognostic Insights. Curr Pharm Des. 2024;30: 1386–1397. doi:10.2174/0113816128291321240329050945

251. Manhas J, Bhattacharya A, Agrawal SK, Gupta B, Das P, Deo SVS, et al. Characterization of cancer stem cells from different grades of human colorectal cancer. Tumor Biology. 2016;37: 14069–14081. doi:10.1007/s13277-016-5232-6

252. Cherciu I, Bărbălan A, Pirici D, Mărgăritescu C, Săftoiu A. Stem cells, colorectal cancer and cancer stem cell markers correlations. Curr Health Sci J. 2014;40: 153–61. doi:10.12865/CHSJ.40.03.01

253. AbuSamra DB, Aleisa FA, Al-Amoodi AS, Ahmed HMJ, Chin CJ, Abuelela AF, et al. Not just a marker: CD34 on human hematopoietic stem/progenitor cells dominates vascular selectin binding along with CD44. Blood Adv. 2017;1: 2799–2816. doi:10.1182/bloodadvances.2017004317

254. Wang Z, Zhao K, Hackert T, Zöller M. CD44/CD44v6 a reliable companion in cancer-initiating cell maintenance and tumor progression. Front Cell Dev Biol. 2018;6: 1–28. doi:10.3389/fcell.2018.00097

255. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov G V., Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569: 503. doi:10.1038/S41586-019-1186-3

256. Xu Y, Zhang L, Wang Q, Zheng M. Comparison of Different Colorectal Cancer With Liver Metastases Models Using Six Colorectal Cancer Cell Lines. doi:10.1007/s12253-020-00805-3/Published

257. Hertweck MK, Erdfelder F, Kreuzer KA. CD44 in hematological neoplasias. Annals of Hematology. 2011. pp. 493–508. doi:10.1007/s00277-011-1161-z

258. Saeed U, Myklebust T, Robsahm TE, Kielland MF, Møller B, Skålhegg BS, et al. Risk and survival in colorectal cancer with increasing body mass index: A nationwide population-based cohort study. Colorectal Disease. 2023;25: 375–385. doi:10.1111/codi.16367

259. Denost Q, Quintane L, Buscail E, Martenot M, Laurent C, Rullier E. Short- and long-term impact of body mass index on laparoscopic rectal cancer surgery. Colorectal Disease. 2013;15: 463–469. doi:10.1111/codi.12026

260. He Y, Wang J, Bian H, Deng X, Wang Z. BMI as a Predictor for Perioperative Outcome of Laparoscopic Colorectal Surgery: A Pooled Analysis of Comparative Studies. Dis Colon Rectum. 2017;60: 433–445. doi:10.1097/DCR.0000000000000760

261. Gu YJ, Chen LM, Gu ME, Xu HX, Li J, Wu LY. Body mass index–based predictions and personalized clinical strategies for colorectal cancer in the context of PPPM. EPMA Journal. Springer Science and Business Media Deutschland GmbH; 2022. pp. 615–632. doi:10.1007/s13167-022-00306-0

262. Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, et al. Obesity and Risk of Colorectal Cancer: A Systematic Review of Prospective Studies. PLoS ONE. 2013. doi:10.1371/journal.pone.0053916

263. Matsuda M, Shimomura I. Increased oxidative stress in obesity: Implications for metabolic syndrome, diabetes, hypertension, dyslipidemia, atherosclerosis, and cancer. Obesity Research and Clinical Practice. 2013. doi:10.1016/j.orcp.2013.05.004

264. Cinciripini PM, Kypriotakis G, Blalock JA, Karam-Hage M, Beneventi DM, Robinson JD, et al. Survival Outcomes of an Early Intervention Smoking Cessation Treatment After a Cancer Diagnosis. JAMA Oncol. 2024. doi:10.1001/jamaoncol.2024.4890

265. Harris E. Cancer Risk Decreased 10 Years After Quitting Smoking. JAMA. 2024;331: 822. doi:10.1001/jama.2024.0732

266. Guraya SY. Pattern, Stage, and Time of Recurrent Colorectal Cancer After Curative Surgery. Clinical Colorectal Cancer. Elsevier Inc.; 2019. pp. e223–e228. doi:10.1016/j.clcc.2019.01.003

267. Boute TC, Swartjes H, Greuter MJE, Elferink MAG, van Eekelen R, Vink GR, et al. Cumulative Incidence, Risk Factors, and Overall Survival of Disease Recurrence after Curative Resection of Stage II–III Colorectal Cancer: A Population-based Study. Cancer Research Communications. 2024;4: 607–616. doi:10.1158/2767-9764.crc-23-0512

268. Qaderi SM, Galjart B, Cornelis Verhoef &, Slooter GD, Koopman M, Robert &, et al. Disease recurrence after colorectal cancer surgery in the modern era: a population-based study. doi:10.1007/s00384-021-03914-w/Published

269. Varlamova E V., Butakova MA, Semyonova V V., Soldatov SA, Poltavskiy A V., Kit OI, et al. Machine Learning Meets Cancer. Cancers. Multidisciplinary Digital Publishing Institute (MDPI); 2024. doi:10.3390/cancers16061100

270. Walcher L, Kistenmacher AK, Suo H, Kitte R, Dluczek S, Strauß A, et al. Cancer Stem Cells—Origins and Biomarkers: Perspectives for Targeted Personalized Therapies. Front Immunol. 2020;11: 1–33. doi:10.3389/fimmu.2020.01280

271. He B, Zhang HQ, Xiong SP, Lu S, Wan YY, Song RF. Changing patterns of serum CEA and CA199 for evaluating the response to first-line chemotherapy in patients with advanced gastric adenocarcinoma. Asian Pacific Journal of Cancer Prevention. 2015;16: 3111–3116. doi:10.7314/APJCP.2015.16.8.3111

272. Forster M, Deal AM, Page A, Vohra S, Wardell AC, Pak J, et al. Dose delay, dose reduction, and early treatment discontinuation in Black and White women receiving chemotherapy for nonmetastatic breast cancer. Oncologist. 2024. doi:10.1093/oncolo/oyae150

273. Schneck H, Gierke B, Uppenkamp F, Behrens B, Niederacher D, Stoecklein NH, et al. EpCAM-independent enrichment of circulating tumor cells in metastatic breast cancer. PLoS One. 2015;10. doi:10.1371/journal.pone.0144535

274. Pascual J, Attard G, Bidard FC, Curigliano G, De Mattos-Arruda L, Diehn M, et al. ESMO recommendations on the use of circulating tumour DNA assays for patients with cancer: a report from the ESMO Precision Medicine

Working Group. Annals of Oncology. 2022;33: 750–768. doi:10.1016/j.annonc.2022.05.520

275. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med. 2016;8. doi:10.1126/scitranslmed.aaf6219

276. Tarazona N, Gimeno-Valiente F, Gambardella V, Zuñiga S, Rentero-Garrido P, Huerta M, et al. Targeted next-generation sequencing of circulating-tumor DNA for tracking minimal residual disease in localized colon cancer. Annals of Oncology. 2019;30: 1804–1812. doi:10.1093/annonc/mdz390

277. Chen G, Peng J, Xiao Q, Wu HX, Wu X, Wang F, et al. Postoperative circulating tumor DNA as markers of recurrence risk in stages II to III colorectal cancer. J Hematol Oncol. 2021;14. doi:10.1186/s13045-021-01089-z

278. Fan W, Xia Z, Chen R, Lin D, Li F, Zheng Y, et al. Circulating tumor DNA analysis predicts recurrence and avoids unnecessary adjuvant chemotherapy in I–IV colorectal cancer. Ther Adv Med Oncol. 2024;16. doi:10.1177/17588359231220607

279. Kayikcioglu E, Onder AH, Bacak B, Serel TA. Machine learning for predicting colon cancer recurrence. Surg Oncol. 2024;54. doi:10.1016/j.suronc.2024.102079

280. Zhang L, Li Y, Yao L, He R, Wu J. Establishment and Clinical Significance of the Patient-Derived Xenograft Model of Colorectal Cancer. Cureus. 2024. doi:10.7759/cureus.71116

281. Kahounová Z, Pícková M, Drápela S, Bouchal J, Szczyrbová E, Navrátil J, et al. Circulating tumor cell-derived preclinical models: current status and future perspectives. Cell Death and Disease. Springer Nature; 2023. doi:10.1038/s41419-023-06059-6

282. Puchalapalli M, Zeng X, Mu L, Anderson A, Glickman LH, Zhang M, et al. NSG mice provide a better spontaneous model of breast cancer metastasis than athymic (nude) mice. PLoS One. 2016;11. doi:10.1371/journal.pone.0163521

283. Pizon M, Schott D, Pachmann U, Schobert R, Pizon M, Wozniak M, et al. Chick Chorioallantoic Membrane (CAM) Assays as a Model of Patient-

Derived Xenografts from Circulating Cancer Stem Cells (cCSCs) in Breast Cancer Patients. Cancers (Basel). 2022;14. doi:10.3390/cancers14061476

284. Rousset X, Maillet D, Grolleau E, Barthelemy D, Calattini S, Brevet M, et al. Embryonated Chicken Tumor Xenografts Derived from Circulating Tumor Cells as a Relevant Model to Study Metastatic Dissemination: A Proof of Concept. Cancers (Basel). 2022;14. doi:10.3390/cancers14174085

285. Krzywinska E, Cornillon A, Allende-Vega N, Vo DN, Rene C, Lu ZY, et al. CD45 isoform profile identifies natural killer (NK) subsets with differential activity. PLoS One. 2016;11. doi:10.1371/journal.pone.0150434

286. Al Barashdi MA, Ali A, McMullin MF, Mills K. Protein tyrosine phosphatase receptor type C (PTPRC or CD45). J Clin Pathol. 2021;74: 548–552. doi:10.1136/jclinpath-2020-206927

287. Ahmed MGT, Limmer A, Sucker C, Fares KM, Mohamed SAB, Othman AH, et al. Differential Regulation of CD45 Expression on Granulocytes, Lymphocytes, and Monocytes in COVID-19. J Clin Med. 2022;11. doi:10.3390/jcm11144219

288. Shimomura Y, Maruoka H, Ochi Y, Koba Y, Ono Y, Hiramoto N, et al. Detection of Cells with Low Side Scatter and Dimmer CD45 Expression after Allogeneic Hematopoietic Stem Cell Transplantation Predicts Good Survival. Blood. 2015;126: 3207–3207. doi:10.1182/blood.v126.23.3207.3207

289. Grillet F, Bayet E, Villeronce O, Zappia L, Lagerqvist EL, Lunke S, et al. Circulating tumour cells from patients with colorectal cancer have cancer stem cell hallmarks in ex vivo culture. Gut. 2017;66: 1802–1810. doi:10.1136/gutjnl-2016-311447

290. Yu H, Kim DJ, Choi HY, Kim SM, Rahaman MI, Kim YH, et al. Prospective pharmacological methodology for establishing and evaluating anti-cancer drug resistant cell lines. BMC Cancer. 2021;21: 1–13. doi:10.1186/S12885-021-08784-7/TABLES/6

291. Xu Y, Pachnikova G, Wang H, Wu Y, Przybilla D, Schäfer R, et al. IC50: an unsuitable measure for large-sized prostate cancer spheroids in drug sensitivity evaluation. Biomolecules and Biomedicine. 2022;22: 580–592. doi:10.17305/bjbms.2022.7279

292. Hafner M, Niepel M, Chung M, Sorger PK. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nature Methods 2016 13:6. 2016;13: 521–527. doi:10.1038/nmeth.3853

293. Larsson P, Engqvist H, Biermann J, Werner Rönnerman E, Forssell-Aronsson E, Kovács A, et al. Optimization of cell viability assays to improve replicability and reproducibility of cancer drug sensitivity screens. Scientific Reports 2020 10:1. 2020;10: 1–12. doi:10.1038/s41598-020-62848-5

294. Hazekawa M, Nishinakagawa T, Kawakubo-Yasukochi T, Nakashima M. Evaluation of IC50 levels immediately after treatment with anticancer reagents using a real-time cell monitoring device. Exp Ther Med. 2019. doi:10.3892/etm.2019.7876

295. Murphy H, McCarthy G, Dobrovolny HM. Understanding the effect of measurement time on drug characterization. PLoS One. 2020;15: e0233031. doi:10.1371/JOURNAL.PONE.0233031

296. Lindström HJG, Friedman R. Inferring time-dependent population growth rates in cell cultures undergoing adaptation. BMC Bioinformatics. 2020;21: 1–13. doi:10.1186/S12859-020-03887-7/TABLES/1

297. Vandermeer J. How Populations Grow: The Exponential and Logistic Equations | Learn Science at Scitable. 2010 [cited 10 Sep 2024]. Available: https://www.nature.com/scitable/knowledge/library/how-populations-grow-the-exponential-and-logistic-13240157/

298. Chenxi YIN, Wenhua LU, Mingzhe MA, Yang Q, Wenzhuo HE, Yumin HU, et al. Efficacy and mechanism of combination of oxaliplatin with PKM2 knockdown in colorectal cancer. Oncol Lett. 2020;20. doi:10.3892/OL.2020.12175

299. Zheng Z, Wu M, Li H, Xu W, Yang M, Pan K, et al. Downregulation of AC092894.1 promotes oxaliplatin resistance in colorectal cancer via the USP3/AR/RASGRP3 axis. BMC Med. 2023;21: 1–16. doi:10.1186/S12916-023-02826-6/FIGURES/8

300. Landrum GA, Riniker S. Combining IC50 or Ki Values from Different Sources Is a Source of Significant Noise. J Chem Inf Model. 2024;64: 1560–1567. doi:10.1021/ACS.JCIM.4C00049

301. Liu X, Deng J, Yuan Y, Chen W, Sun W, Wang Y, et al. Advances in Trop2-targeted therapy: Novel agents and opportunities beyond breast cancer.

Pharmacology and Therapeutics. Elsevier Inc.; 2022. doi:10.1016/j.pharmthera.2022.108296

302. Švec J, Šťastná M, Janečková L, Hrčkulák D, Vojtěchová M, Onhajzer J, et al. TROP2 Represents a Negative Prognostic Factor in Colorectal Adenocarcinoma and Its Expression Is Associated with Features of Epithelial–Mesenchymal Transition and Invasiveness. Cancers (Basel). 2022;14. doi:10.3390/cancers14174137

303. Foersch S, Schmitt M, Litmeyer AS, Tschurtschenthaler M, Gress T, Bartsch DK, et al. TROP2 in colorectal carcinoma: associations with histopathology, molecular phenotype, and patient prognosis. Journal of Pathology: Clinical Research. 2024;10. doi:10.1002/2056-4538.12394

304. Cubas R, Li M, Chen C, Yao Q. Trop2: A possible therapeutic target for late stage epithelial carcinomas. Biochim Biophys Acta Rev Cancer. 2009;1796: 309–314. doi:10.1016/j.bbcan.2009.08.001

305. Guerra E, Trerotola M, Tripaldi R, Aloisi AL, Simeone P, Sacchetti A, et al. Trop-2 induces tumor growth through AKT and determines sensitivity to AKT inhibitors. Clinical Cancer Research. 2016;22: 4197–4205. doi:10.1158/1078-0432.CCR-15-1701

306. Guerra E, Trerotola M, Relli V, Lattanzio R, Tripaldi R, Vacca G, et al. Trop-2 induces ADAM10-mediated cleavage of E-cadherin and drives EMT-less metastasis in colon cancer. Neoplasia (United States). 2021;23: 898–911. doi:10.1016/j.neo.2021.07.002

307. Sun X, Jia L, Wang T, Zhang Y, Zhao W, Wang X, et al. Trop2 binding IGF2R induces gefitinib resistance in NSCLC by remodeling the tumor microenvironment. J Cancer. 2021;12: 5310–5319. doi:10.7150/jca.57711

308. Goldenberg DM, Cardillo TM, Govindan S V., Rossi EA, Sharkey RM. Trop-2 is a novel target for solid cancer therapy with sacituzumab govitecan (IMMU-132), an antibody-drug conjugate (ADC) (Oncotarget (2015) 6 (22496-22512) DOI: 10.18632/oncotarget.4318). Oncotarget. 2020;11: 942. doi:10.18632/oncotarget.27512

309. Kuai X, Jia L, Yang T, Huang X, Zhao W, Zhang M, et al. TROP2 promotes multidrug resistance by regulating Notch1 signaling pathway in gastric cancer cells. Medical Science Monitor. 2020;26. doi:10.12659/MSM.919566

310. Koltai T, Fliegel L. The Relationship between Trop-2, Chemotherapeutic Drugs, and Chemoresistance. Int J Mol Sci. 2024;25. doi:10.3390/ijms25010087

311. Ahmed D, Eide PW, Eilertsen IA, Danielsen SA, Eknæs M, Hektoen M, et al. Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis. 2013;2. doi:10.1038/oncsis.2013.35

312. Ministerio de Sanidad. Documento de instrucciones de la Agencia Española de Medicamentos y Productos Sanitarios (AEMPS) para la realización de ensayos clínicos en España. 2015.

Annexes

A.1. Ethical committee approvals

INSTITUTO DE INVESTIGACIÓN SANITARIA FUNDACIÓN JIMÉNEZ DÍAZ

EOH089-20_FJD

CONFORMIDAD DE LA DIRECCIÓN DEL CENTRO

D. Juan Antonio Álvaro de la Parra, Presidente de la Junta Directiva de la Fundación Instituto de Investigación Sanitaria – Fundación Jiménez Díaz (FIIS-FJD), y vista la autorización del Comité Ético de Investigación Clínica,

CERTIFICA

Que conoce la propuesta para que sea realizado en este Centro el Estudio titulado: **"Detección de posibles células iniciadoras de tumores (CITs) en sangre periférica de pacientes con cáncer colorrectal (CCR)"**, el cual está previsto que sea dirigido en el **Servicio de Oncología del Hospital Fundación Jiménez Díaz** por la **Dra. Ana León Carbonero.**

Que acepta la realización de dicho estudio clínico en este Centro.

Lo que firma en Madrid a 4 de diciembre de 2020

D. Juan Antonio Álvaro de la Parra
Presidente Junta Directiva IIS-FJD

Vº Bº

Dra. Lucía Llanos Jiménez
Secretaria Técnica CEImFJD

Hospital Universitario Ramón y Cajal SaludMadrid	Comité de Ética de la Investigación con medicamentos

DICTAMEN DEL COMITE DE ÉTICA DE LA INVESTIGACIÓN CON MEDICAMENTOS

Dª. Itziar de Pablo López de Abechuco, Secretaria del COMITÉ DE ÉTICA DE LA INVESTIGACIÓN CON MEDICAMENTOS DEL HOSPITAL UNIVERSITARIO RAMÓN Y CAJAL

CERTIFICA

Que se ha evaluado la propuesta del promotor referida al ESTUDIO:

TÍTULO:

DETECCIÓN DE POSIBLES CÉLULAS INICIADORAS DE TUMORES (TICS) EN SANGRE PERIFÉRICA DE PACIENTES CON CÁNCER COLORRECTAL (CRC).

Protocolo versión 2 de 01 de marzo de 2019
HIP/CI versión 3 de 28 de abril de 2019

-El estudio se plantea siguiendo los requisitos legalmente establecidos, y su realización es pertinente.
-Se cumplen los requisitos necesarios de idoneidad del protocolo en relación con los objetivos del estudio y están justificados los riesgos y molestias previsibles para el sujeto.
-El alcance de las compensaciones económicas previstas no interfiere con el respeto a los postulados éticos.
-La capacidad del investigador y sus colaboradores, y las instalaciones y medios disponibles, tal y como ha sido informado, son apropiados para llevar a cabo el estudio.

Este CEIm actuando como comité evaluador, emite dictamen **FAVORABLE** y acepta que dicho estudio sea realizado en los centros siguientes por los investigadores principales que se relacionan a continuación:

Dra. Vanesa Pachón Olmos / Hospital Universitario Ramón y Cajal

Y HACE CONSTAR QUE:

1° En la reunión celebrada el día **28 de enero de 2019 ACTA nº 358** se decidió emitir el informe correspondiente al estudio de referencia.
2° En dicha reunión se cumplieron los requisitos establecidos en la legislación vigente y se cumplieron las normas de funcionamiento interno del Comité para que la decisión del citado CEIm sea válida.
3° El CEIm, tanto en su composición, como en los PNT cumple con las normas de BPC (CPMP/ ICH/ 135/95)
4° La composición actual del CEIm es la siguiente:

PRESIDENTE	Sonsoles Sancho García.	Oncología Radioterápica
VICEPRESIDENTE	Mª Ángeles Gálvez	Farmacología Clínica
SECRETARIO TECNICO	Itziar de Pablo López de Abechuco	Farmacología Clínica
VOCALES	Mª Jesús Blanchard Rodriguez	Hematología
	Irene Cabrera Troye	Miembro Lego, ajeno a la investigación biomédica/asistencia clínica
	Yolanda de Blas Fernández	DUE Servicio de Urgencias
	Guillermo García Ribas	Neurología

Ctra. Colmenar, km. 9,100 - 28034 MADRID (Planta -2 Dcha) *E-mail: ceic.hrc@salud.madrid.org* 1
Tlfº.: 91 336 8322

HOSPITAL UNIVERSITARIO DE TOLEDO

sescam
Servicio de Salud de Castilla-La Mancha

CONFORMIDAD JEFE DE SERVICIO

Dr/Dra: Dra. Dª CARMEN ESTEBAN ESTEBAN
Servicio: ONCOLOGÍA MÉDICA
Nombre del Centro: HOSPITAL UNIVERSITARIO DE TOLEDO

Declaro:

Que en relación al estudio: ***Detección de posibles células iniciadoras de tumores (CITs) en sangre periférica de pacientes con cáncer colorrectal (CCR)***

Y cuyo investigador principal es:

SE HACE CONSTAR:

- Que cuenta con los recursos humanos necesarios para llevar a cabo el Estudio, sin que ello interfiera en la realización de otro tipo de estudios ni en otras tareas que tiene habitualmente encomendadas.
- En prueba de conformidad, firmo el presente documento para su constancia y archivo,

JEFE DE SERVICIO	Firma	Fecha
Dr/Dra: CARMEN ESTEBAN ESTEBAN		14/12/2022

Gerencia Atención Integrada

sescam

INFORME DEL COMITÉ DE ÉTICA DE INVESTIGACIÓN CON MEDICAMENTOS

El Comité de Ética de Investigación con medicamentos de la Gerencia de Atención Integrada de Ciudad Real, en su reunión del día 24/05/2022 (Acta nº 05/2022),

Acordó la:

APROBACIÓN DEFINITIVA
Nota.-Durante la evaluación y decisión final estuvo ausente D. José Ramón Muñoz Rodríguez.

Del estudio descrito a continuación:

Titulo: **"Detección de posibles células iniciadoras de tumores (TICS) en sangre periférica de pacientes con cáncer colorrectal (CRC)".**
Código interno: C-532
Protocolo: sin versión/fecha
HIP/CI: sin versión/fecha
Investigadores principales:

- Leticia Serrano Oviedo. Servicio: Unidad de Investigación Traslacional. GAICR.
- José Carlos Villa Guzmán. Servicio: Oncología Médica. GAICR

COMITÉ DE ÉTICA DE INVESTIGACIÓN CON MEDICAMENTOS

En Ciudad Real, a 24 de Mayo de 2022

Fdo.: Pilar Vicente Sánchez
Presidente CEIm

Comunidad de Madrid

Gerencia Asistencial
de Atención Primaria
CONSEJERÍA DE SANIDAD

COMISIÓN LOCAL DE INVESTIGACIÓN CENTRO

DICTAMEN DE LA COMISIÓN LOCAL DE INVESTIGACIÓN CENTRO

TÍTULO: **““Modelo de machine-learning (aprendizaje automático) encaminado a la detección precoz de recaída en cáncer colorrectal. ONCOREC””.**

CÓDIGO de proyecto: 12/23-C

INVESTIGADORAS: Carmen Ramírez Catillejo. Profesor Titular. Departamento de Biotecnología y Biología Vegetal / Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas. Universidad Politécnica de Madrid. Ivan de los Mozos Hernando Médico de familia CS Daroca
Centro de salud Daroca

La Comisión Local de Investigación Centro, **INFORMA FAVORABLEMENTE** sobre la realización de dicho estudio en los centros de salud pertenecientes a la Dirección Asistencial Centro de la Gerencia de Atención Primaria de la Comunidad de Madrid.

En Madrid, a 28 de junio de 2023

GERENCIA ATENCIÓN PRIMARIA
SaludMadrid
C/ Alberto Palacios, 22 - MADRID

Fdo.: Dr. Ángel Alberquilla Menéndez-Asenjo
Presidente de la Comisión Local de Investigación Centro

UNIVERSIDAD
POLITÉCNICA
DE MADRID

INFORME QUE EMITE EL COMITÉ DE ÉTICA

DE LA UNIVERSIDAD POLITÉCNICA DE MADRID

A petición de la Vicerrectora de Investigación, Innovación y Doctorado de la Universidad Politécnica de Madrid,

Dª Asunción Gómez Pérez, de un dictamen sobre los aspectos éticos del

Proyecto **"Desarrollo de un kit de detección de células iniciadoras de tumores para predicción de recaídas"**, cuyo investigador principal es Dña. Carmen Ramírez Castillejo, del Departamento de Biotecnología BV/ Grupo de Tecnologías para Ciencias de la Salud, en el Centro de Tecnología Biomédica. El Comité emite el siguiente informe:

A la vista que:

- Los investigadores acreditan capacidad y medios suficientes para realizar el estudio, y en concreto en relación con el nivel de seguridad de las instalaciones y los procedimientos utilizados.

- Se cumplen los requisitos necesarios de idoneidad del protocolo en relación con los objetivos del estudio y están justificados los riesgos y molestias previsibles para el sujeto.

- El estudio es una colaboración entre la Universidad Politécnica de Madrid y el Hospital Universitario Ramón y Cajal para la identificación de moléculas marcadoras que permitan identificar la presencia de células iniciadoras de tumores

El solicitante declara conocer los principios éticos y las normas legales que rigen las actividades de investigación y se compromete a respetar estos principios y normas en el desarrollo del estudio propuesto.

En particular declara conocer y atenerse a lo dispuesto en
- la Declaración de Helsinki de la Asociación Médica Mundial (64ª Asamblea General, Fortaleza, Brasil, octubre 2013)
- el Convenio de Oviedo relativo a los Derechos Humanos y la Biomedicina (Consejo de Europa, 1997)
- la Ley 41/2002 básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica.
- la Circular 7/2004 de la AEMPS sobre investigaciones clínicas con productos sanitarios
- la Ley 14/2007 de Investigación biomédica
- el RD 1591/2009 que regula los productos sanitarios

UNIVERSIDAD POLITÉCNICA DE MADRID Vicerrectorado de Investigación, Innovación y Doctorado
Ramiro de Maeztu, 7. 28040 Madrid Tel +34 91 06 70175 secretaria.adjunto.vinvestigacion@upm.es

UNIVERSIDAD
POLITÉCNICA
DE MADRID

POLITÉCNICA

- el RD 1616/2009 que se regula los productos sanitarios implantables activos
-Ley 33/2011, de 4 de octubre, General de Salud Pública
- el RD 1716/2011 de biobancos y muestras biológicas
- el RD 9/2014 sobre la donación, la obtención, la evaluación, el procesamiento, la preservación, el almacenamiento y la distribución de células y tejidos humanos
- el RD 1090/2015 de ensayos clínicos con medicamentos

Así mismo, el solicitante declara conocer la legislación relativa a la protección de datos, en particular el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos, y se compromete a respetar la confidencialidad de los datos personales de este estudio, así como a hacer explícito este compromiso ante las personas que participan en el proyecto.

Igualmente, se compromete a no modificar los protocolos de investigación y a solicitar una nueva autorización en caso de modificación.

Por lo tanto, no hay inconveniente en **informar favorablemente la propuesta**, recomendando el efectivo cumplimiento de los compromisos declarados.

Y para que así conste, firmo el presente informe en Madrid a 25 de septiembre de dos mil veinte.

La Presidenta del Comité

GOMEZ PEREZ ASUNCION DE MARIA - DNI 08821929A

Firmado digitalmente por GOMEZ PEREZ ASUNCION DE MARIA - DNI 08821929A
Fecha: 2020.09.28 08:36:21 +02'00'

Asunción Gómez Pérez

El Secretario del Comité

Juan Carlos Dueñas López

A.2. Publications

Article published 6 May 2022

biomedicines

MDPI

Article

Implication of Different Tumor Biomarkers in Drug Resistance and Invasiveness in Primary and Metastatic Colorectal Cancer Cell Lines

Marta Sánchez-Díez [1,2,*], Nicolás Alegría-Aravena [1], Marta López-Montes [1], Josefa Quiroz-Troncoso [1,2], Raquel González-Martos [1,2], Adrián Menéndez-Rey [1], José Luis Sánchez-Sánchez [3], Juan Manuel Pastor [2] and Carmen Ramírez-Castillejo [1,2,4,*]

1 CTB (CTB-UPM) Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, 28223 Pozuelo de Alarcón, Spain; nicolas.alegria.aravena@gmail.com (N.A.-A.); marta.lopez.montes@alumnos.upm.es (M.L.-M.); josefa.quiiroz@gmail.com (J.Q.-T.); raquel.gonzalez@ctb.upm.es (R.G.-M.); menendez.rey.adrian@gmail.com (A.M.-R.)
2 Grupo de Sistemas Complejos, Universidad Politécnica de Madrid, 28040 Madrid, Spain; juanmanuel.pastor@upm.es
3 Unidad de Oncología, Hospital de Almansa, 02640 Albacete, Spain; jlsanchez_sanchez@hotmail.com
4 ETSIAAB, Departamento Biotecnología-Biología Vegetal, Universidad Politécnica de Madrid, IdISSC, 28040 Madrid, Spain
* Correspondence: marta.sanchez@ctb.upm.es (M.S.-D.); mariadelcarmen.ramirez@upm.es (C.R.-C.)

Article published 27 August 2022

International Journal of *Molecular Sciences*

MDPI

Article

Effectiveness of Gold Nanorods of Different Sizes in Photothermal Therapy to Eliminate Melanoma and Glioblastoma Cells

Javier Domingo-Diez [1], Lilia Souiade [1], Vanesa Manzaneda-González [2], Marta Sánchez-Díez [1,3], Diego Megias [4], Andrés Guerrero-Martínez [2], Carmen Ramírez-Castillejo [1,3,5,6], Javier Serrano-Olmedo [1,7,8] and Milagros Ramos-Gómez [1,7,8,9,*]

1 Center for Biomedical Technology (CTB), Universidad Politécnica de Madrid (UPM), 28040 Madrid, Spain; javier.domingo@ctb.upm.es (J.D.-D.); marta.sanchez@ctb.upm.es (M.S.-D.); carmen.ramirez@ctb.upm.es (C.R.-C.)
2 Departamento de Química Física, Universidad Complutense de Madrid, Avenida Complutense s/n, 28040 Madrid, Spain; aguerrero@quim.ucm.es (A.G.-M.)
3 Grupo de Sistemas Complejos, Universidad Politécnica de Madrid, 28040 Madrid, Spain
4 Advanced Optical Microscopy Unit, UCCTs, Instituto de Salud Carlos III (ISCIII), 28222 Madrid, Spain
5 Departamento Biotecnología-B.V. ETSIAAB, Universidad Politécnica de Madrid, 28040 Madrid, Spain
6 Departamento de Oncología, Instituto de Investigación Sanitaria San Carlos (IdISSC), 28040 Madrid, Spain
7 Centro de Investigación Biomédica en Red para Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, 28029 Madrid, Spain
8 Departamento de Tecnología Fotónica y Bioingeniería, ETSI Telecomunicaciones, Universidad Politécnica de Madrid, 28040 Madrid, Spain
9 Experimental Neurology Unit, Center for Biomedical Technology, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, Pozuelo de Alarcón, 28223 Madrid, Spain
* Correspondence: milagros.ramos@ctb.upm.es

Article published 15 November 2023

cancers MDPI

Article

Self-Renewal Inhibition in Breast Cancer Stem Cells: Moonlight Role of PEDF in Breast Cancer

Carmen Gil-Gas [1,†], Marta Sánchez-Díez [2,†], Paloma Honrubia-Gómez [1], Jose Luis Sánchez-Sánchez [3,4], Carmen B. Alvarez-Simón [1,4], Sebastia Sabater [4], Francisco Sánchez-Sánchez [5,*] and Carmen Ramírez-Castillejo [2,6,*]

1 Centro Regional de Investigaciones Biomédicas, Universidad de Castilla-La Mancha, 02006 Albacete, Spain; mc.gilgas@edu.gva.es (C.G.-G.); paloma.honrubia@uclm.es (P.H.-G.)
2 HST Group, Department Biotechnology-BV, Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, 28040 Madrid, Spain; marta.sanchez@ctb.upm.es
3 Oncology Unit, Hospital General de Almansa, 02640 Albacete, Spain; jluisss@sescam.jccm.es
4 Complejo Hospitalario Universitario de Albacete, 02006 Albacete, Spain
5 Laboratory of Medical Genetic, Faculty of Medicine, Instituto de Investigaciones en Discapacidades Neurológicas (IDINE), University of Castilla La-Mancha, 02006 Albacete, Spain
6 Oncology Group, Instituto de Investigación Sanitaria San Carlos, 28040 Madrid, Spain
* Correspondence: francisco.ssanchez@uclm.es (F.S.-S.); mariadelcarmen.ramirez@upm.es (C.R.-C.)
† These authors contributed equally to this work.

Review published 18 July 2024

frontiers | Frontiers in Immunology

TYPE Review
PUBLISHED 18 July 2024
DOI 10.3389/fimmu.2024.1333150

Check for updates

OPEN ACCESS

EDITED BY
Sina Naserian,
Hôpital Paul Brousse, France

REVIEWED BY
Antonio Di Stasi,
University of Alabama at Birmingham, United States
Mohamed Essameldin Abdelgawad,
Wake Forest University, United States

*CORRESPONDENCE
Carmen Ramírez-Castillejo
carmen.ramirez@ctb.upm.es
Raquel González-Martos
raquel.gonzalez@ctb.upm.es

RECEIVED 04 November 2023
ACCEPTED 14 June 2024
PUBLISHED 18 July 2024

CITATION
Ercilla-Rodríguez P, Sánchez-Díez M, Alegría-Aravena N, Quiroz-Troncoso J, Gavira-O'Neill CE, González-Martos R and Ramírez-Castillejo C (2024) CAR-T

CAR-T lymphocyte-based cell therapies; mechanistic substantiation, applications and biosafety enhancement with suicide genes: new opportunities to melt side effects

Paula Ercilla-Rodríguez[1], Marta Sánchez-Díez[1,2], Nicolás Alegría-Aravena[3,4], Josefa Quiroz-Troncoso[1,2], Clara E. Gavira-O'Neill[2,5], Raquel González-Martos[1,2*] and Carmen Ramírez-Castillejo[1,2,5*]

[1]ETSIAAB, Universidad Politécnica de Madrid, Madrid, Spain, [2]Laboratorio Cancer Stem Cell, HST group, Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, Madrid, Spain, [3]Grupo de Biología y Producción de Cérvidos, Instituto de Desarrollo Regional, Universidad de Castilla-La Mancha, Albacete, Spain, [4]Asociación Española Contra el Cáncer (AECC)-Fundación Científica AECC, Albacete, Spain, [5]Sección de Oncología, Instituto de Investigación Sanitaria San Carlos, Madrid, Spain

Article published 13 February 2025

pharmaceutics

MDPI

Article

Assessment of Cell Viability in Drug Therapy: IC50 and Other New Time-Independent Indices for Evaluating Chemotherapy Efficacy

Marta Sánchez-Díez [1,2,*], Paula Romero-Jiménez [1], Nicolás Alegría-Aravena [3,4], Clara E. Gavira-O'Neill [1,2,5], Elena Vicente-García [1], Josefa Quiroz-Troncoso [1], Raquel González-Martos [1,2,5], Carmen Ramírez-Castillejo [1,2,6,7,*] and Juan Manuel Pastor [2,8]

1 CTB (CTB-UPM) Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, 28223 Pozuelo de Alarcón, Spain; paula.romero.jimenez@alumnos.upm.es (P.R.-J.); clara.gavira@ctb.upm.es (C.E.G.-O.); elena.vicente.garcia@alumnos.upm.es (E.V.-G.); josefa.quiroz@ctb.upm.es (J.Q.-T.); raquel.gonzalez@ctb.upm.es (R.G.-M.)

2 Grupo de Sistemas Complejos, Universidad Politécnica de Madrid, 28040 Madrid, Spain; juanmanuel.pastor@upm.es

3 Instituto de Desarrollo Regional (IDR) and Instituto de Investigación en Recursos Cinegéticos (IREC), Universidad de Castilla-La Mancha (UCLM), 02071 Albacete, Spain; nicolas.alegria@uclm.es

4 Asociación Española Contra el Cáncer (AECC)-Fundación Científica AECC, 02004 Albacete, Spain

5 Nageru S.L., 28045 Madrid, Spain

6 Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Departamento Biotecnología-Biología Vegetal, Universidad Politécnica de Madrid, 28040 Madrid, Spain

7 Departamento de Oncología, Instituto de Investigación Sanitaria San Carlos (IdISSC), 28040 Madrid, Spain

8 Grupo Interdisciplinar de Sistemas Complejos (GISC), 28040 Madrid, Spain

* Correspondence: marta.sanchez@ctb.upm.es (M.S.-D.); mariadelcarmen.ramirez@upm.es (C.R.-C.)