

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  <https://creativecommons.org/licenses/?lang=ca>

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <https://creativecommons.org/licenses/?lang=es>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

Identification of novel candidate biomarkers of response to chemotherapeutic agents in patients with colorectal cancer

By
Li Jing

For the Degree of

DOCTORAL THESIS

Biochemistry, Molecular Biology and Biomedicine program
Faculty of Medicine
Autonomous University of Barcelona

The thesis was carried out at
Group of Biomedical Research in Digestive Tract Tumors
Center for Molecular Biology and Biochemistry (CIBBIM-Nanomedicine)

Vall d'Hebron Research Institute

and

Group of Molecular Oncology
Biomedical Research Institute of Lleida

2023

Director

Dr. Diego Arango del Corro

Director

Dr. Águeda Martínez Barriocana

Tutor

Dr. Ana Meseguer Navarro

PhD candidate

Li Jing

Table of Contents

.....	1
<i>Index of figures</i>	6
<i>Index of tables</i>	9
LIST OF ALL ABBREVIATIONS	10
ABSTRACT	14
1. Introduction	20
1.1 General aspects of cancer	20
1.1.1 Epidemiology	20
1.1.2 Etiology	20
1.1.3 Types of cancer	22
1.1.4 Oncogenes and tumor suppressor genes	22
1.1.5 Epigenetic alterations of cancer.....	25
1.2 Anatomy and histology of the intestinal tract	27
1.2.1 The small intestine.....	27
1.2.2 The large intestine	32
1.3 Colorectal cancer	33
1.3.1 Epidemiology of colorectal cancer.....	33
1.3.2 Histologic and anatomical classification of colorectal cancer	36
1.3.3 Tumorigenesis and deregulation of signaling pathways in colorectal cancer	38
1.3.4 Molecular classification of colorectal cancer.....	51
1.3.5 Treatment of colorectal cancer	57
1.3.5.5 Immunotherapy	62
1.3.6 Resistance in Cancer Chemotherapy	67
1.3.7 Biomarkers: diagnostic, prognostic and predictive	70
2. Hypotheses and aims of the study	<i>Error! Bookmark not defined.</i>
3. Material and Methods	<i>Error! Bookmark not defined.</i>
3.1 Materials	75
3.1.1 Cell lines	75
3.1.2 Tumor samples	78

3.1.3	Antibodies.....	78
3.1.4	Primers.....	79
3.1.5	Plasmids.....	79
3.1.6	Therapeutic Agents.....	79
3.1.7	OMICS Data.....	80
3.1.8	Tissue microarrays (TMAs).....	82
3.2	Methods.....	83
3.2.1	Growth Inhibition Assay.....	83
3.2.2	Classification of cell lines according to drug response.....	85
3.2.3	Functional group enrichment analysis.....	86
3.2.4	Removing batch effects in high-throughput experiments.....	87
3.2.5	Principle component analysis (PCA).....	87
3.2.6	Clustering analysis.....	87
3.2.7	Transfection.....	88
3.2.8	Western blotting.....	89
3.2.9	Immunohistochemistry.....	89
3.2.10	Quantification of the cell line tissue microarrays (TMAs).....	90
3.2.11	Patients grouping.....	90
3.2.12	Normalization of immunohistochemistry in patients samples staining.....	91
3.2.13	Biomarker cutoff determination.....	91
3.2.14	Statistical analysis.....	91
4	Results.....	<i>Error! Bookmark not defined.</i>
4.1	Characterization of the sensitivity of colorectal cancer cell lines to chemotherapeutic agents 93	
4.1.1	- Sensitivity of colorectal cancer cell lines to 5-FU.....	98
4.1.2	- Sensitivity of colorectal cancer cell lines to oxaliplatin.....	100
4.1.3	- Sensitivity of colorectal cancer cell lines to irinotecan.....	101
4.1.4	- Comparison of the different metrics of drug sensitivity.....	108
4.1.5	Classification of colon cancer cell lines into two clinically relevant categories: responders and non-responders.....	109
4.2	Association of drug sensitivity with colorectal cancer molecular subtypes.....	115
4.2.1	- Colorectal cancer molecular subtypes and response to 5-FU.....	115
4.2.2	- Colorectal cancer molecular subtypes and response to oxaliplatin.....	116
4.2.3	- Colorectal cancer molecular subtypes and response to irinotecan.....	117

4.3 Association between drug sensitivity and mutations in “driver” genes.....	118
4.3.1 - Driver gene mutations and sensitivity to 5-FU.....	118
4.3.2 - Driver gene mutations and sensitivity to oxaliplatin	119
4.3.3 - Driver gene mutations and sensitivity to irinotecan.....	120
4.4 Genomic, transcriptomic and proteomic datasets of a large panel of colorectal cancer cell lines	121
4.4.1 – Genomic data of colorectal cancer cell lines.....	121
4.4.2 – Transcriptomic data of colorectal cancer cell lines	121
4.4.3 – Proteomic data of colorectal cancer cell lines.	125
4.5 Association between OMICs data and drug response.....	126
4.5.1 - Associations between OMICs data and sensitivity to 5-FU	127
4.5.2 - Associations between OMICs data and sensitivity to oxaliplatin	129
4.5.3 - Associations between OMICs data and sensitivity to irinotecan	131
4.6 Functional group enrichment analysis	133
4.6.1 Functional group enrichment analysis of genes differentially expressed in responder and non-responder cell lines to 5-FU	134
4.6.2 Functional group enrichment analysis of genes differentially expressed in responder and non-responder cell lines to oxaliplatin	135
4.6.3 Functional group enrichment analysis of genes differentially expressed in responder and non-responder cell lines to irinotecan.....	136
4.6.4 Common enriched functional group.....	138
4.7 Potential biomarkers of drug response.....	139
4.8 Validation of the potential of SLC29A1 protein expression to predict response to irinotecan-based treatment for colorectal cancer patients.....	141
4.8.1 Validation of SLC29A1 mRNA levels as a biomarker of response to irinotecan in an external cohort of 221 metastatic colorectal cancer patients.....	142
4.8.2 Validation of SLC29A1 protein levels as a biomarker of response to irinotecan in a large cohort of metastatic colorectal cancer patients.....	143
4.9 Validation of the potential of GUCY2C expression to predict response to irinotecan-based treatment for colorectal cancer patients.....	160
4.9.1 Validation of GUCY2C mRNA levels as a biomarker of response to irinotecan in an external cohort of 221 metastatic colorectal cancer patients.....	160
4.9.2 Validation of SLC29A1 protein levels as a biomarker of response to irinotecan in a large cohort of metastatic colorectal cancer patients.....	161

5	<i>Discussion</i>	<i>Error! Bookmark not defined.</i>
5.1	Characterization of drug sensitivity of colorectal cancer cell lines	175
5.1.1	Cell line as a tool to identify predictive biomarkers	175
5.1.2	Characterization of cell lines' sensitivity to 5-FU, oxaliplatin, and irinotecan	177
5.2	Consensus molecular subtype 1 and microsatellite instability are associated with a higher response rate to irinotecan	180
5.3	KRAS wild type is associated with a higher response rate to irinotecan	182
5.4	Omics data merging	183
5.5	Association of OMICS data with response.	184
5.6	Validation of potential biomarkers	185
5.7	Validation of potential biomarkers in colorectal cancer patient's samples	186
	<i>Conclusion</i>	<i>Error! Bookmark not defined.</i>
	<i>Appendix 1</i>	192
	<i>Appendix 2</i>	233
	<i>Appendix 3</i>	237
	<i>Appendix 4</i>	242
	<i>Appendix 5</i>	248
	<i>Appendix 6</i>	249
	<i>REFERENCES</i>	<i>Error! Bookmark not defined.</i>

Index of figures

Figure 1: Cancer statistics. _____	20
Figure 2: Schematic diagram of the histology of the small intestine. _____	28
Figure 3: The structure of the small and large intestine and the major cell types of the epithelium. _____	30
Figure 4: Anatomy of the large intestine. _____	32
Figure 5: The trend of colorectal cancer incidence rate across countries. _____	33
Figure 6: Proportion of colorectal cancer cases associated with sporadic and hereditary factors. _____	35
Figure 7: Staging of colorectal cancer. _____	38
Figure 8: The hallmarks of cancer and therapeutic targets. _____	39
Figure 9: Schematic diagram of colorectal cancer development through two approaches. _____	40
Figure 10: Signaling pathways alterations in colorectal cancer. _____	41
Figure 11: Overview of canonical and non-canonical WNT signaling. _____	42
Figure 12: Schematic representation of RTK-RAS signaling pathway in cancer. _____	45
Figure 13: PI3K pathway and its major downstream effectors. _____	47
Figure 14: TGF- β signaling pathway. _____	48
Figure 15: The P53 signaling pathway. _____	50
Figure 16: Diagram showing the mechanism of Microsatellite instability (MSI) and CpG island methylator phenotype (CIMP). _____	54
Figure 17: Consensus molecular subtypes classification of colorectal cancer. _____	56
Figure 18: Mechanism of action and metabolism of 5-FU, oxaliplatin and irinotecan. _____	60
Figure 19: Schedule of the SRB method to characterize the drug sensitivity of cancer cell lines (irinotecan as the example). _____	86
Figure 20: Dose-response curve and drug sensitivity parameters. _____	94
Figure 21: Comparison of 5-FU sensitivity measured in overlapping cell lines between the pharmacologic datasets. _____	96
Figure 22: Comparison of oxaliplatin sensitivity measured in overlapping cell lines between the pharmacologic datasets. _____	97
Figure 23: Comparison of irinotecan sensitivity measured in overlapping cell lines between the pharmacologic datasets. _____	97
Figure 24: 5-FU sensitivity metrics for the colorectal cancer cell lines panel. _____	99
Figure 25: Oxaliplatin sensitivity metrics for the colorectal cancer cell lines panel. _____	101
Figure 26: Irinotecan sensitivity metrics for the panel of colorectal cancer cell lines . _____	102
Figure 27: Comparison of 5-FU sensitivity measured in overlapping cell lines between different datasets. _	104
Figure 28. Figure 28: Comparison of oxaliplatin sensitivity measured in overlapping cell lines between the different datasets. _____	105
Figure 29: Comparison of irinotecan sensitivity measured in overlapping cell lines between the different datasets. _____	107
Figure 30: Spearman's correlation coefficients of the five sensitivity metrics. _____	109

<i>Figure 31: Classification of the sensitivity of colorectal cancer lines to 5-FU, oxaliplatin, and irinotecan according to cancer cell line Response Evaluation Criteria in Solid Tumors (cclRECIST).</i>	112
<i>Figure 32: Distribution of cell line categorized by cancer cell line Response Evaluation Criteria (cclRECIST) and IC50.</i>	114
<i>Figure 33: Comparison of cancer cell line Response Evaluation Criteria (cclRECIST) and IC50.</i>	114
<i>Figure 34: Association of 5-FU response with known colorectal cancer molecular subtypes.</i>	116
<i>Figure 35: Association of oxaliplatin response with known colorectal cancer molecular subtypes.</i>	116
<i>Figure 36: Association of irinotecan response with known colorectal cancer molecular subtypes.</i>	118
<i>Figure 37: Association of mutations in colorectal cancer driver genes with the 5-FU response.</i>	119
<i>Figure 38: Association of mutations in colorectal cancer driver genes with the oxaliplatin response.</i>	119
<i>Figure 39: Association of mutations in colorectal cancer driver genes with the irinotecan response.</i>	120
<i>Figure 40: Principal component analysis (PCA) plots showed mRNA expression variability across 285 cell lines from four datasets (RNAseq-1, RNAseq-2, microarray-1, microarray-2) before (a) and after (b) conducting ComBat.</i>	123
<i>Figure 41: Hierarchical clustering of transcriptomes based on 5000 genes with the highest variations.</i>	124
<i>Figure 42: The similarity between Proteomics-1 and Proteomics-2 shotgun proteomics data and batch correction.</i>	125
<i>Figure 43: Hierarchical clustering analysis of proteome from two datasets after ComBat merging.</i>	126
<i>Figure 44: Association of transcriptome and proteome expression with responder or non-responder colorectal cancer cell lines to 5-FU.</i>	129
<i>Figure 45: Association of transcriptome and proteome expression with responder or non-responder colorectal cancer cell lines to oxaliplatin.</i>	131
<i>Figure 46: Association of transcriptome and proteome expression with responder or non-responder colorectal cancer cell lines to irinotecan.</i>	133
<i>Figure 47: Functional group enrichment analysis of genes differentially expressed in responder and non-responder cell lines to 5-FU.</i>	134
<i>Figure 48: GO terms and KEGG pathway enrichment plot using proteomes between responder and non-responder cell lines to irinotecan.</i>	135
<i>Figure 49: GO terms and KEGG pathway enrichment plot using proteomes between responder and non-responder cell lines to irinotecan.</i>	136
<i>Figure 50: GO terms and KEGG pathway enrichment plot using proteomes between responder and non-responder cell lines to irinotecan.</i>	137
<i>Figure 51: Intersection between the 5-FU, oxaliplatin, and irinotecan functional group analysis.</i>	138
<i>Figure 52: Validation of SLC29A1 predictive power in an external cohort of 221 metastatic colorectal cancer patients treated with irinotecan.</i>	142
<i>Figure 53: Biomarker validation workflow.</i>	143
<i>Figure 54: Validation of SLC29A1.</i>	145
<i>Figure 55: Representative images of the pattern of SLC29A1 immunostaining in tissue microarrays from colorectal cancer patients.</i>	148

<i>Figure 56: Distribution of SLC29A1 immunohistochemistry staining score in colorectal cancer patients from the three cohorts.</i>	148
<i>Figure 57: Survival analysis of SLC29A1 based on colorectal cancer patient samples from three cohorts.</i>	149
<i>Figure 58: Association between objective response to irinotecan-based treatment and SLC29A1 protein levels.</i>	152
<i>Figure 59: Survival analysis of SLC29A1 based on colorectal cancer patients samples from Madrid patients cohort.</i>	155
<i>Figure 60: Survival analysis of SLC29A1 based on colorectal cancer patients samples from Barcelona patients cohort.</i>	156
<i>Figure 61: Survival analysis of SLC29A1 based on colorectal cancer patients samples from Germany cohort.</i>	157
<i>Figure 62: Association between objective response to irinotecan-based treatment and SLC29A1 expression.</i>	159
<i>Figure 63: Validation of GUCY2C predictive power in an external cohort of 221 metastatic colorectal cancer patients treated with irinotecan.</i>	161
<i>Figure 64: Validation of the anti-GUCY2C antibody in formalin-fixed, paraffin-embedded samples.</i>	163
<i>Figure 65: Representative images of the pattern of GUCY2C immunostaining in tissue microarrays from colorectal cancer patients.</i>	164
<i>Figure 66: Distribution of GUCY2C immunohistochemistry staining score in colorectal cancer patients from the three cohorts.</i>	165
<i>Figure 67: Survival analysis of colorectal cancer patients based on GUCY2C levels on tumor patient samples from three cohorts.</i>	165
<i>Figure 68: Association between objective response to irinotecan-based treatment and GUCY2C protein levels.</i>	168
<i>Figure 69: Survival analysis of GUCY2C based on colorectal cancer patients samples from Madrid patient cohort.</i>	170
<i>Figure 70: Survival analysis of GUCY2C based on colorectal cancer patients samples from Barcelona patient cohort.</i>	171
<i>Figure 71: Survival analysis of GUCY2C based on colorectal cancer patients samples from Germany patient cohort.</i>	172
<i>Figure 72: Association between objective response to irinotecan-based treatment and SLC29A1 expression.</i>	174

Index of tables

<i>Table 1: Frequent therapeutic regimens used for advanced colorectal cancer treatment and response rates.</i>	61
<i>Table 2: Colorectal cancer treatment regimens by stage.</i>	64
<i>Table 3: Cell lines used in this study.</i>	76
<i>Table 4: Primers used in this study.</i>	79
<i>Table 5: Comparison of drug response measures from different datasets for 5-FU.</i>	95
<i>Table 6: Comparison of drug response measures from different datasets for oxaliplatin.</i>	95
<i>Table 7: Comparison of drug response measures from different datasets for irinotecan.</i>	95
<i>Table 8: Association of mutations with responders and non-responders to 5-FU.</i>	128
<i>Table 9: Association of mutations with responders and non-responders to oxaliplatin.</i>	130
<i>Table 10: Association of mutations with irinotecan responder and non-responder cell lines.</i>	131
<i>Table 11: Representative antibodies recognizing the potential biomarkers in immunohistochemistry.</i>	140
<i>Table 12: Baseline patients and tumors clinicopathological features from three cohorts of samples analyzed.</i>	146
<i>Table 13: Association between patients and tumors clinicopathological features and patient survival.</i>	150
<i>Table 14: Multivariate Cox regression of SLC29A1 in overall survival analysis.</i>	151
<i>Table 15: Multivariate Cox regression of SLC29A1 in progression-free survival analysis.</i>	151
<i>Table 16: Association between patient and tumors clinicopathological features and irinotecan-based treatment responders and non-responders.</i>	153
<i>Table 17: Multivariate logistic regression of SLC29A1.</i>	154
<i>Table 18: Multivariate Cox regression of GUCY2C in overall survival analysis.</i>	166
<i>Table 19: Multivariate Cox regression of GUCY2C in progression-free survival analysis.</i>	167
<i>Table 20: Multivariate logistic regression of GUCY2C.</i>	168

LIST OF ALL ABBREVIATIONS

ALL: Acute lymphoblastic leukemia
AML: Acute myeloid leukemia
APC: Adenomatous polyposis coli
AGRF: Australian Genome Research Facility
ABCB1: ATP Binding Cassette Subfamily B Member 1
AURKA: Aurora kinase A
BRAF: B-Raf Proto-Oncogene
bRPLC: Basic reversed-phase high-pressure liquid chromatography
CES: Carboxylesterases
CCLE: Cancer Cell Line Encyclopedia
CIN: Chromosomal instability
CLL: Chronic lymphocytic leukemia
CML: Chronic myeloid leukemia
CRC: Colorectal cancer
CMS: Consensus Molecular Subtypes
CT: Computerized tomography
CDKN1A: Cyclin dependent kinase inhibitor 1A
CDK: Cyclin-dependent kinase
CKI: Cyclin-dependent kinase inhibitor
CR: Complete response
DACH: Diamino cyclohexane
DAVID: Database for Annotation, Visualization and Integrated Discovery
DMEM: Dulbecco's Modified Eagle Medium
DNMT: DNA methyltransferases
EGF: Epidermal growth factor
EGFR: Epidermal growth factor receptor
EGFR: Epidermal growth factor receptors
EMT: Epithelial-mesenchymal transition
EBV: Epstein-Barr virus
ERCC1: Excision repair cross-complementation group 1
FAP: Familial adenomatous polyposis

FBS: Fetal bovine serum
FdUMP: Fluorodeoxyuridine monophosphate
FdUTP: Fluorodeoxyuridine triphosphate
FUTP: Fluorouridine triphosphate
FFPE: Formalin-fixed paraffin-embedded tissue
G2/M: G2/M checkpoint arrest
GR: Growth rate
E: Glutamic acid
Q61: Glutamine at position 61
G12: Glycine at position 12
GRB2: Growth factor receptor-bound protein 2
GAPs: GTPase activating proteins
GEFs: Guanine nucleotide exchange factors
HDAC1: Histone deacetylases 1
HPLC: High-performance liquid chromatography
IHC: Immunohistochemistry
IBD: Inflammatory bowel disease
ICD-O: International Classification of Diseases for Oncology
KRAS: Kirsten Rat Sarcoma Viral Oncogene Homologue
LV: Leucovorin
LFS: Li-Fraumeni syndrome
LC-MS/MS: Liquid chromatography-tandem mass spectrometry
LD: Longest diameter
LOH: Loss of heterozygosity
MRI: Magnetic resonance imaging
MEM: Minimum Essential Media
MSI: Microsatellite instability
MSI-H: Microsatellite instability-high
MSI-L: Microsatellite instability-low
MSS: Microsatellite stable
MMR: Mismatch repair
MSH2: Mismatch repair protein 2

MLH1: MutL Homolog 1
MUT: Mutant
OPRT: Orotate phosphoribosyl transferase
OD510: Optical Density at 510 nm
OS: Overall survival
PBS: Phosphate-buffered saline
PCA: Principal component analysis
PR: Partial Response
PI3K: Phosphoinositide 3-kinase
PD: Progressive Disease
PFS: Progression-free survival
PKB: Protein kinase B
PTK: Protein tyrosine kinase
RTKs: Receptor tyrosine kinases
RECIST: Response Evaluation Criteria in Solid Tumors
SETD1A: SET Domain Containing 1A
RMA: Robust Multichip Average
SSRs: Simple sequence repeats
SNV: Single-nucleotide variant
SMAD4: Mothers Against Decapentaplegic Homolog 4
SOS: Son of sevenless
SD: Stable Disease
SRB: Sulforhodamine B
TMT: Tandem Mass Tag
TMA: Tissue microarrays
TP: Thymidine phosphorylase
TOP1: Topoisomerase I
TP53 The Tumor Protein P53:
TGF- β : Transforming growth factor-beta
TCA: Trichloroacetic acid solution
TAS-102: Trifluridine and tipiracil
TNF: Tumor necrosis factor

V: V600E - valine

VEGF: Vascular endothelial growth factor

VEGFR: Vascular endothelial growth factor receptor

WHO: World Health Organization

WB: Western blotting

WES: Whole-exome sequencing

WT: Wild type

TITLE

Identification of novel candidate biomarkers of response to chemotherapeutic agents in patients with colorectal cancer

ABSTRACT

Colorectal cancer (CRC) is the third most common type of cancer worldwide. Most CRC patients receive chemotherapeutic treatment during the course of their disease. However, the efficacy of chemotherapeutic compounds as single agents is not greater than 30%. To improve the clinical management of CRC patients, and due to significant side effects and high medical costs, it is essential to identify predictive biomarkers of response to treatment in patients with CRC.

To identify novel molecular features of the tumors that could predict response to treatment, first, we determined the *in vitro* sensitivity as a single agent to 5-Fluorouracil (5-FU) and oxaliplatin in a panel of 57 colorectal cancer cell lines and to irinotecan in an extended panel of 90 cell lines. Then, we adapted the standard RECIST (Response Evaluation Criteria in Solid Tumors) classification of patients' response to treatment to classify the cell line's sensitivity (cancer cell line Response Evaluation Criteria in Solid Tumors – cclRECIST). In addition, high-throughput expression (RNA-sequencing, microarray and shogun proteomics) and mutation (exome-sequencing) data were obtained for the CRC cell lines to identify candidate molecular biomarkers of response/resistance to these chemotherapeutic agents.

The response rate to the irinotecan was found to be significantly lower in cell lines with microsatellite stable (MSS) compared to microsatellite instable (MSI) tumors, lower in KRAS mutant compared to KRAS wild type tumors, and lower in Consensus Molecular subtype (CMS) 2 compared to CMS1 and CMS4 tumors. Moreover, we identified 16 proteins with >2-fold significant (FDR <0.2) difference in their expression between cell lines that are resistant and sensitive to irinotecan.

Two of these proteins, the solute carrier SLC29A1 and the guanylate cyclase GUCY2C, were then selected for further validation using a large cohort of 548 tumors from metastatic

colorectal cancer patients that received irinotecan-based treatment. Before assessing the relative levels of protein expression using an immunohistochemical approach on sections of a tissue microarray, the antibodies used were thoroughly validated. The expression of these two proteins was then determined by immunohistochemistry in the cohort of 548 primary colorectal tumors and scored by a trained pathologist. The expression of SLC29A1 or GUCY2C was not associated with the objective response rates or the progression-free survival in this cohort of samples. However, a significant association was observed between the expression of GUCY2C and the overall survival of the patients in this cohort.

In conclusion, this study provides a thorough characterization of the sensitivity to 5-fluorouracil, oxaliplatin and irinotecan in a large panel of CRC cell lines and potential predictive biomarkers of response to these agents were identified. Although the predictive value of two of these biomarkers could not be confirmed in a large cohort of colorectal cancer patients, future studies are needed to validate the capacity of other candidate biomarkers in the clinical setting. Importantly, the expression of GUCY2C was significantly correlated with the overall survival of colorectal cancer patients, and could therefore be used as a prognostic biomarker capable of identifying a subset of patients with poor prognosis and that are good candidates to receive a more aggressive treatment.

TÍTULO

Identificación de nuevos biomarcadores candidatos de respuesta a agentes quimioterapéuticos en pacientes con cáncer colorrectal

RESUMEN

El cáncer colorrectal es el tercer tipo de cáncer más común en todo el mundo. La mayoría de los pacientes con cáncer colorrectal reciben tratamiento quimioterapéutico durante el curso de su enfermedad. Sin embargo, la eficacia de los compuestos quimioterapéuticos como agentes únicos no supera el 30%. Para mejorar el manejo clínico de los pacientes con cáncer colorrectal, y debido a los importantes efectos secundarios y los altos costes médicos, es fundamental identificar biomarcadores predictivos de respuesta al tratamiento en pacientes con cáncer colorrectal.

Para identificar nuevas características moleculares de los tumores que podrían predecir la respuesta al tratamiento, primero determinamos la sensibilidad como agente único a 5-fluorouracilo (5-FU) y oxaliplatino en un panel de 57 líneas celulares de cáncer colorrectal y a irinotecán en un panel extendido de 90 líneas celulares. Luego, adaptamos la clasificación estándar RECIST (Response Evaluation Criteria in Solid Tumors) de la respuesta de los pacientes, para clasificar la sensibilidad de las líneas celulares (cancer cell line Response Evaluation Criteria in Solid Tumors -ccIRECIST). Además, se obtuvieron datos de alto rendimiento de expresión (secuenciación de ARN, microarrays y proteómica de shogun) y mutación (secuenciación del exoma) para las líneas celulares de cáncer colorrectal con el objetivo de identificar posibles biomarcadores moleculares de respuesta/resistencia a estos agentes quimioterapéuticos.

Los resultado del estudio demostraron que la tasa de respuesta al irinotecán era significativamente más baja en las líneas celulares con microsatélites estables (MSS) en comparación con las líneas con microsatélites inestables (MSI), más baja en los tumores KRAS mutantes en comparación con los tumores KRAS de tipo salvaje, y más baja en el subtipo

molecular de consenso (CMS) 2 en comparación con los tumores CMS1 y CMS4. Además, identificamos 16 proteínas con una diferencia significativa (FDR <0,2) de >2 veces en su expresión entre líneas celulares que son resistentes y sensibles a irinotecán.

A continuación, se seleccionaron dos de estas proteínas, el transportador de soluto SLC29A1 y la guanilato ciclasa GUCY2C, para una validación en mayor profundidad utilizando una cohorte de 548 tumores de pacientes con cáncer colorrectal metastásico que recibieron tratamiento basado en irinotecán. Antes de evaluar los niveles relativos de expresión de proteínas mediante un enfoque inmunohistoquímico en secciones de un microarray de tejido (TMA), los anticuerpos utilizados se validaron minuciosamente. A continuación, la expresión de estas dos proteínas se determinó mediante inmunohistoquímica en la cohorte de 548 tumores colorrectales primarios que fue valorada por un patólogo. No se encontraron asociaciones entre la expresión de SLC29A1 o GUCY2C y las tasas de respuesta objetiva o la supervivencia libre de progresión en esta cohorte de muestras. Sin embargo, se observó una asociación significativa entre la expresión de GUCY2C y la supervivencia global de los pacientes de esta cohorte.

En conclusión, este estudio proporciona una caracterización exhaustiva de la sensibilidad a 5-fluorouracilo, oxaliplatino e irinotecán en un panel de líneas celulares de cáncer colorrectal y se identificaron posibles biomarcadores predictivos de respuesta a estos agentes. Aunque no se pudo confirmar el valor predictivo de dos de estos biomarcadores en una gran cohorte de pacientes con cáncer colorrectal, en estudios futuros se validará la capacidad de otros biomarcadores candidatos en el entorno clínico. Es importante destacar que la expresión de GUCY2C se correlacionó significativamente con la supervivencia de los pacientes con cáncer colorrectal y, por lo tanto, podría usarse como un biomarcador de pronóstico capaz de identificar un subgrupo de pacientes con mal pronóstico y que son buenos candidatos para recibir un tratamiento más agresivo.

TÍTOL

Identificació de nous biomarcadors candidats de resposta a agents quimioterapèutics en pacients amb càncer colorectal

RESUM

El càncer colorectal és el tercer tipus de càncer més comú a tot el món. La majoria dels pacients amb càncer colorectal reben tractament quimioterapèutic durant el curs de la seva malaltia. Tot i això, l'eficàcia dels compostos quimioterapèutics com a agents únics no supera el 30%. Per millorar el maneig clínic dels pacients amb càncer colorectal, i a causa dels importants efectes secundaris i els alts costos mèdics, és fonamental identificar biomarcadors predictius de resposta al tractament en pacients amb càncer colorectal.

Per identificar noves característiques moleculars dels tumors que podrien predir la resposta al tractament, primer vàrem determinar la sensibilitat com a agent únic a 5-fluorouracil (5-FU) i oxaliplatí en un panell de 57 línies cel·lulars de càncer colorectal, i a irinotecà en un panell estès de 90 línies cel·lulars. Després, adaptàrem la classificació estàndard RECIST (Response Evaluation Criteria in Solid Tumors) emprada per definir el grau de resposta dels pacients, per classificar la sensibilitat de les línies cel·lulars (cancer cell line Response Evaluation Criteria in Solid Tumors -ccIRECIST). A més, vàrem obtenir dades d'alt rendiment d'expressió (seqüenciació d'ARN, microarrays i proteòmica de shotgun) i mutació (seqüenciació de l'exoma) per a les línies cel·lulars de càncer colorectal per identificar possibles biomarcadors moleculars de resposta/resistència a aquests agents quimioterapèutics.

Els resultats de l'estudi van demostrar que la taxa de resposta a l'irinotecà era significativament més baixa a les línies cel·lulars amb estabilitat de microsatèl·lits (MSS) en comparació amb les línies amb inestabilitat de microsatèl·lits (MSI), també més baixa als tumors *KRAS* mutants en comparació amb els tumors *KRAS* salvatge, i més baixa en el subtipus molecular de consens (CMS) 2 en comparació amb els tumors CMS1 i CMS4. A més, vàrem identificar 16 proteïnes amb una diferència significativa (FDR <0,2) de >2 vegades en la seva expressió entre línies cel·lulars resistents i sensibles a irinotecà.

A continuació, es van seleccionar dues d'aquestes proteïnes, el transportador de solut SLC29A1 i la guanilat ciclasa GUCY2C, per a una validació en més profunditat utilitzant una cohort de 548 tumors de pacients amb càncer colorectal metastàtic que van rebre tractament basat en irinotecà. Abans d'avaluar els nivells relatius d'expressió de proteïnes mitjançant un enfocament immunohistoquímic en seccions histològiques de microarray de teixit (TMA), els anticossos utilitzats van ser validats minuciosament. A continuació, l'expressió d'aquestes dues proteïnes es va determinar mitjançant immunohistoquímica a la cohort de 548 tumors colorectals primaris que va ser valorada per un patòleg. No es van trobar associacions entre l'expressió de SLC29A1 o GUCY2C i les taxes de resposta objectiva o la supervivència lliure de progressió en aquesta cohort de mostres. Tot i això, es va observar una associació significativa entre l'expressió de GUCY2C i la supervivència global dels pacients d'aquesta cohort.

En conclusió, aquest estudi proporciona una caracterització exhaustiva de la sensibilitat a 5-fluorouracil, oxaliplatí i irinotecà en un panell de línies cel·lulars de càncer colorectal i ha permès identificar possibles biomarcadors predictius de resposta a aquests agents. Tot i que no es va poder confirmar el valor predictiu de dos d'aquests biomarcadors en una gran cohort de pacients amb càncer colorectal, en estudis futurs es validarà la capacitat d'altres biomarcadors candidats a l'entorn clínic. És important destacar que l'expressió de GUCY2C es va correlacionar significativament amb la supervivència dels pacients amb càncer colorectal i, per tant, es podria fer servir com un biomarcador de pronòstic capaç d'identificar un subgrup de pacients amb mal pronòstic i que són bons candidats per rebre un tractament més agressiu.

1. INTRODUCTION

1.1 General aspects of cancer

1.1.1 Epidemiology

The burden of cancer morbidity and mortality is rapidly growing. An estimated 19.3 million new cancer cases and nearly 10 million cancer deaths occurred globally, according to the World Health Organization (WHO) data released in 2020 (<https://gco.iarc.fr/>) (Sung et al., 2021). In 112 out of the 183 countries analyzed in the study, cancer was the first or second leading cause of death in the population younger than 70 years old. Breast cancer remains the most commonly diagnosed cancer (11.7%), followed by lung (11.4%), colorectal (10.0%), prostate (7.3%), and stomach (5.6%) cancer. The top five causes of cancer death are lung (18%), colorectal (9.4%), liver (8.3%), stomach (7.7%), and breast (6.9%) cancer (**Figure 1**).

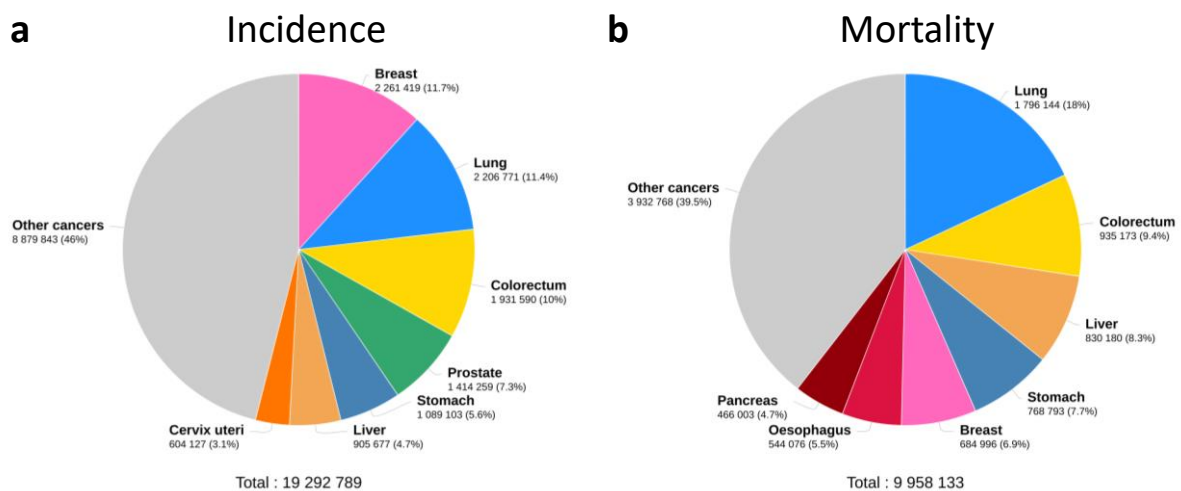


Figure 1: Cancer statistics.

Distribution of the incidence (**a**) and mortality (**b**) for the most common cancer types worldwide in 2020 for both sexes (GLOBOCAN 2020, <https://gco.iarc.fr/>).

1.1.2 Etiology

Cancer is a group of diseases characterized by the uncontrolled growth and spread of abnormal cells. From the standpoint of pathology, tumor cells exhibit a proliferation capacity that exceeds that of normal cells; they damage the organ where they originate and then spread to other parts

of the body, where they resume growth, causing further damage (National Cancer Institute <https://www.cancer.gov/>, 2021). From a cell biology point of view, cancer is a genetic disease (Vogelstein & Kinzler, 2004). Cancer is caused by specific changes in genes that control key cell functions, such as cell growth and division.

The search for the cause of cancer has been carried out for centuries. Active research in this field has identified many risk factors that can cause cancer. Environmental risk factors are one of the most important aspects contributing to cancer onset and development. The major ones are described below.

Age. Although cancer can be diagnosed at any age, most patients are frequently diagnosed after 60 years old. The median age of diagnosis of breast, colorectal, lung and prostate cancer is 62, 67, 71 and 66 years, respectively (National Cancer Institute <https://www.cancer.gov/>, 2021).

Alcohol. Alcohol consumption increases cancer risk, including mouth, throat, larynx, esophagus, colorectal, liver and breast cancer. It has a synergistic carcinogenic effect with tobacco.

Smoking. Tobacco consumption is not only associated with lung cancer but also with other types of cancer, including bladder, kidney, pancreatic and cervical cancers.

Obesity. It has been identified as a risk factor for prostate, endometrium, and colorectal cancer. Free radicals caused by excess body fat may play a role in carcinogenesis.

Chronic inflammation. It can be triggered by chemical, bacterial, viral, and irritants, among others. Chronic inflammation can lead to DNA damage and increase cancer risk in the long-term. For example, colon cancer is highly associated with chronic inflammatory bowel diseases like ulcerative colitis and Crohn's disease.

Infectious agents. They include viruses, bacteria, and parasites, contributing to carcinogenesis by affecting cell signaling and disrupting the immune system. For example, the *Epstein-Barr virus* (EBV) increases the risk of Burkitt's lymphoma by causing DNA damage, dysfunction of telomeres, and genome instability (Graham & Lynch, 2021).

1.1.3 Types of cancer

Cancer can be classified histologically according to the cell of origin. The International Classification of Diseases for Oncology (ICD-O), grouped cancer into six major categories based on their histology.

- 1) **Carcinomas** originate from epithelial cells. They account for 80-90% of all cancer cases and include adenocarcinoma and squamous cell carcinoma subtypes. Adenocarcinomas develop in organs or glands, while squamous cell carcinomas originate in the squamous epithelia. The most common carcinomas include breast, lung, colon, and prostate cancer.
- 2) **Sarcomas** arise in supportive and connective tissues such as bones, tendons, cartilage, muscle, and fat. There are about 70 types of sarcomas.
- 3) **Myelomas** derive from plasma cells in the bone marrow. Plasma cells, also called plasma B cells, are leukocytes that generate antibodies to build the immune system. In myeloma, abnormal plasma cells produce non-functional antibodies. In turn, anemia, infections such as pneumonia and pyelonephritis may occur along the course of the disease.
- 4) **Leukemias** originate in the bone marrow and are manifested as a high proportion of abnormal blood cells. The most common leukemias are acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL) and chronic myeloid leukemia (CML).
- 5) **Lymphomas** generate from the glands or nodes of the lymphatic system. Lymphomas affect the lymph nodes, spleen, thymus gland and bone marrow.
- 6) **Mixed Types** include cancers with at least two histological origins, *e.g.*, adenosquamous carcinoma.

1.1.4 Oncogenes and tumor suppressor genes

Cancer is a disease caused by genetic and environmental factors (*e.g.*, age, alcohol, tobacco, diet, chemicals, and infections). Two types of genes called proto-oncogenes and tumor suppressor genes play essential roles in regulating cell growth and proliferation and, thus, are

key in the onset and progression of the carcinogenic process. Proto-oncogenes refer to genes that exist in the genome of normal cells to promote cell growth and division but can be transformed into cancerous oncogenes under certain stimuli. Under normal circumstances, the expression of these proto-oncogenes is tightly controlled and performs critical physiological functions. However, upon viral infection, exposure to chemical carcinogens or radiation, proto-oncogenes can be abnormally activated or mutated and transformed into oncogenes, contributing to the cancerous transformation of cells. Tumor suppressor genes regulate several cellular functions, such as cell proliferation, cell differentiation, migration, tumor angiogenesis and DNA repair. Tumor suppressor genes prevent damaged cells from replicating (Sherr, 2004; Joyce et al., 2021). Without functional tumor suppressor genes, the risk of dysregulated cell growth is high, thus leading to the development of cancers.

1.1.4.1 Oncogenes

A large number of oncogenes have been identified to date, many of which are responsible for providing positive signals leading to enhanced cell proliferation. Mutation in either one or two alleles of an oncogene can promote uncontrolled cell growth. That is why oncogenes are regarded as dominant. In contrast, tumor suppressor genes normally must display defects in both alleles to contribute to the oncogenic process and therefore, are regarded as recessive.

There are different mechanisms by which a proto-oncogene can be converted into an oncogene (Haite, 2001):

- 1) **Gene amplification.** Gene duplication leads to gene overexpression which in turn can cause increased protein levels within a cell. For instance, *MYC* is a proto-oncogene that codes for a transcription factor and is found frequently upregulated in breast, colorectal, pancreatic, gastric, and uterine cancers. *MYC* is thought to function by increasing the expression of many genes involving cell proliferation and growth, and is therefore, regarded as a promising target for anti-cancer drugs (Duffy et al., 2021).
- 2) **Chromosomal Rearrangements.** Through chromosomal translocation, the transcriptional control of proto-oncogenes can be lost and thereby cause a high level of the gene product. This is the case of the t(8;14) translocation, which is observed in 70%

of Burkitt's lymphoma patients and leads to the translocation of the *MYC* proto-oncogene to chromosome 14, which in turn activates *MYC* and induces abnormal cell proliferation (Haluska et al., 1987).

- 3) **Activation by point mutations.** A mutation within a proto-oncogene can lead to an increase in protein activity or the loss of protein regulation. In approximately 27% of all human cancers, the Ras family of small GTPases are activated by point mutations at codons 12, 13 or 61, which will cause enhanced Ras signaling (Prior et al., 2012).
- 4) **Activation by the production of chimeric gene products.** Chimeric genes form by combining portions of two or more coding sequences to produce new genes. In 95% of chronic myeloid leukemia patients, the translocation t(9; 22) results in the *BCR/ABL* gene fusion, which causes a cell to grow and divide uncontrollably (Heisterkamp et al., 1985).

1.1.4.2 Tumor suppressor genes

Unlike oncogenes, most tumor suppressor genes must have defects in both alleles to cause a phenotype change, also known as the “two-hit” hypothesis (Knudson Jr, 1971). If only one allele of the gene is damaged, the other can typically still produce enough of the right protein to maintain a proper function. Non-hereditary retinoblastoma usually develops late in children compared to hereditary retinoblastoma because two mutations must occur before the tumor develops. In children with hereditary retinoblastoma, the first mutation in the *RB1* gene is inherited, and the second mutation is acquired, which is why children with this type develop it earlier. Tumor suppressor genes inhibit cell proliferation when cells grow uncontrollably, and therefore, the loss of function due to mutations or other mechanisms contributes to cancer initiation or progression. They can be classified into four types according to their various mechanisms of action (Wang et al., 2018).

- 1) **Cell cycle-related genes.** These genes regulate gene expression at a specific stage of the cell cycle. The *INK4* tumor suppressor gene is a cyclin-dependent kinase inhibitor (CKI). It encodes the Cyclin-dependent kinase (CDK) inhibitor p15, p16, p18, p19 and blocks CDK4/6, leading to G1 arrest (Drexler, 1998).

- 2) **Proliferation inhibitor genes.** This type of gene secretes hormones or developmental signals that inhibit cell proliferation. Adenomatous polyposis coli (*APC*) encodes *APC* protein providing a scaffold for the destruction complex to unstable β -catenin and inhibiting transcription. Mutations of *APC* lead to tumorigenesis (Smith et al., 2012).
- 3) **Pro-apoptotic genes.** These genes encode proteins that induce apoptosis when DNA damage cannot be repaired. *TP53* specifically activates the extrinsic apoptotic pathway by the tumor necrosis factor (TNF) receptor family or intrinsic apoptosis pathway via the BCL-2 family (Nayak et al., 2009).
- 4) **DNA repair genes.** These genes encode proteins involved in repairing mistakes in DNA, such as DNA mismatch repair protein 2 (*MSH2*) and MutL Homolog 1 (*MLH1*), encoding for proteins that repair mutations in the genome, preventing cells from replication errors (Rahman & Scott, 2007).

1.1.5 Epigenetic alterations of cancer

Epigenetic control provides an alternative way to regulate gene activity without changing the nucleotide sequence. Typically, epigenetic regulation includes DNA and histone modifications, which are essential to maintaining tissue-specific gene expression. However, epigenetic changes also play an essential role in the development of cancer.

1.1.5.1 DNA modification

The most common epigenetic DNA modification is methylation, which is defined as the addition of a methyl group (-CH₃) to cytosine and adenine, two of the DNA's nucleotide bases. CpG dinucleotides or CpG sites are regions where guanosine bases follow cytosine bases. CpG islands (or CG islands) are regions rich in CpG dinucleotides.

DNA methylation is a normal mechanism by which cells regulate gene expression. CpG sites located in promoter regions of the human genome are generally hypomethylated, whereas most CpG sites located outside the promoter region are highly methylated. DNA methylation is regulated by DNA methyltransferases (DNMT) and is important for maintaining genome

stability. Compared with normal somatic cells, the CpG islands in the promoter region of cancer cells are often hypermethylated, while the overall DNA is hypomethylated (Lao & Grady, 2011).

Hypermethylation in CpG islands of tumor suppressor genes may contribute to the inactivation of genes involved in tumorigenesis (Feinberg & Tycko, 2004). One major mechanism by which DNA methylation regulates gene expression is through synergistic interactions with enzymes that regulate chromatin structure, which can induce a compressed chromatin environment that inhibits gene expression. Methyl-binding proteins are important proteins in this mechanism. They constitute a family of proteins that bind to methylated DNA with high affinity, leading to the recruitment of proteins that regulate histone acetylation, histone methylation, and chromatin remodeling to alter chromatin structure, thereby condensing chromatin and blocking transcription factors from entering promoter regions. Hypermethylation of the promoter region of tumor suppressor genes, such as *CDKN2A*, *MLH1*, *CDH1*, and *VHL*, contribute to the cancer formation process (Esteller, 2007).

1.1.5.2 Histone modification

Histone modification is more complex than DNA modification as it involves histone methylation and acetylation, among others (Chen et al., 2020; Zhao & Shilatifard, 2019). The basic structural unit of eukaryotic chromatin is the nucleosome. Each nucleosome comprises DNA and four histones: H2A, H2B, H3, and H4. Gaining or losing functional groups onto histones impacts gene expression, nucleosome localization, chromatin structure and/or DNA replication.

Histone methylation modification involves adding or removing methyl groups from histones and plays an important role in transcriptional regulation by affecting chromatin structure, recruitment of transcription factors, interaction with initiation and elongation factors, and RNA processing (Greer & Shi, 2012). Histone methylation is regulated by histone methyltransferases. SET Domain Containing 1A (SETD1A), a histone methyltransferase, is highly expressed in breast cancer patients and promotes the genesis and metastasis of breast cancer by positively regulating cancer cell proliferation, invasion, and migration (Salz et al., 2015).

Histone acetylation modifications involve adding or removing acetyl groups from histones. Histone acetylation is involved in cell cycle regulation, cell proliferation and apoptosis. Histone deacetylases are enzymes that remove the acetyl groups from an ϵ -N-acetyl-lysine amino acid of a histone, resulting in more condensed nucleosomes and reduced gene transcription (Milazzo et al., 2020). In prostate cancer, after removing acetyl groups from histones, overexpression of histone deacetylases 1 (*HDAC1*) reforms chromatin structure and reduces tumor suppressor gene cyclin dependent kinase inhibitor 1A (*CDKN1A*) expression. *CDKN1A* promotes cell cycle arrest, and the lack of *CDKN1A* leads to prostate cancer progression (Halkidou et al., 2004).

1.2 Anatomy and histology of the intestinal tract

The digestive system is responsible for processing and digesting food as well as absorbing nutrients. It comprises the digestive tract, liver, pancreas, and gallbladder. In turn, the digestive tract comprises the mouth, esophagus, stomach, small intestine, large intestine and anus. The small intestine comprises three parts: duodenum, jejunum, and ileum from proximal to distal. The large intestine includes the cecum, appendix, colon, and rectum. This thesis is focused on colorectal cancer, which arises in the large intestine. However, the large and small intestine share many similarities at the anatomical and histological levels, which are described in detail.

1.2.1 The small intestine

The duodenum is the small intestine's shortest part (20-25cm), which begins at the pyloric sphincter. The bile from the gallbladder and the pancreatic juices are secreted into the duodenum to neutralize the acids from the stomach and initiate the enzymatic digestion of the food. The jejunum is about 2.5 meters long and connects the duodenum and the ileum. The ileum is the longest part of the small intestine, about 3 meters in length (**Figure 2**). The primary function of the small intestine is to help further digest food, absorb water and nutrients (vitamins, minerals, carbohydrates, lipids, and proteins) and release them into the bloodstream.

It also plays an immunity role, acting as a barrier to many microorganisms in the gut lumen, thus ensuring no harmful bacteria and viruses enter the blood.

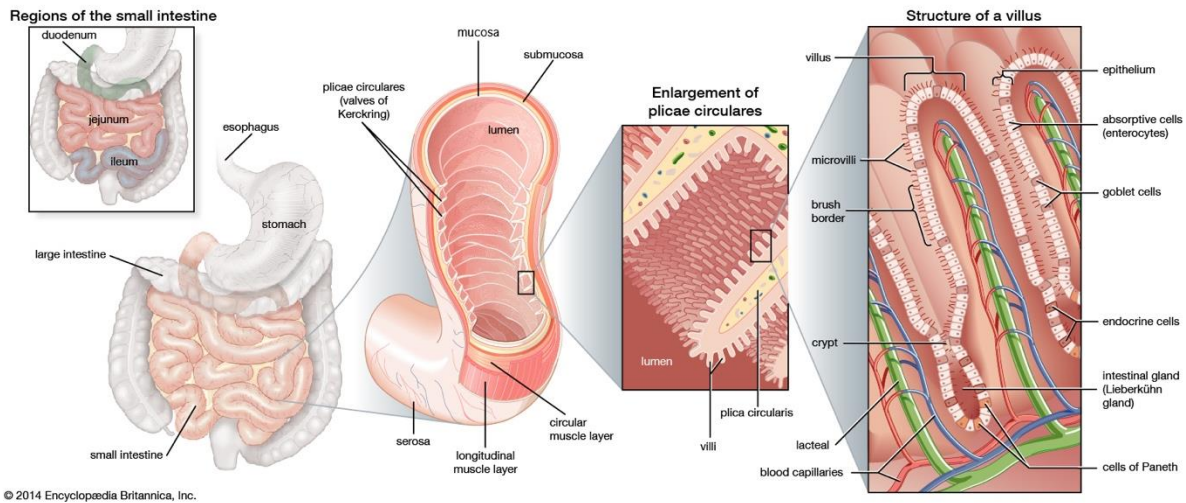


Figure 2: Schematic diagram of the histology of the small intestine.

The small intestine's great length and surface area facilitate the effective digestion and absorption of nutrients. The plicae circulares (circular folds), villi, and microvilli create a large internal surface area (Letricia Dixon, 2014; Encyclopaedia, 2022).

There are four distinct functional layers of the small intestine: mucosa, submucosa, muscularis propria and serosa.

- 1) **Mucosa:** The mucosa is the innermost layer of the small intestine, which consists of the epithelial monolayer, lamina propria and muscularis mucosa. The epithelium has two fundamental structures: villi and crypts (intestinal glands or Lieberkühn glands). The mucous membrane lining the intestinal wall of the small intestine is highly folded, and villi grow into the cavity. These structures significantly contribute to increasing the contact surface with the intestinal contents to enhance secretion and absorption. The mucosa also contains epithelial invaginations termed crypts in which multipotent cells reside, renewing and proliferating continuously. Epithelial cells migrate up the crypt-villus axis and differentiate into several cell types, including enterocytes, goblet cells, Paneth cells, and enteroendocrine cells (Clevers, 2013). Every 4-5 days, most of these cells will be substituted by new cells before sloughing into the lumen. The lamina propria is a loose connective tissue. It carries blood supply to the epithelium and supports the delicate mucosal epithelium. The lamina propria also contains many infiltrating lymphocytes and lymphoid nodules to form an intestinal immune barrier. The outer layer is the muscularis mucosae, a thin muscle layer that maintains the constant exchange of substances between the mucosal surface and the underlying

glands. The muscularis mucosae maintain the contact between the epithelium and the lumen contents and accelerate the elimination of the intestinal content from the glandular crypts.

- 2) **Submucosa:** The submucosa consists of a layer of dense connective tissue containing blood and lymphatic vessels, and nerves to support the mucosa and transfer substances, including electrolytes, vitamins and minerals.
- 3) **Muscularis:** The muscularis comprises two smooth muscle layers. The thinner outer longitudinal layer shortens and elongates the gut, while the thicker inner circular layer of smooth muscle causes constriction. Nerves and muscles between these two layers allow these muscle layers to work together to propagate food from the proximal to the distal direction.
- 4) **Serosa:** The serosa (also termed adventitia) is the outside supporting layer of the small intestine, which secretes fluids to allow sliding movements between distinct surfaces (Stacey E. Mills, 2007).

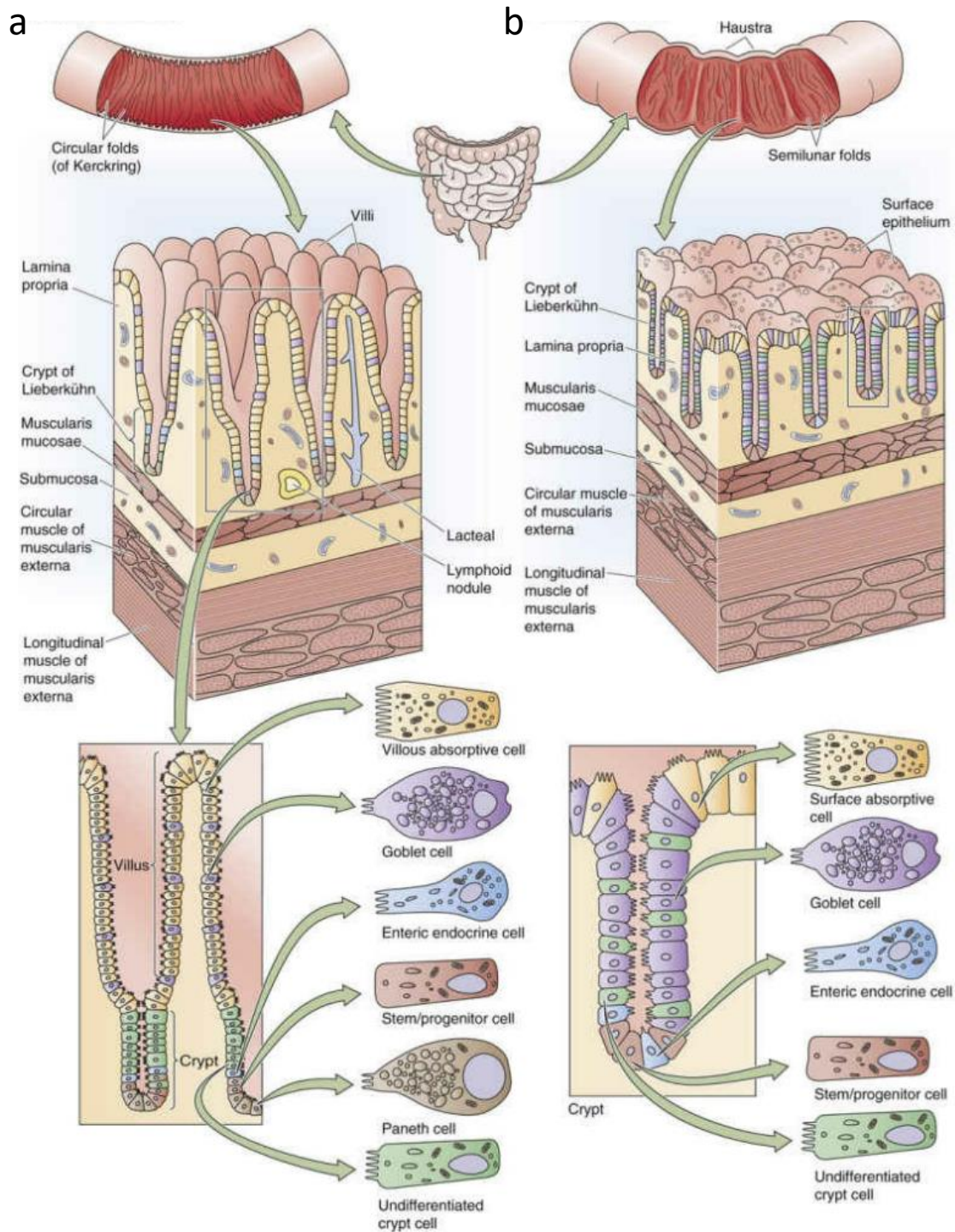


Figure 3: The structure of the small and large intestine and the major cell types of the epithelium.

a) The surface area of the small intestine is amplified at three levels: (1) macroscopic circular folds, also known as circular folds of Kerckring, (2) microscopic villi, crypts, and small intestine histology structure, (3) Small intestine six polarized cell types: villous absorptive cell (enterocytes), goblet cell, enteric endocrine cell, stem/progenitor cell, Paneth cell, and undifferentiated crypt cell. **b)** The surface area of the colon is amplified at the same three levels as the small intestine: (1) macroscopic semilunar folds, (2) crypts (but not villi), and (3) five polarized cell types: absorptive cell, goblet cell, enteric endocrine cell, stem/progenitor cell, Paneth cell, and undifferentiated crypt cell (Boron & Boulpaep, 2016).

As mentioned above, the inner surface of the small intestinal wall is made of simple columnar epithelium. Several distinct cell types are present in the intestinal epithelium, including villous absorptive cells (enterocytes), goblet cells, enteric endocrine cells, stem/progenitor cells, Paneth cells, and undifferentiated crypt cells (Bruno Sarmiento, 2015) (**Figure 3**).

Villous absorptive cells or enterocytes are the most abundant cells within the epithelium and are responsible for absorbing nutrients from the intestinal lumen. These cells display microscopic cellular membrane protrusions toward the lumen, known as microvilli, to increase the contact surface with the intestinal contents significantly. The enterocytes also express many catabolic enzymes on the luminal side to break down molecules to sizes suitable for uptake into the cell, such as ions, water, carbohydrates, peptides and lipids.

Goblet cells account for about 10% of cells in the epithelium. They secrete the mucus that lubricates the passage of food through the intestinal tract and protects the intestinal wall from physical and chemical injuries.

Enteric endocrine cells represent about 1% of all epithelial cells. These cells secrete various gastrointestinal hormones (*e.g.*, secretin, pancreaticozymins, and enteroglucagon) to regulate secretions into the intestinal lumen, influenced by the chime, which is the semi-fluid mass of partly digested food.

Stem/progenitor cells and undifferentiated crypt cells, located towards the bottom of the crypts, are pluripotent cells and have the ability to undergo asymmetric cell division for self-renewal and to give rise to daughter cells committing to lineage-specific differentiation to give rise to all cells in the mature epithelium. Cell differentiation into absorptive cells, goblet cells, and enteric endocrine cells occurs during the process of cell migration along the crypt-villus axis.

Paneth cells are highly specialized secretory epithelial cells in the small intestine. They produce multiple antimicrobial peptides and immunomodulating proteins, such as antibacterial α -defensins, which have an immune function by modulating the gut microbiota composition (Bevins & Salzman, 2011). Paneth cells also mediate stem cell renewal and regeneration by secreting essential niche signals, including EGF, TGF- α , Wnt3 and the Notch ligand Dll4 (Sato et al., 2011). Unlike other cells in the small intestine, Paneth cells move toward the base of the

crypts after differentiation and are renewed every 18-23 days (Porter et al., 2002); Clevers, 2013.

1.2.2 The large intestine

The colon forms the major part of the large intestine (or large bowel) and connects the ileum and the rectum. The colon is much wider but shorter than the small intestine. It is about 1.5 meters long and consists of four parts: ascending, transverse, descending and sigmoid colon (**Figure 4**). The function of the large intestine is to absorb water from the remaining indigestible food residue and store the feces before defecation.

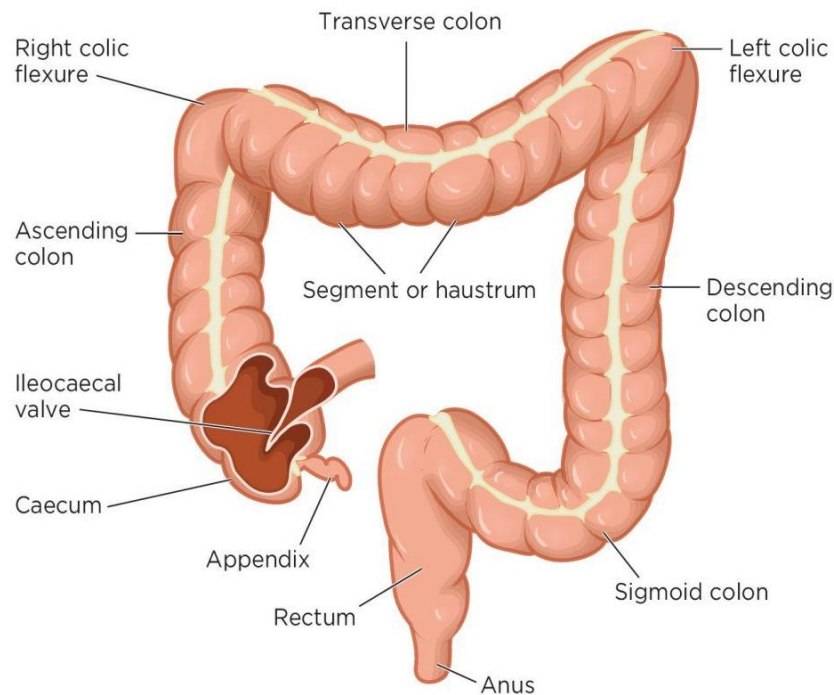


Figure 4: Anatomy of the large intestine.

The main components include ascending, transverse, descending and sigmoid colon (<https://healthknownow.com/gastrointestinal-tract-anatomy-and-functions-of-the-large-intestine/>).

The histology of the large intestine is similar to the small intestine. It is also divided into four layers: mucosa, submucosa, muscularis and serosa. It differs from the small intestine in that villi are absent; instead, the large intestine has a flat absorptive mucosa and invaginations (also called colonic crypts or intestinal glands), which are deeper than the crypts in the small intestine. The epithelium of the large intestine is also made up of a simple columnar epithelium

containing absorptive cells, goblet cells, which are more abundant than in the small intestine, endocrine cells, and basal stem cells, but no Paneth cells (**Figure 3b**). The lamina propria and submucosa display the same properties explained for the small intestine. The longitudinal smooth muscle in the muscularis externa is reduced to three separate bands known as taenia coli. Serosa is the thin exterior layer of connective tissue.

1.3 Colorectal cancer

1.3.1 Epidemiology of colorectal cancer

Colorectal cancer (CRC) has a high incidence and mortality rate worldwide, especially in high-income countries. It is the third most diagnosed and the second cause of death among all types of cancer. More than 1.9 million large bowel cancer patients were diagnosed, and approximately 935,000 people died of colorectal cancer in 2020 worldwide (Sung et al., 2021). Moreover, colorectal cancer has shown rapid growth in incidence and mortality, especially in developing countries, due to the “westernization” of lifestyle. Although the global colorectal cancer burden is still increasing, the incidence is decreasing in some countries where early detection of pre-neoplastic lesions (*e.g.*, endoscopic) is commonly carried out (Arnold et al., 2020)(**Figure 5**).

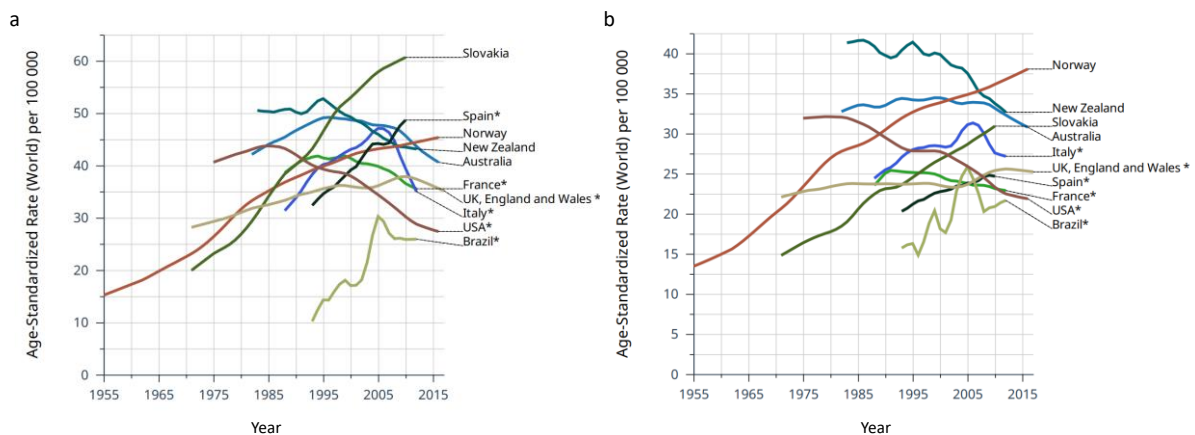


Figure 5: The trend of colorectal cancer incidence rate across countries.

In high-income countries like the USA, France and New Zealand, the incidence rate of colorectal cancer is decreasing, while in most countries around the world, the incidence is still increasing. **a)** Colorectal cancer incidence in men. **b)** Colorectal cancer incidence in women (Sung et al., 2019).

Colorectal cancer is highly associated with age. Both colorectal cancer incidence and mortality increase after the age of 40 and increase sharply at the age of 50. However, an increasing rate of early-onset colorectal cancer (age at diagnosis is < 50 years old) has been found in some high-income countries, including the United States, Canada, and Australia, with the incidence rising by 1-4% per year (Siegel et al., 2019). Approximately 90% of global cases occur in people aged 50 or older (Hagggar & Boushey, 2009). Thus, in many countries, people aged 50-75 are recommended to be regularly screened, including colonoscopy, stool-based tests, or imaging techniques (Shaukat et al., 2021).

Concerning sex-associated risk factors, age-adjusted incidence and mortality rates of colorectal cancer are higher in men than in women, with an approximate 1.4-fold incidence difference (23.4 versus 16.2 cases per 100,000 persons per year) and 1.5-fold for death (11.0 versus 7.2 deaths per 100,000 persons per year) (Sung et al., 2021).

In addition, colorectal cancer is associated with several environmental and lifestyle risk factors, such as high intake of fat, sugar, alcohol, red meat, processed meats, obesity, smoking, and a lack of physical exercise. Inflammatory bowel disease (IBD), including ulcerative colitis and Crohn's disease, also contributes to colorectal cancer (Hagggar & Boushey, 2009).

Most colorectal cancer cases are caused by sporadic mutations (60-65%) or hereditary components (35-40%) influenced by environmental factors. There are three patterns of colorectal cancer, which reflect different risk factors: sporadic, hereditary, and familial.

- 1) **Sporadic colorectal cancer.** An estimated 60-65% of colorectal cancer occurs in people acquiring somatic genomic alterations in cells from the bowel, which means that the alterations develop individually without inherited defects (**Figure 6**).
- 2) **Inherited colorectal cancer.** Approximately 5-10% of colorectal cancer patients have known hereditary conditions, including hereditary cancer syndromes, known colorectal cancer low-penetrance genetic variations and other unknown inherited aberrations (**Figure 6**). 5% of colorectal cancer cases are associated with **hereditary cancer syndromes**, including familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC), also known as Lynch syndrome (Keum & Giovannucci, 2019), caused by inherited germline mutations in rare but high-penetrance susceptibility genes (for example, *APC* and *MLH1*, respectively). The rest

of the inherited genetic risks are known as **low-penetrance genetic variations** (rarely express an associated symptom) but predisposing to colorectal cancer and other **unknown inherited genomic aberrations** (Jiao et al., 2014).

3) **Family history.** 25% of colorectal cancer patients have a family history of colorectal cancer without obvious hereditary cancer syndromes, also known as familial colorectal cancer. Individuals in these families have an increased risk of colorectal cancer, but the risk is not as high as in patients with hereditary cancer syndromes. Familial colorectal cancer arises from a number of different, lower to intermediate-penetrance susceptibility genes. Familial colorectal cancer is also different from known colorectal cancer with low-penetrance genetic variations, as familial colorectal cancer susceptibility genes are rare. In contrast, the susceptibility genes of known colorectal cancer low-penetrance genetic variations are common, so they can be identified by genome-wide association studies (Jasperson et al., 2010).

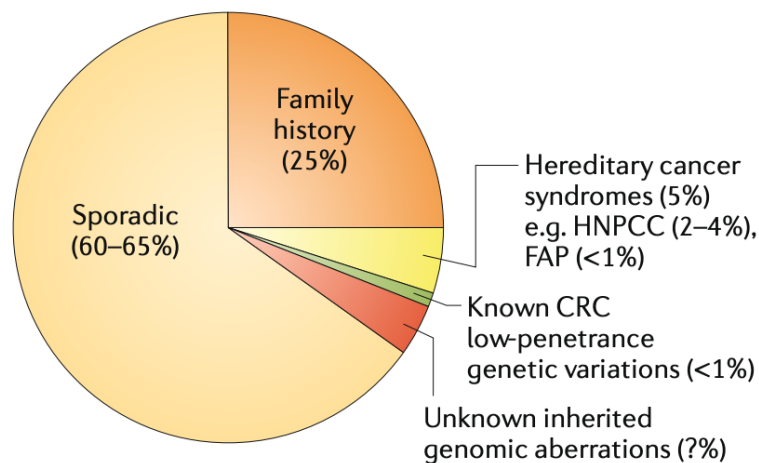


Figure 6: Proportion of colorectal cancer cases associated with sporadic and hereditary factors.

60-65% of colorectal cancer patients are associated with somatic genomic alterations. 25% of the cases have a family history of colorectal cancer without any obvious genetic cancer syndrome. 5% of colorectal cancer patients have hereditary cancer syndromes (hereditary nonpolyposis colorectal cancer and familial adenomatous polyposis). Less than 1% of colorectal cancer patients have known common but low-penetrance genetic variations. The rest are other inherited aberrations yet to be discovered (Keum & Giovannucci, 2019).

1.3.2 Histologic and anatomical classification of colorectal cancer

There are several ways to classify colorectal cancer. In terms of histology, colorectal cancer can be divided into adenocarcinoma (97.4%), carcinoid (1.5%), lymphoma (0.6%), non-carcinoid neuroendocrine (0.3%), squamous cell (0.3%), and sarcoma (<0.1%) (Kang et al., 2007). **Adenocarcinomas** originate from the epithelial cells of the intestine, including mucinous adenocarcinoma, signet-ring cell carcinoma, and medullary carcinoma. **Carcinoid tumors** are a type of neuroendocrine carcinoma that arise from enterochromaffin cells, also known as the silvery Kulchitsky cells, of the intestinal mucosal glands. These tumors originate from cells of endodermic origin and are arranged in nests, and although they have low rates of metastatic spread, carcinoid tumors have been classified as malignant tumors (Eggenberger, 2011). The most common histological subtype of colorectal **lymphoma** is diffuse large B-cell lymphoma, which consists of rapidly proliferating cells of B-cell origin (Quayle & Lowney, 2006). Like carcinoid tumors, **non-carcinoid neuroendocrine tumors** originate from neuroendocrine cells, but these cells are high-grade and more aggressive (Bernick et al., 2004). **Squamous cell carcinoma** colorectal squamous cell carcinoma is a poorly differentiated cancer with squamous features resulting from squamous metaplasia caused by external stimuli such as viral infections.

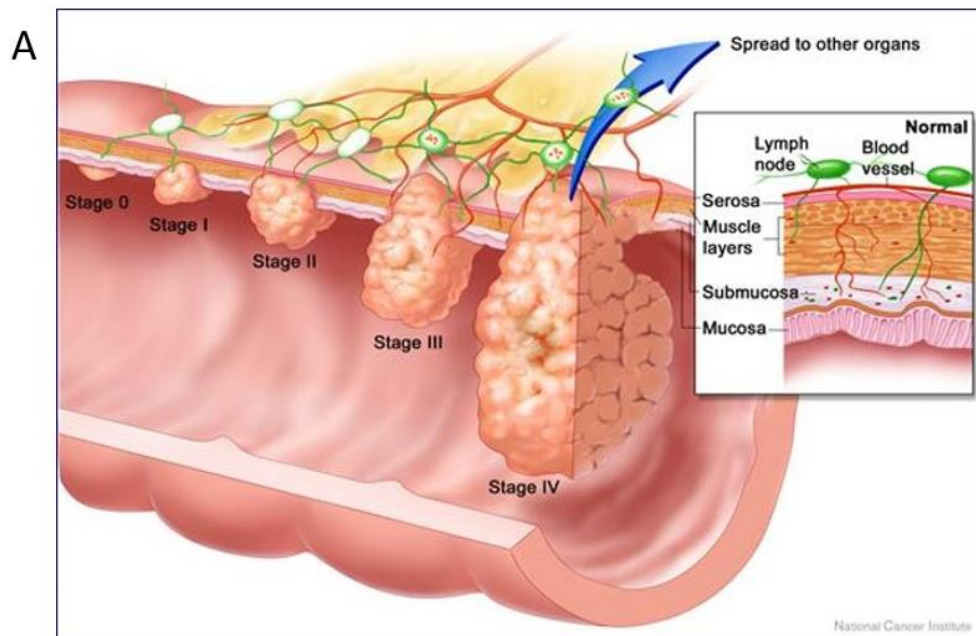
Tumors of different stages have significant prognostic differences (**Figure 7B**). The 5-year survival rate is often used to assess the prognosis of patients because more than 90% of tumor recurrences occur within five years of receiving treatment. If no tumor is detected by screening within five years, it is considered that the cancer has been cured.

The most widely used classification to estimate the prognosis of colorectal cancer patients is based on their staging. The first method was proposed by C.E. Dukes in 1932 and is known as the Dukes' staging system. After several revisions, the Dukes' classification is as follows: Dukes' A cancer is limited in the inner lining of the bowel or slightly affects the muscle layer. Dukes' B colorectal cancer has invaded through the muscle layer of the bowel. Dukes' C cancer grows deeper and has spread to at least one lymph node close to the bowel. Dukes' D cancer has spread to other organs, such as the liver, lungs, or bones Ciccolallo et al., 2005). The Dukes' staging system is now mainly of historical interest as it has largely been substituted by the TNM staging system.

In the TNM staging system, “T” stands for the local extent of the primary tumor at the time of diagnosis, “N” refers to the spread to nearby lymph nodes, while “M” indicates distant metastatic disease (metastasis); **Figure 7A**).

The TNM staging system is vital for guiding treatment and an essential indicator for estimating the prognosis. As the TNM staging may change during the treatment, the staging is always based on the stage at diagnosis.

In addition, the tumor grading system indicates how much cancer cells resemble healthy colon tissue under a microscope. In this classification system, tumors are graded as 1 (well differentiated, low grade), 2 (moderately differentiated; intermediate grade); 3 (poorly differentiated; high grade); and 4 (undifferentiated, high grade) (Compton, 2003).



B

STAGE	T	N	M	DUKES	5-YEAR RELATIVE SURVIVAL RATE (AGE 20 - 64)	5-YEAR RELATIVE SURVIVAL RATE (AGE > 64)
0	Tis	N0	M0	/	/	/
I	T1	N0	M0	A	95.2% (n = 26341)	89.1% (n = 33848)
	T2	N0	M0	A		
IIA	T3	N0	M0	B	89.6% (n = 21680)	84.4% (n = 33564)
IIB	T4a	N0	M0	B	67.6% (n = 4169)	55% (n = 5972)
IIC	T4b	N0	M0	B	/	/
IIIA	T1–T2	N1/N1c	M0	C	91.3% (n = 4433)	85.1 (n = 3872)
	T1	N2a	M0	C		
IIIB	T3–T4a	N1/N1c	M0	C	76.9% (n = 17945)	64.6% (n = 20198)
	T2–T3	N2a	M0	C		
	T1–T2	N2b	M0	C		
IIIC	T4a	N2a	M0	C	61.8% (n = 11853)	45.5% (n = 11376)
	T3–T4a	N2b	M0	C		
	T4b	N1–N2	M0	C		
IVA	AnyT	AnyN	M1a	D	14.2% (n = 30669)	7.4% (n = 30246)
IVB	AnyT	AnyN	M1b	D		

Figure 7: Staging of colorectal cancer.

A) Graphical representation of the different stages of colorectal cancer. (NCI, 2015). **B)** TNM classification. T - primary tumor, Tis - Carcinoma *in situ*, T1 - Tumor invades submucosa, T2 - Tumor invades muscularis propria, T3 - Tumor invades subserosa or pericolic/perirectal tissue, T4a - Tumor grows into the surface of the visceral peritoneum, T4b - Tumor invades other organs or structures. N - No spread to regional lymph nodes, N1 - Metastasis in 1 (N1a), 2 (N1b), or 3 (N1c) regional lymph nodes, N2a - Metastasis in 4 - 6 regional lymph nodes, N2b - Metastasis in 7 or more regional lymph nodes, M0 - No spread to a distant part of the body, M1a - Metastasis spreads to one organ, M1b - Metastasis spreads to 2 or one organs. Data were taken from (Edge et al., 2010) and (Crooke et al., 2018).

1.3.3 Tumorigenesis and deregulation of signaling pathways in colorectal cancer

Tumorigenesis is a multiple-step process influenced by environmental and genetic factors. In this process, normal cells acquire multiple hallmarks of cancer (**Figure 8**) to form malignant tumors (Ogino et al., 2011).

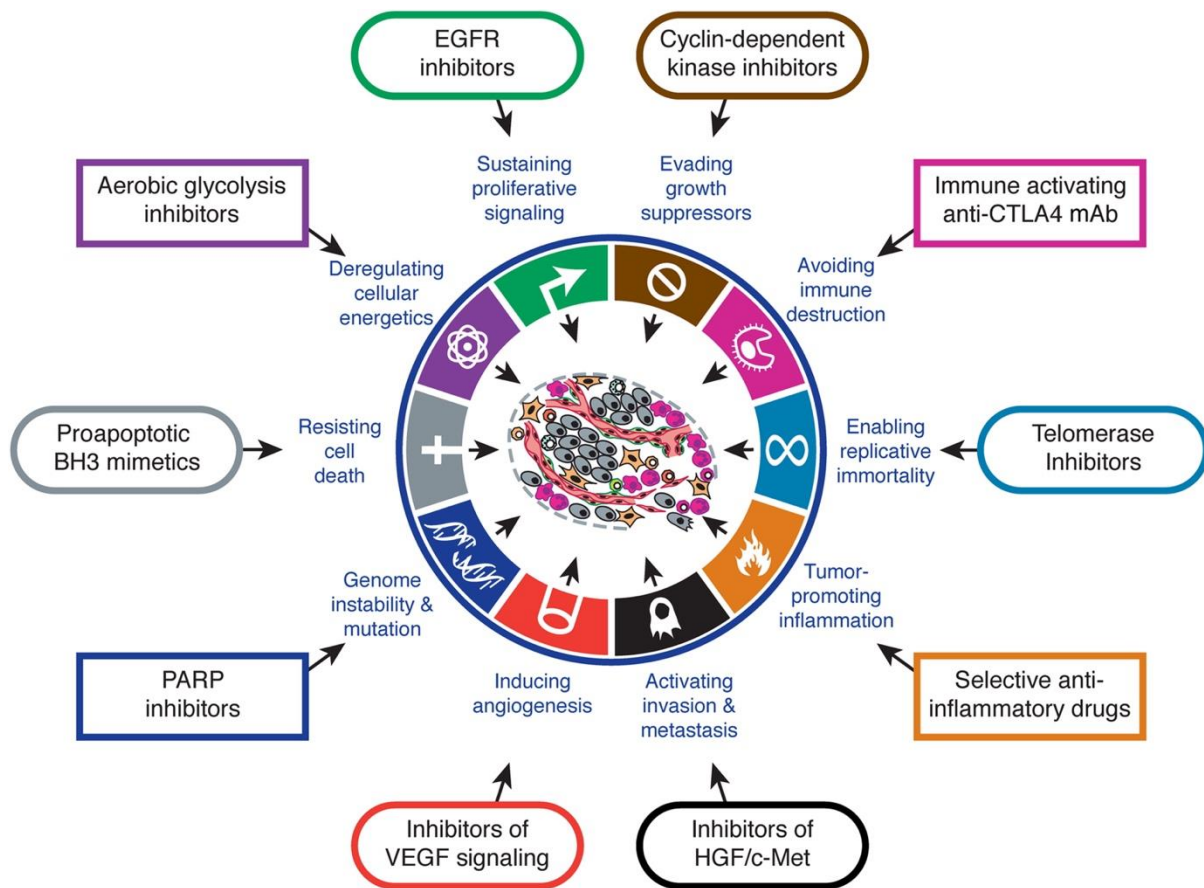


Figure 8: The hallmarks of cancer and therapeutic targets.

The illustration of the ten hallmark capabilities of cancer proposed in 2000 and 2011. Drugs that interfere with every hallmark for tumor growth and progression have been developed and are in clinical trials. EGFR, Epidermal growth factor receptor; CTLA4, Cytotoxic T-Lymphocyte Associated Protein 4; HGF, Hepatocyte growth factor; VEGF, Vascular endothelial growth factor; PARP, Poly (ADP-ribose) polymerase. (Hanahan & Weinberg, 2011).

In the development of colorectal cancer, abnormal cell signaling exists at all stages of tumorigenesis. In most cases, a disturbance occurs by a mutation in key genes within the pathway. Genetic mutations and epigenetic alterations may activate oncogenes and inactivate tumor suppressor genes. These alterations may cause genomic or epigenomic instability, which will, in turn, generate more mutations or epigenetic alterations (Kuipers et al., 2015). In the development of colorectal cancer, although mutations could occur on thousands of genes, only those mutations that can confer a selective growth advantage to the cells drive the malignant transformation of colon cells. These genes are called cancer driver genes (Fearon & Vogelstein, 1990) (**Figure 9**).

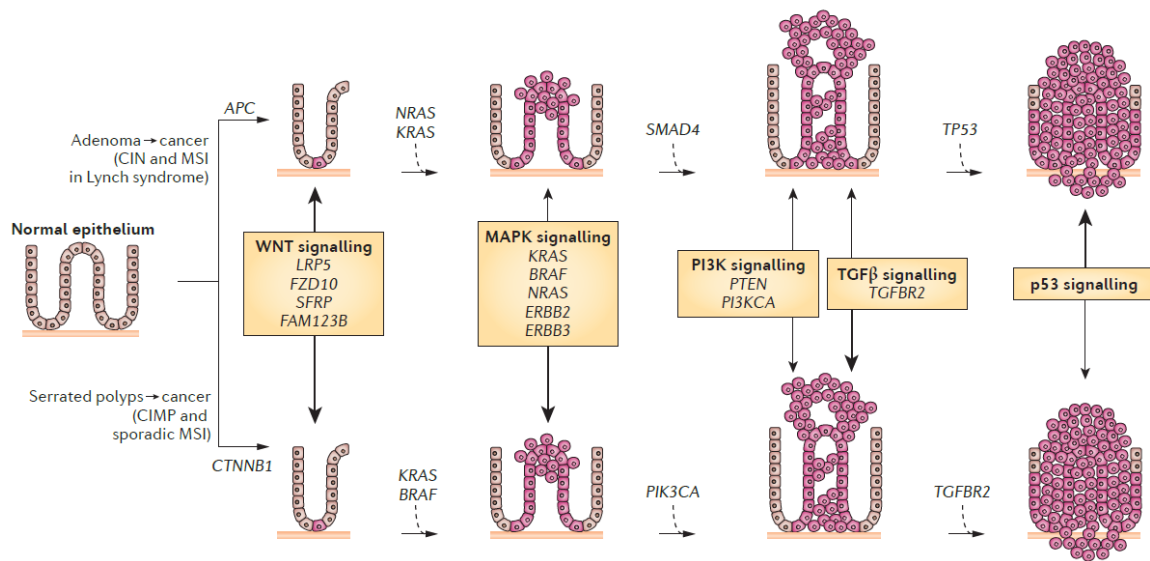


Figure 9: Schematic diagram of colorectal cancer development through two approaches.

The progression of colorectal tumors follows the: **1) adenoma-cancer pathway**, which is driven by inactivating mutations in *APC*, *NRAS/KRAS*, *SMAD4* and *TP53*; **2) serrated polyps-cancer pathway**, which is triggered by inactivating mutations in *CTNNB1*, *KRAS/BRAF*, *PIK3CA* and *TGFBR2*. CIN, chromosomal instability; MSI, microsatellite instability; CIMP, CpG island methylator phenotype; *APC*, adenomatous polyposis coli; *CTNNB1*, catenin- β 1; *FAM123B*, family with sequence similarity 123B (also known as *AMER1*); *FZD10*, frizzled class receptor 10; *LRP5*, low-density lipoprotein receptor-related protein 5; MAPK, mitogen-activated protein kinase; MSI, microsatellite instability; PI3K, phosphatidylinositol 3-kinase; *PI3KCA*, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit- α ; *PTEN*, phosphatase and tensin homolog; *SFRP*, secreted frizzled-related protein; *SMAD4*, SMAD family member 4; TGF β , transforming growth factor- β ; *TGFBR2*, TGF β receptor 2. Figure adapted from (Kuipers et al., 2015).

Specifically, five main pathways play key roles in colorectal cancer development: the WNT signaling pathway, RTK-RAS signaling pathway, PI3K signaling pathway, TGF- β signaling pathway and P53 signaling pathway (**Figure 10**). These pathways share some genes with each other and thus generate interaction between pathways. Studying the pathways in depth can help us understand the mechanism of carcinogenesis, thereby allowing the development of new treatments or prevention methods for colorectal cancer (Muzny et al., 2012).

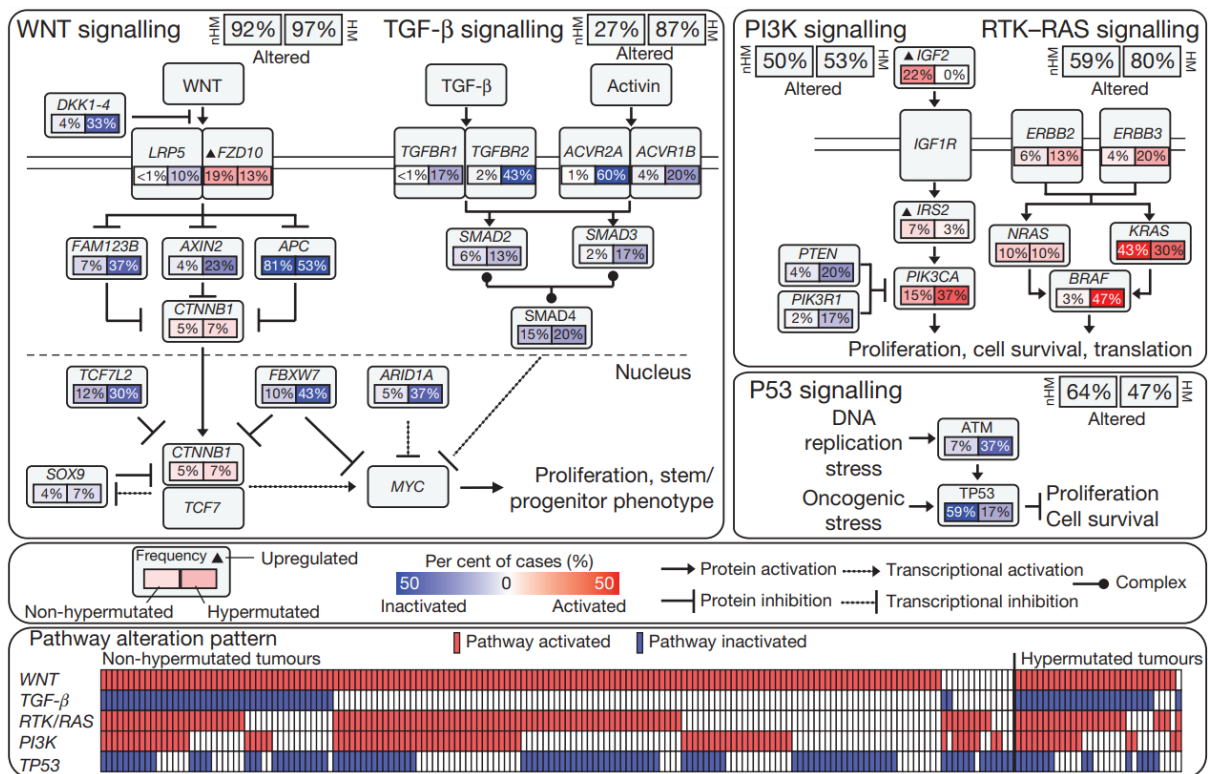


Figure 10: Signaling pathways alterations in colorectal cancer.

This figure shows non-hypermuted (nHM; n = 165) and hypermuted (HM; n = 30) samples, respectively. All alteration data are based on somatic mutations, homozygous deletions, high levels of local amplification, and in some cases, significant up- or down-regulation of gene expression. The frequency of changes is expressed as a percentage of all cases. Red indicates activated genes, while blue denotes inactivated genes. The lower panel shows that each sample has at least one gene altered in all five pathways (Muzny et al., 2012).

1.3.3.1 WNT signaling pathway

The WNT signaling pathway is a highly evolutionary conserved pathway found in almost all animal species. It plays an essential role in several cellular functions, such as proliferation, migration and polarity. The abnormal activation of the WNT signaling pathway is related to the pathogenesis of multiple types of tumors, including colorectal cancer, leukemia, melanoma and breast cancer.

Two WNT signaling pathways have been identified: the WNT-β-catenin pathway, also called the canonical WNT pathway, and the non-canonical WNT pathway, including the planar cell polarity (WNT/PCP) and calcium signaling (WNT/Ca²⁺) pathways (Huelsenken & Behrens, 2002; Gordon & Nusse, 2006) (**Figure 11**). WNT protein is a secreted cysteine-rich glycoprotein that functions through autocrine or paracrine action.

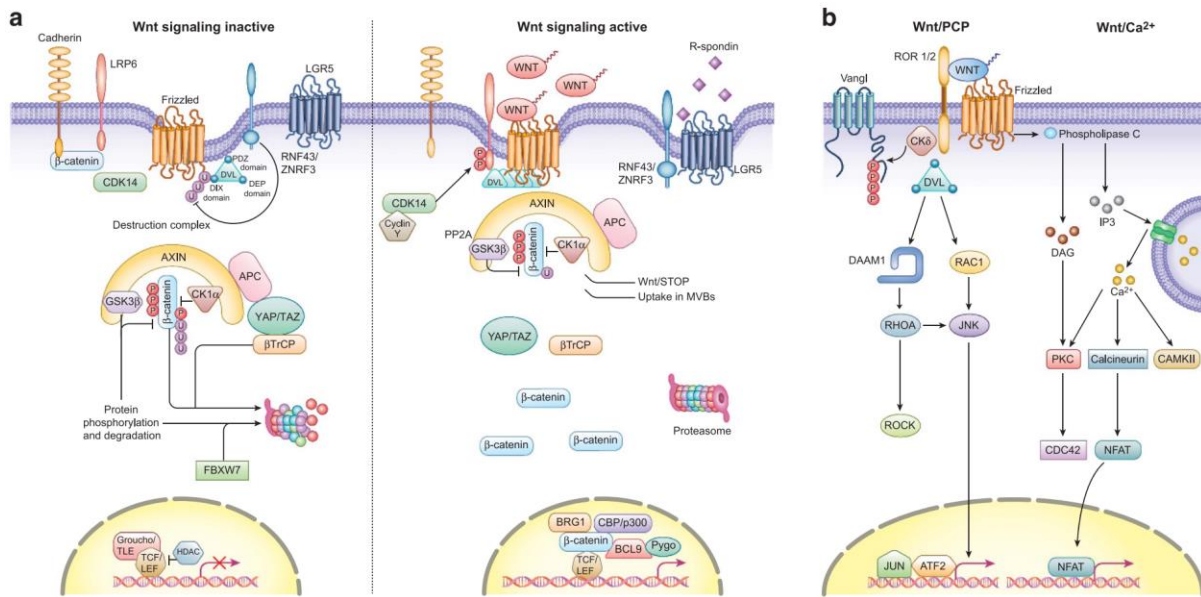


Figure 11: Overview of canonical and non-canonical WNT signaling.

a). Canonical WNT signaling. **b).** Non-canonical WNT signaling. LRP6, Low-Density Lipoprotein Receptor-Related Protein 6; LGR5, Leucine Rich Repeat Containing G Protein-Coupled Receptor 5; CDK14, Cyclin Dependent Kinase 14; RNF43, Ring Finger Protein 43; ZNRF3, Zinc And Ring Finger 3; AXIN, Axis Inhibition Protein; APC, Adenomatous Polyposis Coli Protein, GSK3 β , Glycogen Synthase Kinase 3 Beta; YAP, Yes Associated Protein; TAZ, Phospholipid-Lysophospholipid Transacylase; β TrCP, Beta-Transducin Repeat Containing E3 Ubiquitin Protein Ligase; FBXW7, F-Box And WD Repeat Domain Containing 7; TLE, Transcriptional Corepressor; TCF, Transcription Factor; LEF, Lymphoid Enhancer Binding Factor; HDAC, Histone Deacetylase; BRG1, Actin Dependent Regulator Of Chromatin; CBP/p300, Cyclic adenosine monophosphate Response Element Binding protein Binding Protein (CREB-binding protein); BCL9, B-cell CLL/lymphoma 9 protein; ROR, Receptor Tyrosine Kinase Like Orphan Receptor; DVL, Dishevelled Segment Polarity Protein; DAAM1, Dishevelled Associated Activator Of Morphogenesis; RAC1, Rac Family Small GTPase 1; RHOA, Ras Homolog Family Member A; JNK, Mitogen-Activated Protein Kinase 8; ROCK, Rho Associated Coiled-Coil Containing Protein Kinase; JUN, Jun Proto-Oncogene; ATF2, Activating Transcription Factor 2; DAG, diacylglycerol; IP3, Inositol trisphosphate; PKC, Protein Kinase C; CAMKII, Calcium/Calmodulin Dependent Protein Kinase II, CDC42, Cell Division Cycle 42; NFAT, Activating Protein With ITAM Motif 1 (Zhan et al., 2017).

Intestine stem cells can continuously proliferate and differentiate at the bottom of the crypts. The intestinal epithelial cells migrate upward and stop proliferating when they reach the top of the crypt under the control of the WNT and BMP signaling, an opposing pathway to WNT signaling to keep the epithelial homeostasis. However, epithelial cells proliferate aberrantly in colorectal cancer patients due to constant over-activation of the WNT- β -catenin pathway, mostly due to gene mutations.

In the **canonical WNT pathway** (WNT/ β -catenin; **Figure 11a**), when cells are not stimulated by WNT ligands, β -catenin is phosphorylated by the β -catenin destruction complex composed of scaffold protein AXIN, APC, kinases GSK3 β and casein kinase (CK1 α). Phosphorylated β -

catenin is ubiquitinated by β -TrCP and subsequently degraded by the proteasome. The absence of β -catenin causes a complex in the nucleus containing TCF/LEF, and transducing-like enhancer protein (Groucho/TLE) recruits Histone Deacetylases (HDACs) to repress target genes. The canonical pathway is activated when WNT proteins like WNT3a and WNT1 stimulate cells, the binding of WNT protein and the low-density lipoprotein receptor-related protein (LRP) family disrupt the destruction complex, preventing the phosphorylation of β -catenin and its degradation by proteases. The accumulated β -catenin in the cytoplasm translocates to the nucleus and binds to the transcription factor TCF/LEF complex. This transcriptional switch leads to the activation of WNT target genes (Schneikert & Behrens, 2007).

The **noncanonical WNT pathway (Figure 11b)** does not involve β -catenin. In **WNT/PCP** (planar cell polarity) signaling, the pathway is activated by WNT ligands binding to the ROR-Frizzled receptor. Then cascade can be activated involving RHOA and Jun-N-terminal kinase (JNK), and Nemo-like kinase (NLK) signaling to induce intracellular cytoskeleton rearrangements and impact cell behaviors (Y. Wang, 2009). **WNT/Ca²⁺ signaling** is activated by G-proteins activating phospholipase C activity and increasing Intracellular calcium levels. This affects calcium dependent-transcription factors like NFAT and leads to a change in multiple cellular processes (De, 2011).

The inactivation of the **adenomatous polyposis coli (APC)** tumor suppressor is one of the pivotal events in malignant WNT signaling activation. The *APC* gene is located on chromosome 5q22. Its product, APC, acts as a key negative regulator in the WNT-signaling pathway. Its mutation is present in about 64% of patients (non-hypermuted: 81%; hypermutated: 53%), according to TCGA data (Cerami et al., 2012; Gao et al., 2013). During embryonic development, this pathway regulates cell fate determination, cell migration, cell polarity, and neural patterning. Mutations in *APC* are considered the initiating event in the multi-step colorectal cancer development process. The mutation can occur in both germline and somatic cells. Germline mutations lead to familial adenomatous polyposis (FAP), the main inherited colorectal cancer predisposition syndrome, while somatic mutations are found in sporadic colorectal cancer (Rowan et al., 2000).

In normal cells, APC provides a scaffold for the destruction complex that stimulates β -catenin phosphorylation and subsequent ubiquitin-dependent degradation of β -catenin (B Rubinfeld et

al., 1996). However, the mutations in *APC* lead to the loss of the interaction site with *AXIN* (Polakis, 2000), driving the instability of the destruction complex and preventing the formation of the complex. Besides, loss of *APC* function also affects cell adhesion (Bienz & Hamada, 2004), cell cycle (Baeg et al., 1995), DNA repair (Neufeld et al., 2000), and contributes to chromosome instability (CIN).

1.3.3.2 RTK-RAS signaling pathway

The RTK-RAS pathway integrates extracellular signals and transmits them to the nucleus to regulate cell growth and proliferation. This signaling pathway has four key factors: RAS, RAF, MEK1/2, and ERK1/2 proteins (**Figure 12**). Any alterations of these factors may cause a tumor. Receptor tyrosine kinases (RTKs), such as epidermal growth factor receptors (EGFR), are located on the cell surface. They are activated by extracellular ligands, such as the epidermal growth factor (EGF), and exhibit protein tyrosine kinase (PTK) activity. EGFR consists of an extracellular receptor containing a ligand-binding site, a transmembrane domain, and an intracellular domain with tyrosine kinase activity. RTKs homo- or hetero-dimerize and transphosphorylation, thus recruiting the growth factor receptor-bound protein 2 (GRB2), leading to its tyrosine phosphorylation and subsequent activation of the son of sevenless (SOS) protein. The activated SOS proteins subsequently promote RAS protein removing GDP and binding to GTP, causing its activation. In addition, RAS activation can further activate RAF. RAF phosphorylates and activates MEK1/2, which in turn phosphorylates and activates MAPK Kinases (also known as ERK1/2), inducing signal transduction to the nucleus, leading to the engagement of factors to induce gene expression (W. Zhang & Liu, 2002).

The RTK-RAS pathway contains known colorectal cancer oncogenes such as Kirsten Rat Sarcoma Viral Oncogene Homologue (*KRAS*) and B-Raf Proto-Oncogene (*BRAF*). These two genes belong to the RAS and RAF family, respectively. The mutations of *KRAS* and *BRAF* are considered early events of colorectal cancer. In addition, the detection of *KRAS* and *BRAF* gene mutations has been used in the clinic to guide subsequent choice to administer targeted therapy with anti-epidermal growth factor receptor antibodies.

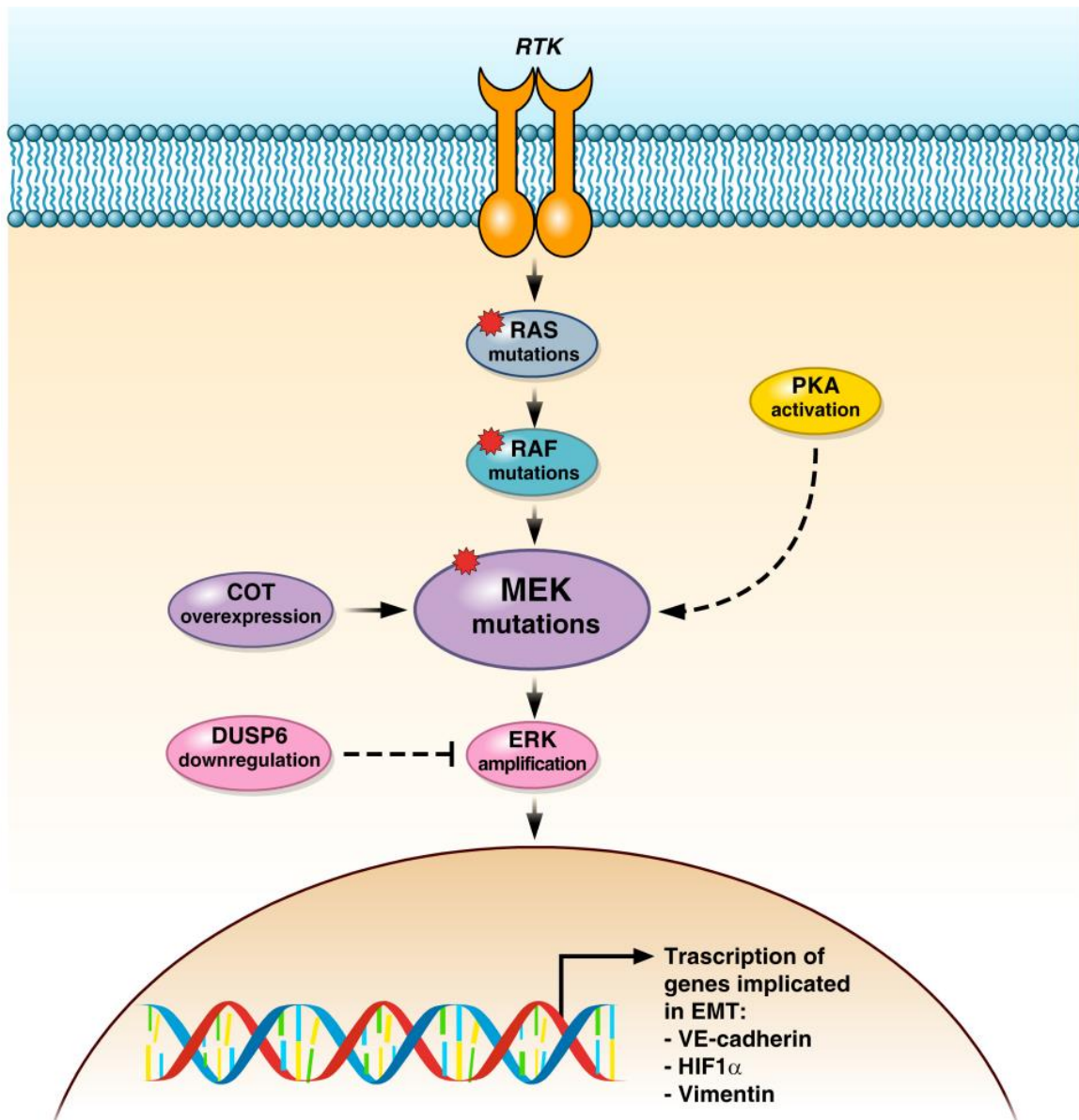


Figure 12: Schematic representation of RTK-RAS signaling pathway in cancer.

RTK, Tyrosine kinase receptor; RAS, Rat sarcoma virus, RAF, Rapidly accelerated fibrosarcoma; MEK, Mitogen-activated protein kinase; ERK, Extracellular signal-regulated kinases; PKA, Protein kinase A. (Martinelli et al., 2017).

This signaling pathway is often activated in colorectal tumors by mutations in *KRAS*, a proto-oncogene, located in chromosome 12p12 that belongs to the RAS superfamily. As mentioned above, the protein encoded by *KRAS* relays extracellular signals to the nucleus via the RTK/RAS signaling. *KRAS* is mutated in approximately 36% of colorectal cancer, among which the majority of mutations are at glycine at position 12 (G12), glycine at position 13 (G13), and glutamine at position 61 (Q61) (Chang et al., 2013). In normal conditions, *KRAS* is activated when binding to GTP (RAS-GTP) and inactivated after converting GTP to GDP

(RAS-GDP), regulated by GTPase activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs). Mutant *KRAS* is impaired in GTP hydrolysis activity, resulting in RAS-GTP accumulation, which activates downstream targets related to tumorigenesis (*e.g.*, PI3K, BRAF) (Bos et al., 2007). *KRAS* mutation can also be found at high frequency in other cancer types like non-small cell lung cancer and pancreatic cancer.

Alternatively, the RTK/RAS pathway can be activated by mutations in *BRAF*, a proto-oncogene that belongs to the RAF family and is located on chromosome 7q34. It plays a vital role in cell proliferation, differentiation, and secretion via RTK/RAS signaling. *BRAF* can be activated by RAS-GTP, thus activating the transcription of multiple genes and promoting cell proliferation. *BRAF* mutations have been found in 5-10% of colorectal cancer. Approximately 96% of *BRAF* mutations are V600E - valine (V) is substituted by glutamic acid (E) at amino acid residue 600 (Cantwell-Dorris et al., 2011). V600E mutation increases BRAF activity approximately 10-fold, phosphorylating and activating downstream targets of the RTK/RAS signaling cascade (Nazemalhosseini Mojarad et al., 2013; Caputo et al., 2019)

1.3.3.3 PI3K signaling pathway

The phosphoinositide 3-kinase (PI3K) pathway regulates cell proliferation, differentiation, apoptosis, and protein transport, among other cell functions. PI3K protein consists of a regulatory and catalytic subunit, which can be activated by various RTK receptors, including IGF1R, EGFR, and HER2. The activated PI3K can convert PIP2 (Phosphatidylinositol (4,5)-bisphosphate) into PIP3 (Phosphatidylinositol (3,4,5)-trisphosphate). PIP3 acts as a second messenger to drive a variety of downstream signaling pathways (**Figure 13**).

The downstream effectors of the PI3K pathway include protein kinase B (PKB), also known as AKT, and mTORC1 (mammalian target of rapamycin), participating in regulating cell proliferation, survival, apoptosis, migration and metabolism. Their abnormal activation can increase the translation of cell cycle regulatory factors, including MYC and CCND1, and cause excessive cell proliferation. AKT can also phosphorylate key target proteins through various downstream pathways to exert anti-apoptotic effects.

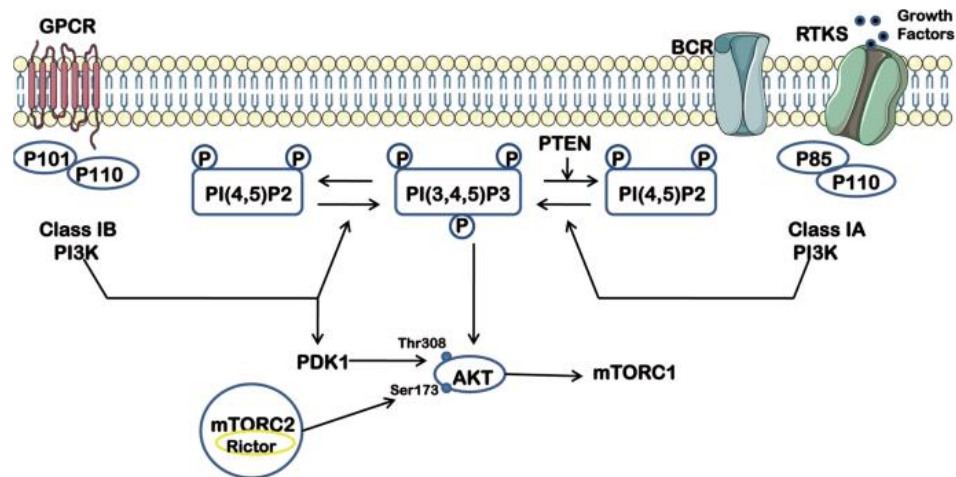


Figure 13: PI3K pathway and its major downstream effectors.

RTK recruits PI3K following activation and phosphorylates PI(4,5)P2 to PI(3,4,5)P3, which activates AKT by recruiting PDK1, thereby activating mTORC1 and regulating cell growth. RTK, receptor tyrosine kinase; PI3K, phosphatidylinositol 3-kinase; AKT, protein kinase B; PI, phosphatidylinositol; PDK1, 3-phosphoinositol-dependent protein kinase-1; PTEN, phosphatase and tensin homolog; mTORC1, mammalian target of rapamycin complex 1 (R. Liu et al., 2020)

PIK3CA is located on chromosome 3q26, and it encodes the protein Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha (also called p110 α), mutated in 32% of colorectal tumors. Besides colorectal cancer, *PIK3CA* mutations are also found in the brain (27%), gastric (25%), breast (8%), and lung (4%) cancers (Karakas et al., 2006). About 80% of *PIK3CA* mutations in colorectal cancer are activating mutations in codons E542K, E545K, and H1047R (Samuels et al., 2004). *PIK3CA* is activated by tyrosine kinase receptors (RTK) and the regulatory subunit (p85 α). Downstream effectors include AKT (protein kinase B), serine-threonine kinase and mTOR (mammalian target of rapamycin). PTEN (Phosphatase and tensin homolog) catalyzes the opposite reaction of *PIK3CA* and thus operates as a tumor suppressor (Engelman, 2009). Similarly, *PTEN* has a relatively high mutation rate in colorectal cancer patients (18%) (Nassif et al., 2004).

1.3.3.4 TGF- β signaling pathway

TGF- β (Transforming growth factor-beta) belongs to the transforming growth factor superfamily, which inhibits cell proliferation, induces apoptosis, activates autophagy, and prevents tumor formation. TGF- β signaling initiates with the TGF- β superfamily ligands binding to the TGF- β type II receptor (TGFBR2/TGF β RII). TGFBR2 recruits, phosphorylates

and activates TGFBR1 (TGFβRI). Then, activated TGFBR1 induces phosphorylation of SMAD2 and SMAD3. SMAD2/3 complex, in turn, binds SMAD4 (Massagué, 2012). The complex then shuttles to the nucleus, activating transcriptional regulators, including cyclin-dependent kinase (CDK) inhibitors *CDKN1A*, *CDKN1B*, and *CDKN2B*, activating cell proliferation (Figure 14) (Wu et al., 2005).

TGF-β inhibits early tumor formation in normal intestinal epithelial cells by inhibiting cell proliferation and inducing apoptosis. Colorectal cancer cells evade apoptosis by accumulating point mutations or gene deletions in *SMAD2*, *SMAD3*, *SMAD4*, *TGFBR2* and *TGF-βR2* (Xu & Pasche, 2007; Akhurst & Hata, 2012). Besides, it has also been found that in the advanced stage of the tumor, TGF-β can promote tumorigenesis by regulating genome instability, epithelial-mesenchymal transition (EMT), angiogenesis, immune escape and metastasis (Mishra et al., 2005).

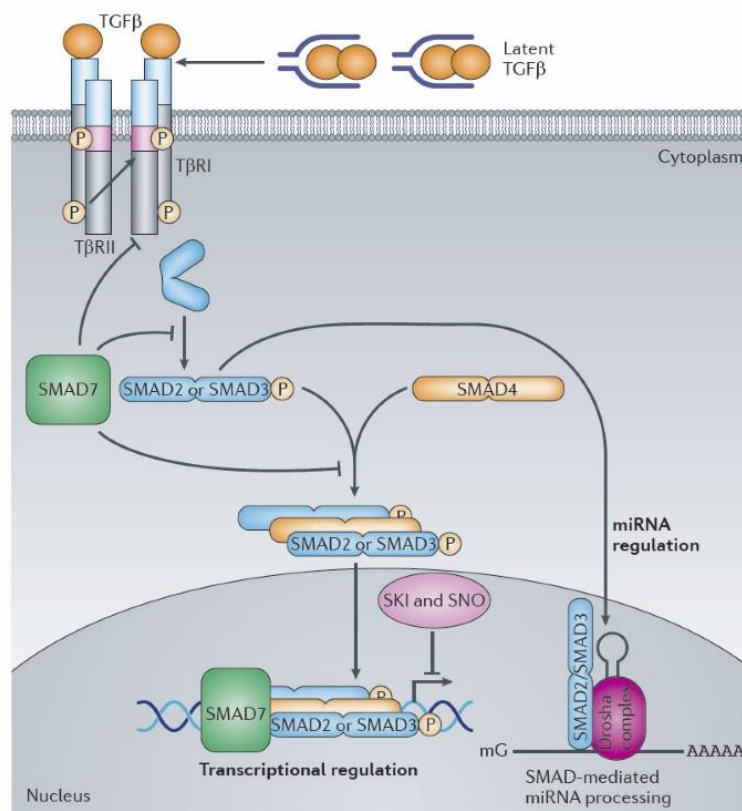


Figure 14: TGF-β signaling pathway.

The TGF-β signaling pathway is activated by the binding of the TGF-β ligand to TGFBR2 (TβRII). Activated TGFBR2 recruits and phosphorylates TGFBR1 (TβRI). Upon phosphorylation by TGFBR1, SMAD2 and SMAD3 form a transcriptional complex with SMAD4. This complex then translocates into the nucleus, activating transcriptional regulators and regulating target gene expression. SKI and SNO function as negative regulators of TGF-β signaling. SMAD, Mothers against decapentaplegic homolog (Akhurst & Hata, 2012).

SMAD4 (Mothers Against Decapentaplegic Homolog 4), located in chromosome 18q21, has been identified as a tumor suppressor gene in the transforming growth factor (TGF)- β signaling pathway. Approximately 10% of sporadic colorectal tumors have *SMAD4* mutations. As mentioned before, SMAD4 binds the SMAD2/3 complex to regulate the transcription of the genes involved (Fleming et al., 2013). It regulates the transcription of genes involved in cell cycle arrest in epithelial cells, wound healing, extracellular matrix production, immunosuppression and carcinogenesis. *SMAD4* mutation also contributes to other cancers, including pancreatic, esophageal and gastric cancer, in 55%, 6%, and 6% of cases, respectively. Notably, *SMAD4* downregulation is found in 60% of patients with metastatic colorectal cancer (Miyaki et al., 1999; Alazzouzi et al., 2005). Moreover, loss of SMAD4 function is associated with sensitivity to chemotherapy in colorectal cancer (Alhopuro et al., 2005; Yu et al., 2015).

1.3.3.5 P53 signaling pathway

The P53 signaling pathway responds to stress signals that impact cellular homeostasis by regulating apoptosis, cell-cycle arrest, DNA repair, senescence, and anti-angiogenesis (Harris & Levine, 2005). TP53 (The Tumor Protein P53) plays a central role in the P53 pathway. It responds to various cellular stresses leading to DNA damage, including exacerbating oncogenic signaling. In normal cells, TP53 protein is negatively regulated by MDM2 by targeting it for ubiquitylation. *TP53* mutations cause loss of protein function (**Figure 15**).

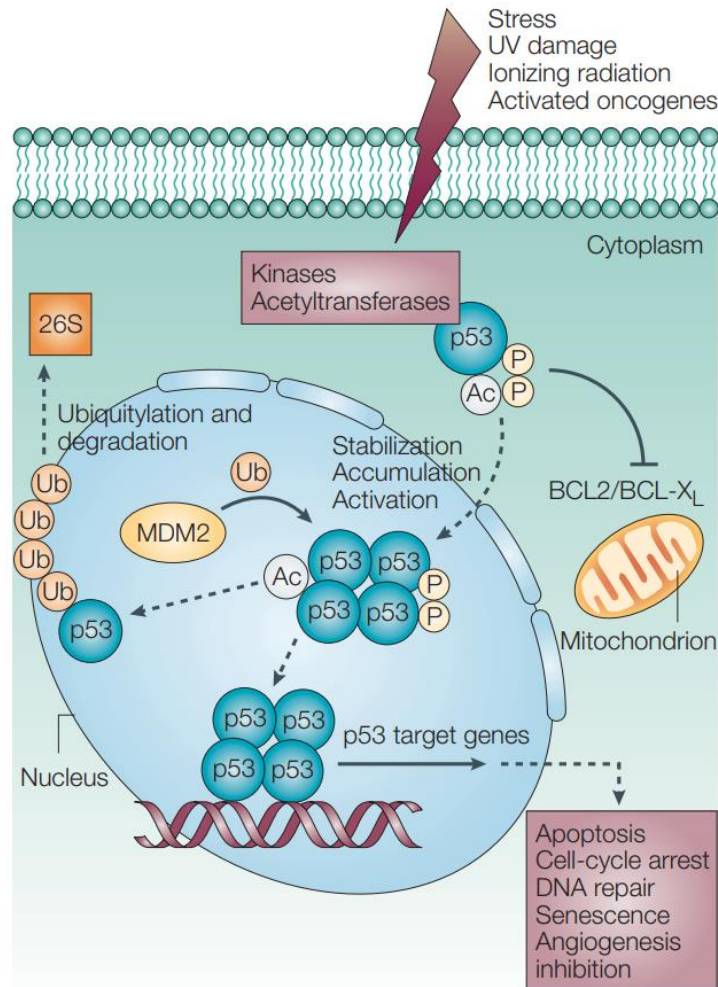


Figure 15: The P53 signaling pathway.

TP53 is activated by stress signals such as UV damage, ionizing radiation, and oncogenes. The activation leads to diverse cellular responses such as apoptosis, cell-cycle arrest, DNA repair, senescence, and anti-angiogenesis. When TP53 is no longer required, it is targeted for ubiquitylation by MDM2, exported out of the nucleus, and degraded by the 26S proteasome (Bode & Dong, 2004).

TP53, located on chromosome 17p, is one of the most frequently mutated genes in cancer. Its mutation has been found in various types of cancer (Saha et al., 2013). In colorectal cancer patients, the frequency of *TP53* mutations is about 50%. Li-Fraumeni syndrome (LFS), an inherited familial predisposition to a wide range of cancers, is also due to *TP53* mutations. *TP53* is a transcription factor that regulates the arrest in the G1 phase of the cell cycle (G1/S) and also G2/M checkpoint arrest (G2/M) by targeting p21, thereby facilitating DNA repair in case of an error in DNA replication. If the DNA damage is irreversible, TP53 activates the extrinsic apoptotic pathway by the tumor necrosis factor (*TNF*) receptor family or intrinsic apoptosis pathway via the BCL-2 family, including BAX, BBC3 (PUMA) and PMAIP1 (NOXA) (Haupt et al., 2003).

1.3.4 Molecular classification of colorectal cancer

Mutations are the "driving force" of biological evolution. During the evolution of life on our planet, undesired mutations are eliminated by natural selection, while beneficial mutations are selected. As mutations conferring a selective advantage are passed down from generation to generation, they are retained in the human genome and evolve from genetic variation to a normal gene. Tumors function much like biological species, and Darwinian evolution is a key driving force of the oncogenic process. As mentioned, colorectal cancer is caused by a multi-step process with cumulative genomic alterations that activate proto-oncogenes and inactivate tumor suppressor genes. Low mutation rates in somatic cells are important to prevent tumor formation within the lifespan of the host. However, increased genomic instability mechanisms that accelerate the mutation rate have been identified and constitute important tumorigenic drivers. A description of the main types of genomic instability observed in colorectal cancer can be found below.

1.3.4.1 Chromosomal instability (CIN)

CIN is the most common genomic instability mechanism, occurring in approximately 85% of sporadic colorectal cancers. It is manifested as the duplication or deletion of whole chromosomes or large chromosomal regions, resulting in aneuploidy and loss of heterozygosity (LOH). Loss of heterozygosity is an abnormal chromosomal event that losses one allele in heterozygous regions, caused by unbalanced chromosomal rearrangements and mitotic recombination. Three different mechanisms, including chromosome segregation defects, telomere dysfunction, and defective DNA damage response are known to mediate CIN (Pino & Chung, 2010).

Chromosome segregation defects. CIN can be driven by telomere dysfunction. The mitotic checkpoint ensures the accuracy of chromosome segregation by delaying the onset of anaphase until all duplicated chromatids are properly aligned on the metaphase plate. Defects of mitotic checkpoint signaling cause unbalanced segregation of chromosomes into daughter cells. Aurora kinase A (AURKA) plays an important role in centrosome function and duplication,

mitotic entry, and bipolar spindle assembly by regulating the segregation of chromatids at mitosis, histone modification, and cytokinesis. Its overexpression positively correlates with the CIN rate in colorectal cancer (Anand et al., 2003).

Telomere dysfunction. Telomeres are regions of repetitive DNA sequences at the end of eukaryotic chromosomes that protect the ends of chromosomes from fusing and breaking during segregation. In somatic cells, a portion of telomeric DNA is lost after each cell division. Stable shortening of telomeres plays a role in aging and cancer prevention. Telomerase is an enzyme complex that lengthens telomeres but normally is only active in stem cells. However, increased telomerase activity confers immortality to cells, thus allowing the tumorigenic process to proceed. The overexpression of telomerase has been found to increase the depth of tumor invasion in colorectal cancer (Gertler et al., 2004).

Defective DNA damage response. DNA damage response mechanisms involve a complex set of genes leading to DNA repair, cell cycle arrest and apoptosis in response to DNA damage. DNA double-strand breaks lead to chromosomal rearrangements, resulting in the loss, amplification and exchange of chromosomal segments. *TP53* is an essential gene in this process, as previously described. Others, like *MRE11*, have been found to have a crucial role in DNA double-strand break repair and are often inactivated in colorectal cancer (Giannini et al., 2002).

1.3.4.2 Microsatellite instability (MSI)

Microsatellites, also known as simple sequence repeats (SSRs), are short tandemly repeated sequences of DNA ranging from one to six base pairs in length. DNA polymerases produce errors within microsatellites at a frequency 10-100 fold higher than non-repetitive sequences. In normal cells, mismatch repair (MMR) proteins protect the genome integrity from incorrect insertion, deletion, and misincorporation of bases. Uncorrected insertions/deletions in coding microsatellites result in reading frame shifts during the translation process and truncated proteins that can contribute to the oncogenic process. Therefore, mismatch repair deficiency can lead to widespread mutations mostly in microsatellite regions, causing what is known as microsatellite instability (**Figure 16**). Microsatellite instability accounts for 15-20% of sporadic colorectal cancer and more than 95% of Lynch syndrome cases (Grady & Carethers,

2008). Microsatellite instability prevalence is approximately 20% in Stage I and II, 12% in stage III and 5% in stage IV (Battaglin et al., 2018). Microsatellite instability has also been considered a predictive biomarker for response to chemotherapy in colorectal cancer, given that microsatellite instability shows a significantly different overall survival (OS) and response rate to chemotherapy in colorectal cancer patients.

In sporadic colorectal cancer cases, microsatellite instability is most often caused by silencing *MLH1*, a mismatch repair gene, via CpG island promoter hypermethylation. In contrast, microsatellite instability in Lynch syndrome results from germline point mutations on MMR genes, including *MLH1*, *MSH2*, *MSH6*, and *PMS2* (Hampel et al., 2005). Microsatellite status can be determined by analyzing specific genes through immunohistochemistry (IHC) or PCR. *MLH1*, *MSH2*, *MSH6*, and *PMS2* protein staining on paraffin tissue sections and five nucleotide repeats on four different genes (*KIT* [BAT25], *MSH2* [BAT26 and D2S123], *APC* [D5S346], and *BRCA1* [D17S250]) detected through PCR, have been well established as markers to detect microsatellite instability in the clinics (Loukola et al., 2001). The loss of protein expression or the change in the length of these markers is considered a surrogate marker of microsatellite instability. Colorectal cancer tumors can be stratified into microsatellite instability-high (MSI-H), microsatellite instability-low (MSI-L), and microsatellite stable (MSS). Tumors that lose one or more of the mismatch repair proteins are referred to as MSI-H by using IHC as the detection. In PCR detection, MSI-H is defined when more than 30% of the markers assessed are unstable. In comparison, MSI-L exhibits lower than 30% of instability and MSS is considered when tumor material exhibits the same microsatellite length as normal cells. When used in parallel, IHC and PCR testing methods can increase accuracy.

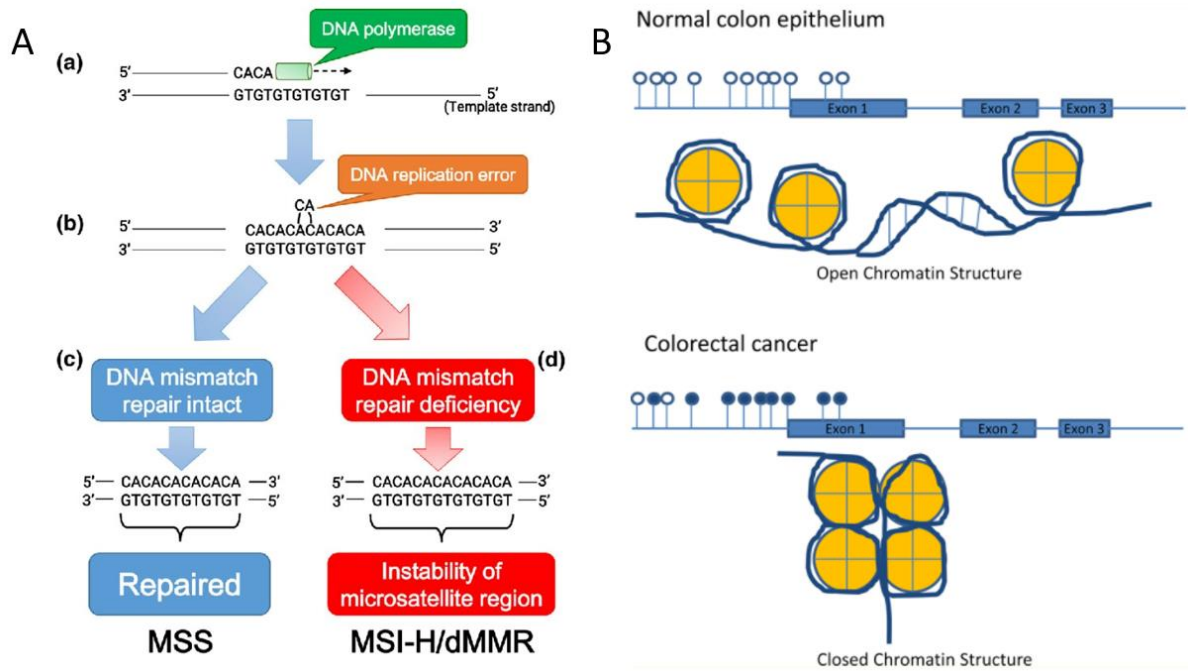


Figure 16: Diagram showing the mechanism of Microsatellite instability (MSI) and CpG island methylator phenotype (CIMP).

A) MSI: (a) The DNA polymerase begins to lengthen the single strand of DNA according to another original DNA molecule (cytosine/adenine [CA] × 6 repeats). (b) A set of excess CA is mistakenly incorporated into the new DNA duplex during the replication process due to slippage of the DNA polymerase. (c) An intact mismatch repair system can fix the error, rendering the same sequence as the template. (d) When mismatch repair is deficient, there will be 7 CA repeats. B) CIMP: In the upper diagram, open white circles represent unmethylated cytosines at promoter regions, maintaining the chromatin in an open structure, thus allowing gene transcription. However, hypermethylation of the promoter (filled circles in the lower panel) causes chromatin condensation, preventing transcriptional activity (Lao & Grady, 2011; Eso et al., 2020).

1.3.4.3 CpG island methylator phenotype (CIMP)

CpG dinucleotides or CpG sites are regions where guanosine bases follow cytosine bases. CpG islands (or CG islands) are regions rich in CpG dinucleotides. CIMP is defined by simultaneous hypermethylation of numerous CpG islands surrounding the promoter regions of several genes (Jover et al., 2011), which may silence the expression of tumor suppressor genes like *CDKN2A*, *MGMT* and *MLH1* and induce colorectal cancer (**Figure 16**) (Y. S. Kim & Deng, 2007). Approximately 17% of all colorectal tumors are classified as CIMP + (Ogino et al., 2006). For this subset of colorectal cancers, CIMP is a critical event in the early development of colorectal cancer. CIMP+ is associated with higher age at diagnosis, poor tumor differentiation, lower *TP53* mutation rates and high-rate of mutation of *KRAS* and *BRAF* (Issa, 2004). CIMP occurs

mainly in sporadic colorectal cancer patients with microsatellite instability and is rarely detected in Lynch syndrome cases. In addition, proximal colon cancer has a higher rate of CIMP compared to distal (30-40% vs. 3-12%) colorectal cancer (Curtin et al., 2011).

Like microsatellite instability, tumors can be classified as no-CIMP, CIMP-low and CIMP-high, or alternatively CIMP+ and CIMP-, by detecting a panel of markers. However, how to define the border between the different subtypes is not well established. CIMP status is determined by performing methylation-specific PCR amplification of the promoters of several genes. The most accepted panel of genes for CIMP screening contains *MLH1*, *CDKN2A*, *MINT1*, *MINT2* and *MINT31*, but additional panels containing genes such as *CACNA1G*, *CRABP1*, *IGF2*, *NEUROG1*, *RUNX3*, *SOCS1*, *HIC1*, *IGFBP3*, and *WRN* have been developed (Nazemalhosseini Mojarad et al., 2013). Besides, the development of high-throughput approaches is also an alternative way for CIMP detection, as it converts the perspective from single-gene methylation to whole-genome methylation profiling (Ang et al., 2010).

1.2.5.4 Consensus Molecular Subtypes (CMS)

Although many molecular features related to colorectal cancer have been identified, including changes in the DNA sequence, transcription alterations, and epigenetic modifications, all these characteristics describe the complexity of colorectal cancer based on a limited number of genetic or epigenetic alterations. In 2015, Guinney *et al.* retrieved the gene expression levels of 4151 patients from 18 independent patient datasets and used comprehensive bioinformatics analysis methods to propose an integrated molecular classification system for colorectal cancer, which yielded four different consensus molecular subtypes (CMS; **Figure 17**) (Guinney et al., 2015).

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

Figure 17: Consensus molecular subtypes classification of colorectal cancer.

CIMP, CpG island methylator phenotype; MSI, microsatellite instability; SCNA, somatic copy number alterations (Guinney et al., 2015).

CMS1 contains most microsatellite instability and CpG island methylator phenotype high colorectal cancers and is characterized by high mutation rates, *BRAF* mutations, and increased immune infiltration and activation. It is usually associated with decreased survival rate after recurrence. **CMS2** displays epithelial transcriptomic signatures and chromosomal instability, as measured by somatic copy number alteration counts and upregulation of WNT and MYC signaling. **CMS3** is usually characterized by abnormal metabolic pathways, mutations in *KRAS*, low levels of CpG island methylator phenotype and chromosomal instability. **CMS4** is characterized by chromosomal instability, stromal infiltration, activation of TGF-β signaling, and angiogenesis, and shows worse relapse-free and overall survival.

Although this classification method was proposed in 2015, it still faces considerable challenges in clinical application. For instance, there is a lack of robust biomarkers related to colorectal cancer subtyping. Determining biomarkers requires fundamental research, subsequent verification, and clinical validation. Meanwhile, the Consensus Molecular Subtypes classification has also introduced new prospects and challenges, including integrating other data types such as DNA methylation and miRNA expression profiles and shifting single-omic molecular subtyping into multi-omic molecular subtyping (Wang et al., 2019).

1.3.5 Treatment of colorectal cancer

Treatment of colorectal cancer includes surgery, radiotherapy and medication treatment. Medication therapy (systemic therapy) can also be divided into chemotherapy, targeted therapy and immunotherapy (Wolpin & Mayer, 2008; Yaghoubi et al., 2019).

1.3.5.1 Surgery

Surgery constitutes the first line of treatment for non-metastatic CRCs (stage 0 to stage III). Since the tumor has spread to other organs in Stage IV patients, surgery cannot remove the tumors in most cases. However, surgery is also recommended if the metastasis only applies to a limited area (Cervantes et al., 2023). The most common types of surgery are polypectomy, local excision and colectomy, and the main goal is to remove cancer tissue in the intestinal wall and nearby lymph nodes. Polypectomy is the surgical removal of a polyp, usually done through the colonoscope. Local excision is performed through the colonoscope to remove small colon cancers and a small amount of surrounding healthy tissue on the colon's wall. Colectomy is the removal of the cancerous part of the colon and a small section of the normal colon on both sides, then reattaching the remaining colon sections. Nearby lymph nodes are also removed to evaluate tumor spreading.

Depending on the patient's physical condition, size and location of the tumor, different surgical approaches can be chosen: colonoscopy, conventional open surgery, and laparoscopic surgery. Colonoscopy is performed by using a long, thin, flexible tube coupled to a small camera. Conventional open surgery is done through a single long incision in the abdomen. Laparoscopic surgery is done in the abdomen using small incisions (usually 0.5–1.5 cm) with the aid of a camera. The specific surgical instruments that perform operations are put in through other small incisions to remove tumors.

1.3.5.2 Radiotherapy

Radiotherapy is mainly adopted for rectal cancer patients, accounting for 28% of all colorectal tumors. Radiotherapy uses X-rays to cause cancer cell death. It can be used before surgery to reduce the size of the primary tumor, thus facilitating surgical removal (neoadjuvant radiotherapy). Radiotherapy can also be performed after surgery to eliminate potential remaining cancer cells after an operation and reduce the chance of recurrence (Häfner & Debus, 2016).

1.3.5.3 Chemotherapy

Despite recent advances, chemotherapy remains the mainstay of drug therapy for primary and metastatic colorectal cancer. The principle of chemotherapeutic drugs is to block cell division to inhibit the growth of cancer cells and induce apoptosis. However, it also affects normal cells causing side effects such as infections, anemia, sore mouth, diarrhea and/or constipation, among others (Schirmacher, 2019). Chemotherapy used after surgery is named adjuvant chemotherapy, and when used before surgery is called neoadjuvant chemotherapy. The commonly used chemotherapeutic drugs for colorectal cancer include:

Fluorouracil (5-FU) was the first drug to be approved by regulatory agencies as an effective treatment for colorectal cancer. Since its clinical application in 1957, it has been the most important drug for treating colorectal cancer for decades. Fluorouracil is a pyrimidine analog that, after entering the cell, can be converted either by orotate phosphoribosyl transferase (OPRT) or thymidine phosphorylase (TP) into three primary active metabolites, including fluorodeoxyuridine monophosphate (FdUMP), fluorodeoxyuridine triphosphate (FdUTP) and fluorouridine triphosphate (FUTP). FdUMP interferes with the synthesis of DNA by inhibiting thymidylate synthase activity. FdUTP is incorporated into DNA and disrupts normal DNA replication, and FUTP is incorporated into RNA and interferes with the pre-rRNA maturation and post-transcriptional modification (Longley et al., 2003) (**Figure 18**). Fluorouracil has been used to treat various cancers, especially gastrointestinal (colorectal and stomach) and other solid tumors such as liver, breast, cervical, bladder and skin cancer. It can be administered intravenously, intracavitary or orally. **Capecitabine** is an oral drug converted to 5-FU when

reaching the tumor site. The response rate of 5-FU, when used as a single agent, is about 20% for colorectal cancer patients (Bleiberg, 1996; Capitain et al., 2008). Fluorouracil efficacy can be significantly improved by combining it with leucovorin (LV, also known as folinic acid, FA). LV is a folate analog, a vitamin required for DNA synthesis, DNA repair, and methylation. LV increases the anti-tumor effects of 5-FU and reduces the damage of 5-FU to normal cells (Machover, 1997).

Oxaliplatin is a third-generation diamino cyclohexane (DACH) platinum compound. Platinum atoms of oxaliplatin form DNA adducts, blocking DNA replication and transcription. Oxaliplatin also induces apoptosis through upregulating *TP53*. The increased level of *TP53* protein leads to a subsequent accumulation of PUMA and NOXA pro-apoptotic proteins, leading to cell death and producing anti-tumor activity (Raymond et al., 1998; Arango et al., 2004) (**Figure 18**). Oxaliplatin displays good efficacy in colorectal and ovarian cancer and a moderate clinical benefit in gastric cancer, non-Hodgkin's lymphoma, non-small cell lung cancer, and head and neck tumors. Oxaliplatin exhibits a synergistic effect when combined with 5-FU leading to a reduction in the recurrence rate of colorectal cancer (de Gramont et al., 1997). Accordingly, it is often used to treat colorectal cancer patients combined with 5-FU (FOLFOX).

Irinotecan (CPT-11) is a DNA topoisomerase inhibitor. It can be converted into the active metabolite SN-38 by carboxylesterases (CES), both outside and inside the cell. Topoisomerase I is an enzyme that cuts the DNA strands required for DNA replication and prevents supercoiling. Irinotecan selectively acts on topoisomerase I, causing DNA single-strand and double-strand breaks and inducing apoptosis of cancer cells (Xu & Villalona-Calero, 2002) (**Figure 18**). Irinotecan is mainly used to treat metastatic colorectal cancer in combination with 5-FU and LV (FOLFIRI), including advanced colorectal cancer patients who have failed FOLFOX treatment. Irinotecan improves the efficacy and survival rate of advanced colorectal cancer patients when combined with 5-FU/LV. It is approved as one of the first-line treatment options for advanced colorectal cancer patients who cannot tolerate 5FU and/or LV (Cervantes et al., 2023).

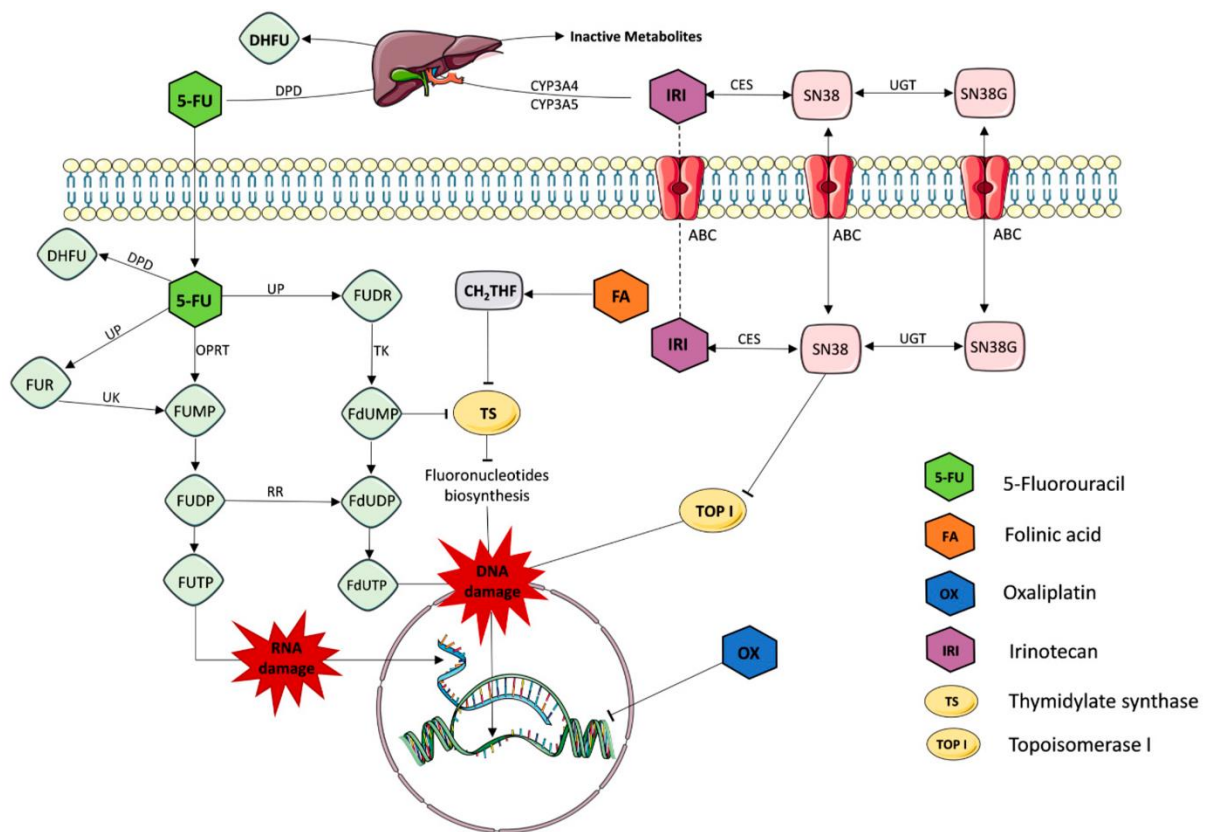


Figure 18: Mechanism of action and metabolism of 5-FU, oxaliplatin and irinotecan.

Thymidylate synthase (TS) is essential for thymidine synthesis, which is needed for DNA replication. 5-FU can be converted to FdUMP by uridine phosphorylase (UP) and uridine kinase (UK), inhibiting TS and causing DNA damage. Another active metabolite, FUDP, transformed by OPRT or UP and UK, can be phosphorylated to FUTP or converted to FdUDP by ribonucleotide reductase (RR). These fluoronucleotides compete with nucleotides and induce DNA and RNA damage. 5-FU is also converted to a less toxic dihydrofluorouracil (DHFU) by dihydropyrimidine dehydrogenase (DPD) or broken down by DPD-mediated conversion in the liver. Folic acid (FA) also binds to TS and competes with natural dUMP by converting to 5,10-methylene-tetrahydrofolate (CH₂THF), improving the anti-tumor effect. Carboxylesterases (CES) and uridine diphosphate glucuronosyltransferase (UGT) metabolize irinotecan (IRI) to SN38, inhibiting TOP I to prevent DNA synthesis. Alternatively, irinotecan can be inactivated by cytochrome P450 3A4 (CYP3A4) and Cytochrome P450 3A5 (CYP3A5) in the liver. Oxaliplatin (OX) induces DNA damage by forming DNA-Platinum adducts (Zoetemelk et al., 2020).

Other chemotherapeutic drugs used in the treatment of colorectal cancer are **trifluridine and tipiracil** (TAS-102), a combination oral treatment used in metastatic and refractory colorectal cancer. TAS-102 is composed of trifluridine (a thymidine analog) and tipiracil hydrochloride (a thymidine phosphorylase inhibitor). Trifluridine interferes with the DNA replication of tumor cells, blocking their proliferation. Tipiracil hydrochloride inhibits TPase, an enzyme that rapidly degrades trifluridine, restricting its activity.

In general clinical practice, oncologists typically use two or more chemotherapeutic drugs to increase efficacy (**Table 1**). However, the toxic side effects can be severe. Patients unable to tolerate combination therapy are treated with a single agent (monotherapy). Chemotherapeutic drugs can also be combined with targeted agents and immunotherapy to increase response rates (described below).

Table 1: Frequent therapeutic regimens used for advanced colorectal cancer treatment and response rates.

Regimen	Drugs	Response rate	Reference
5-FU/LV	Fluorouracil / Leucovorin	5% – 48%	(Petrelli et al., 1987, 1989; Poon et al., 1989; Saltz et al., 2000)
IFL	Irinotecan / Fluorouracil / Leucovorin	31% – 39%	(Petrelli et al., 1989; Saltz et al., 2000)
FOLFOX	Leucovorin (FOL) / Fluorouracil (F) / Oxaliplatin (OX)	34% – 45%	(Colucci et al., 2005; Goldberg et al., 2004)
FOLFIRI	Leucovorin / Fluorouracil / Irinotecan	31% – 34%	(Colucci et al., 2005; Falcone et al., 2007; Souglakos et al., 2006)
FOLFOXIRI	Leucovorin / Fluorouracil / Oxaliplatin / Irinotecan	43% – 60%	(Falcone et al., 2007; Souglakos et al., 2006)

Chemotherapy regimens are administered cyclically to allow the patients to recover from the toxic side effects. The chemotherapeutic cycle usually lasts 2 to 4 weeks, and the number of cycles received by a patient is conditioned by the course of the disease. The side effects caused by chemotherapies have been classified in severity levels according to Adverse Event Reporting Systems.

1.3.5.4 Targeted therapy

The mechanism of action of targeted drugs differs from chemotherapeutic drugs, which impact all rapidly dividing cells in the body. Targeted therapy drugs are designed based on the corresponding identified molecular targets that control the growth of cancer cells. Once the targeted drug enters the body, it binds specifically to the selected molecular targets upregulated by tumor cells, being less harmful to normal cells.

Approved targeted drugs for treating colorectal cancer include vascular endothelial growth factor (VEGF) inhibitors, epidermal growth factor receptor (EGFR) inhibitors, BRAF inhibitors, and multikinase inhibitors (Xie et al., 2020).

VEGF contributes to the formation of new blood vessels (neoangiogenesis) within tumors, which are needed to sustain the nutrients and oxygen supporting the increased cell growth. Drugs blocking VEGF/vascular endothelial growth factor receptor (VEGFR) inhibit angiogenesis. The drugs include monoclonal antibodies (Bevacizumab and Ramucirumab and recombinant proteins (Ziv-aflibercept).

EGFR is a protein with tyrosine kinase activity that contributes to cancer cell growth. By combining with epidermal growth factor (EGF), EGFR activates the transcription of genes promoting cell division and proliferation, survival, migration and invasion and angiogenesis. EGFR-targeting drugs include the monoclonal antibodies Cetuximab and Panitumumab.

Regorafenib is an oral multikinase inhibitor of angiogenic, stromal and oncogenic receptor tyrosine kinases that target VEGFR, tyrosine-protein kinase receptor TIE-2, proto-oncogene *RAF-1*, *BRAF* and MAP kinase.

1.3.5.5 Immunotherapy

Immunotherapeutic approaches enhance the patient's immune system or alter the immune system to boost the recognition, activation and elimination of cancer cells.

The immune system constantly screens the body and attacks pathogens or foreign cells, such as bacteria or cancer cells. Since cancer cells originate from normal cells, the immune system

does not always recognize them as foreign. In addition, cancer cells evade immunological surveillance by targeting regulatory T cells, antigen presentation, tolerance and immune deviation (Vinay et al., 2015).

Immunotherapy includes oncolytic virus therapies, cancer vaccines, cytokine therapies, adoptive cell transfer, and immune checkpoint inhibitors (Y. Zhang & Zhang, 2020). The later has been approved to treat many types of cancer exhibiting high mutation burden. Therefore, immunotherapy is not yet as widely available as surgery, chemotherapy or radiation therapy. Immunotherapy can be used alone or with other cancer therapies, and some agents are associated with severe side effects. It is of note the long-term therapeutic effects of chemotherapeutic drugs compared to other therapies.

Currently, three monoclonal antibodies, Ipilimumab, Pembrolizumab and Nivolumab, which are immune checkpoint inhibitors, have been approved by US Food and Drug Administration (FDA), but only for treating metastatic colorectal cancer patients with microsatellite instability-high tumors (Ganesh et al., 2019). Therefore, the following description regarding immunotherapy in colorectal cancer is restricted to checkpoint inhibitors.

Immune checkpoints are essential in the regulation of the immune system. They prevent immune cells from overreacting against healthy cells or tissues in the body. In the context of cytotoxic T cells, when immune checkpoint proteins on the surface of T cells recognize binding proteins on the surface of antigen-presenting cells or cancer cells, the priming and cytotoxic activity of immune cells is impaired, respectively. Immune checkpoint inhibitors block receptors on the cell membrane. They prevent cancer cells' immune surveillance and the activation of the T cells to engage a proper cytotoxic response (van Parijs & Abbas, 1998; Jennifer, 2013). Checkpoint inhibitors include CTLA-4 (cytotoxic T lymphocyte-associated protein 4) inhibitors, PD-1 (programmed cell death protein 1) inhibitors, and PD-L1 (programmed cell death ligand 1) inhibitors. CTLA-4 and PD-1 are expressed in T cells. PD-L1 is found in cancer and other cells.

CTLA-4 inhibitors. CTLA-4 is expressed on the surface of activated T cells. CTLA-4 binds to B7-1 and B7-2 proteins on antigen-presenting cells and transmits inhibitory signals. When CTLA-4 is blocked, inhibitory signals on cytotoxic T cells are weakened, and ultimately their ability to kill cancer cells is enhanced (Ganesh et al., 2019). Ipilimumab is an anti-CTLA-4

monoclonal antibody approved for the treatment of colorectal cancer patients in combination with PD-1 inhibitors.

PD-1/PD-L1 inhibitors. PD-1 on the surface of T cells restricts the duration of the immune responses. When it binds to PD-L1 on the surface of cells, the activity of T cells is inhibited. Tumor cells upregulate PD-L1 and lead to the engagement of inhibitory signals on T cells. By blocking the PD-1/PD-L1 axis, these drugs prevent PD-1/PD-L1 binding, thereby recovering immune responses against cancer cells (Oliveira et al., 2019). Pembrolizumab and Nivolumab are monoclonal antibodies that target and block PD-1.

Alternative promising immunotherapies for colorectal cancer patients include adoptive T-cell transfer therapies and cancer vaccine-based therapies. Although the clinical research is still in the early phases, these novel approaches alone or in combination with the approved agents offer optimistic chances for future immunotherapies for colorectal cancer patients.

1.3.5.6 Therapeutic strategies for colorectal cancer by stage

The treatment options for colon cancer are mainly determined by tumor staging (**Table 2**). For non-metastatic tumors, surgery is the primary treatment. Radiotherapy and chemotherapy can also be prescribed before (nonadjacency) and/or after surgery (adjutancy). Contrary, metastatic tumors are mainly treated with pharmacological therapy. The survival rate of colorectal cancer patients largely depends on the tumor stage. The 5-year survival rate is 91% for the localized stage (stage 0-II), 72% for the regional stage, and 16% for stage IV (Miller et al., 2019).

Table 2: Colorectal cancer treatment regimens by stage.

Stage	Surgery	Chemotherapy/Biologics	Radiation
Stage 0	Yes	No	No
Stage I	Yes	No	No
Stage II	Yes	Yes, for rectal and high risk colon cancers. FOLFOX or CapeOx	Yes, for rectal cancer. Given in tandem with 5-FU or Xeloda
Stage III	Yes	FOLFOX, CapeOx	Yes, for rectal cancer. Given in tandem with 5-FU or Xeloda
Stage IV	Yes, if the tumor is obstructive or blocking the bowel Some patients become surgical candidates for liver, lung or peritoneal surgery Usually not, if the tumor is not blocking the bowel	FOLFOX or FOLFIRI plus Avastin or Erbitux or Vectibix or Zaltrap, Stivarga, Lonsurf	Yes, for rectal cancer and in certain other cases Interventional radiology for liver and lung metastases

Notes: Colon Cancer Alliance (<https://www.ccalliance.org/>). FOL, leucovorin calcium (folinic acid); OX, oxaliplatin; Cape, capecitabine; F, 5-FU; IRI, irinotecan; Avastin, bevacizumab; Erbitux, cetuximab; Vectibix, panitumumab; Zaltrap, aflibercept; Stivarga, regorafenib; Lonsurf, Trifluridine/Tipiracil (TAS-102).

Treatment options for Stage 0 and Stage I colorectal cancer

As the tumors in these two stages are confined to the intestinal wall, tumor tissues are excised by surgical resection for the majority of the patients, and no further treatment is required.

Treatment options for Stage II colorectal cancer

Tumors at this stage may have penetrated the intestinal wall and invaded the surrounding tissues. Surgical treatment for these patients involves the excision of the intestinal tract where the tumor resides and surrounding lymph nodes. Adjuvant chemotherapy is not routinely recommended for patients in stage II (Böckelman et al., 2015). Patients should be tested for microsatellite instability or mismatch repair deficiency to guide adjuvant/neoadjuvant chemotherapy to eliminate residual tumor cells if there is a higher risk of recurrence, characterized by: 1) tumor is high grade; 2) tumor has grown into nearby blood or lymph vessels; 3) surgery did not remove at least 12 lymph nodes; 4) tumor cells were found in/or near the margin of the excised tissue, meaning that some cancer tissue may have been left behind; 5) tumor had obstructed the colon; 6) tumor caused a perforation in the colon's wall. If the patients with high-risk factors display proficient mismatch repair/microsatellite stable

tumors, adjuvant chemotherapy is recommended (Böckelman et al., 2015). Adjuvant regimens include 5-FU/leucovorin, capecitabine, or combination chemotherapy with FOLFOX (5-FU, leucovorin, oxaliplatin) or CapeOx (capecitabine and oxaliplatin). If the patients with high-risk factors display mismatch repair deficiency/microsatellite instability tumors, oxaliplatin-containing adjuvant chemotherapy is recommended (André et al., 2004). Adjuvant radiotherapy or neoadjuvant radiotherapy for patients with rectal cancer is also recommended (Baxter et al., 2021).

Treatment options for Stage III colorectal cancer

In stage III, tumors have spread to nearby lymph nodes. The treatment regimen includes surgery, radiotherapy (rectal cancer), and chemotherapy. Approved adjuvant chemotherapy includes FOLFOX (5-FU, leucovorin calcium, and oxaliplatin) or CapeOx (capecitabine and oxaliplatin). Adjuvant oxaliplatin-containing chemotherapy for 6 months is recommended for stage III colorectal cancer patients with a high-risk (T4 and/or N2). For patients with low-risk (T1, T2, or T3 and N1), adjuvant oxaliplatin-containing chemotherapy is recommended for 3 or 6 months. It is conducted after discussing with the patient the potential benefits and risks (Lieu et al., 2019) (**Table 2**). If the tumor is too large to be surgically removed, neoadjuvant chemotherapy is considered to reduce the tumor size and/or reduce the difficulty of surgery.

Treatment options for Stage IV colorectal cancer

In stage IV, tumor cells have reached distant organs through hematogenous dissemination. Surgery is recommended for patients with locally recurrent and/or liver-only, and/or lung-only metastatic disease. Surgery is also recommended if there is bowel obstruction, impending perforation, uncontrolled bleeding, or uncontrolled pain (van Cutsem et al., 2016). In addition, for patients with resectable metastases, 6 months of adjuvant chemotherapy is mandatory. Approved regimens include FOLFOX and CapeOx. Patients with unresectable liver metastases but sensitive to neoadjuvant chemotherapy can become candidates for tumor resection occasionally. This is called rescue surgery. Besides surgery, systemic therapy, including chemotherapy, targeted therapy and immunotherapy, is the standard treatment to control tumor

growth. Before the palliative therapy starts, molecular profiling (*RAS* and *BRAF* status), tumor location (left versus right), and patient fitness should be considered. FDA-approved chemotherapeutic and targeted agents for stage IV colorectal cancer patients include cytotoxic agents (5-FU, capecitabine, irinotecan and oxaliplatin), anti-VEGF/R (bevacizumab, aflibercept and ramucirumab), anti-EGFR (cetuximab and panitumumab), regorafenib and TAS-102. First-line regimens for colorectal cancer patients include: a cytotoxic doublet plus an anti-EGFR antibody for (left-sided *RAS* WT), a cytotoxic triplet plus bevacizumab (right-sided *RAS* WT) or a cytotoxic doublet/triplet plus bevacizumab (*RAS* or *BRAF*-mutant) (van Cutsem et al., 2016). In addition, immunotherapeutic agents pembrolizumab (PD-1 inhibitor) (Diaz et al., 2020), nivolumab (PD-1 inhibitor) (Overman et al., 2017), and ipilimumab (CTLA-4 inhibitor) (Overman, Lonardi, et al., 2018) have been approved by the FDA for the treatment of deficient mismatch repair/microsatellite instability-high metastatic colorectal cancer.

1.3.6 Resistance to Cancer Chemotherapy

Although different therapeutic agents for colorectal cancer have emerged in the past few decades, chemotherapy remains the mainstay of colorectal treatment. The average response rate of regimens composed of 5-FU, oxaliplatin and irinotecan is only about 50% (**Table 1**). The main cause of therapeutic failure is drug resistance. The resistance mechanisms can be intrinsic (before drug exposure) and acquired (generated as a consequence of drug treatment). Frequent molecular mechanisms of drug resistance include enhanced efflux of drugs, alteration in genetic factors and elevated growth factors (Luqmani, 2005; Bukowski et al., 2020).

Enhanced efflux of drugs. The ATP Binding Cassette Subfamily B Member 1 (*ABCB1*) transports various substances over the cell membrane, including chemotherapeutic agents. For example, P-glycoprotein, encoded by *ABCB1*, can eliminate irinotecan and its metabolites from cancer cells. Studies have shown genetic variation in *ABCB1* was associated with a lower response to irinotecan-based treatment (Glimelius et al., 2011).

Alteration in genetic factors. Some drugs, such as oxaliplatin, can induce *TP53* activation by damaging DNA, leading to cell death. Conversely, the inactivation of *TP53* mutations in

cancer cells allows continued replication, making them resistant to chemotherapeutic drugs (Y. Huang et al., 2019; F. Mantovani et al., 2019).

Increased DNA repair capacity. Platinum-based drugs induce interstrand and intrastrand DNA cross-links leading to apoptotic cell death. In the nucleotide excision repair pathway, the excision repair cross-complementation group 1 (ERCC1) protein can recognize the DNA-platinum adducts and incise the DNA double-strand to remove the adducts. A significant correlation was shown between the overexpression of ERCC1 proteins and the development of oxaliplatin resistance in colorectal cancer patients (P. Li et al., 2013).

Elevated growth factors. Studies have shown transforming growth factor- β , a growth factor with anti-inflammatory properties is a poor prognostic factor in colorectal cancer and promotes colorectal cancer initiation and progression (Brunen et al., 2013). These are common mechanisms of chemotherapy resistance. The specific resistant mechanisms of 5-FU, oxaliplatin, and irinotecan are further discussed below.

Other mechanisms include genetic factors such as gene mutation, amplification, epigenetic alterations, elevated xenobiotics metabolism, drug distribution, drug inactivation, and cell cycle arrest (Holohan et al., 2013; Bukowski et al., 2020).

1.3.6.1 Mechanisms of resistance to 5-FU

The resistance mechanisms of 5-FU result from various causes, including alteration in target molecules and alteration of drug influx and efflux (N. Zhang et al., 2008). As described before, 5-FU is able to inhibit thymidylate synthase, which catalyzes the conversion of deoxyuridine monophosphate to deoxythymidine monophosphate, thereby leading to DNA damage (Longley et al., 2003). Gene mutations and polymorphism of Thymidylate Synthetase (*TYMS*), the gene encoding thymidylate synthase, leading to acquired resistance (Berger et al., 1988; Lecomte et al., 2004). Studies have shown that colorectal cancer patients with low thymidylate synthase expression are more sensitive to 5-FU and have a higher overall survival (Qiu et al., 2008; Abdallah et al., 2015). Furthermore, thymidine phosphorylase, uridine phosphorylase, and orotate phosphoribosyltransferase promote the conversion of 5-FU into active metabolites in the cell. Therefore, high levels of these enzymes can increase the sensitivity of 5-FU treatment

(Yanagisawa et al., 2007). Dihydropyrimidine dehydrogenase contributes to the degradation of 5-FU in the liver. Studies have shown that dihydropyrimidine dehydrogenase expression levels negatively correlate with drug sensitivity (Sakowicz-Burkiewicz et al., 2016).

1.3.6.2 Mechanisms of resistance to Oxaliplatin

The resistance to oxaliplatin can occur mainly by detoxification, increased drug efflux and enhanced DNA repair mechanism (Martinez-Balibrea et al., 2015; Rottenberg et al., 2021). Oxaliplatin is a third-generation platinum agent, and it works by inducing interstrand and intrastrand DNA cross-links leading to apoptotic cell death (Hector et al., 2001). One of the known resistant mechanisms is related to glutathione, a tripeptide (cysteine, glycine, and glutamic acid) that can bind to Pt and form glutathione-platinum conjugates (Meijer et al., 1992). The glutathione-platinum conjugates are effectively eliminated from the cell by drug efflux pumps (Ishikawa & Ali-Osman, 1993). High levels of glutathione contribute to oxaliplatin resistance, but this observation has not been validated in *in vivo* models (Kelland, 1993). Cellular transporters such as the solute carrier (SLC) transporters and ATP-binding cassette (ABC) transporters also contribute to pumping out the drug. Regarding oxaliplatin resistance mechanisms related to DNA repair, the nucleotide excision repair pathway removes the DNA damage caused by oxaliplatin. In the nucleotide excision repair pathway, the excision repair cross-complementation group 1 (ERCC1) protein can recognize the DNA-platinum adducts and incise the DNA double-strand to remove the adducts. Retrospective studies have shown colorectal cancer patients with high levels of ERCC1 have a poor prognosis after oxaliplatin-based therapy (P. Li et al., 2013; Shirota et al., 2001). Other mechanisms involving oxaliplatin resistance include triggering cell death (apoptotic and nonapoptotic), arresting cell cycle and epigenetic mechanisms (Wang & Lippard, 2005; Martinez-Balibrea et al., 2015).

1.3.6.3 Mechanisms of resistance to Irinotecan

The resistance mechanisms of irinotecan involve the metabolism of irinotecan, drug efflux, DNA Topoisomerase I (*TOPI*) level and mutations (Xu & Villalona-Calero, 2002; Hammond et al., 2016). Irinotecan is a semisynthetic analog of camptothecin, converted to SN-38 outside

or inside the cell. SN-38 contributes to the formation of the topoisomerase 1-inhibitor-DNA complex that affects DNA synthesis. Studies have shown that the *TOP1* copy number positively correlated with the sensitivity of cells to irinotecan (Palshof et al., 2017). Mutations of *TOP1* reduce the affinity of the *TOP1* enzyme to SN-38, decreasing the sensitivity to irinotecan. Carboxylesterase, uridine diphosphate glucuronosyltransferase, hepatic cytochrome P-450 enzyme CYP3A, β -glucuronidase and ATP binding cassette transporters are involved in the irinotecan uptake and metabolism and may affect the sensitivity to irinotecan (de Man et al., 2018). For example, carboxylesterase participates in the conversion of irinotecan to SN-38. Wang et al. found different single-nucleotide polymorphisms affecting the efficiency of carboxylesterase conversion of irinotecan (D. Wang et al., 2018). However, whether carboxylesterase directly affects the sensitivity of irinotecan still needs further study.

1.3.7 Biomarkers: diagnostic, prognostic and predictive

Oncologists commonly use a trial-and-error strategy for cancer patients until they find the treatment therapy most effective. In this process, monitoring the response rate is essential. At present, the response of anti-tumor drugs is evaluated according to RECIST (Response Evaluation Criteria in Solid Tumors), which was created in 2000 and recently revised (RECIST V. 1.1) (Eisenhauer et al., 2009). RECIST sets a standard method for measuring how patients respond to treatment. It enables oncologists to make informed decisions to maintain medication or switch to another treatment. The RECIST guideline uses imaging technology, computerized tomography (CT) and Magnetic resonance imaging (MRI) to measure the longest diameters of tumor lesions. The number of target lesions should be up to five. Tumors are imaged as closely as possible before and after several treatment cycles to compare the baseline lesion volume. The response is defined as Complete Response (CR) if all target lesions disappear. Partial Response (PR) indicates at least a 30% decrease in the sum of target lesions' longest diameter (LD). More than a 20% increase in the sum of the LD of target lesions is considered Progressive Disease (PD). Stable Disease (SD) refers to target lesions with neither sufficient shrinkage to 30% nor sufficient increase to 20% (Eisenhauer et al., 2009).

As mentioned previously, the response rate for most common regimens used for colorectal cancer treatment is relatively low. Although medical oncologists can modify the treatment

strategy according to the monitoring of the lesions by RECIST, about half of the patients do not benefit from the first line of treatment (**Table 1**). The current tendency is to separate patients into subgroups by molecular features of their specific tumors and prescribe a selected regimen with a higher response rate for the individual patients. Biomarkers are crucial in this selection process.

A biomarker is an objective medical sign that can be measured to indicate the state of patients, including disease outcome or treatment response. A biomarker can be a gene, protein, or other substance measured from patients' samples (Group et al., 2001). Biomarkers can be classified as prognostic, diagnostic, predictive, pharmacodynamics, and recurrence biomarkers. Since this thesis is aimed at the identification of novel predictive biomarkers for treating colorectal cancer patients, the following mainly focuses on predictive biomarkers.

Predictive biomarkers in oncology refer to molecular profiles that can predict whether an individual patient will be sensitive to the drug treatment for counteracting tumor growth or promoting tumor elimination. The clinically approved biomarkers capable of predicting treatment response for colorectal cancer patients are *KRAS* mutations in targeted therapy with anti-EGFR agents and microsatellite instability-high (MSI-H) for immunotherapy. *KRAS* is a downstream effector of EGFR. Those metastatic colorectal cancer patients with *KRAS* mutations are resistant to anti-EGFR therapy (Lievre et al., 2006; Amado et al., 2008). Therefore, it is recommended to assess *KRAS* mutation status currently at the time of diagnosis. In addition, since studies have shown that anti-programmed cell death-1 (PD-1) therapy confers a clinical benefit to patients with MSI-H mCRC, pembrolizumab and nivolumab were approved for the treatment of patients with MSI-H metastatic colorectal cancer patients (Overman, Lonardi, et al., 2018).

Since chemotherapeutic agents currently constitute the backbone of the treatment for advanced colorectal patients, many promising predictive biomarkers of response to 5-FU, oxaliplatin and irinotecan have been studied. Although, until nowadays, no biomarker of response to chemotherapy for colorectal cancer patients has reached clinical use due to the lack of validation on a large cohort of patients, limited tissue quality, and no standardization on testing (Koncina et al., 2020b), recent studies have shown that some promising predictive biomarkers may be used soon. The following are some emerging chemotherapeutic biomarkers.

1.3.7.1 Microsatellite instability

Microsatellite instability has been shown to potentially predict the efficacy of 5-FU and irinotecan. Studies have indicated that microsatellite instability-high colorectal cancer patients with stages II and III tumors do not benefit from 5-FU therapy alone (Popat et al., 2005; Sargent et al., 2010). However, due to the difference in chemotherapy regimen among studies and various assays for microsatellite instability detection, it is challenging to compare the results from different studies and therefore develop a consistent conclusion on whether microsatellite status can be used as a reliable biomarker. Studies with irinotecan showed promising results. Colorectal cancer patients with microsatellite stable tumors do not benefit from irinotecan treatment, while patients with microsatellite instable tumors have a significantly higher response rate (Fallik et al., 2003; Bertagnoli et al., 2009). Given that there is a limited number of studies regarding the correlation between irinotecan efficacy and microsatellite status, more studies need to be conducted to support this observation.

1.3.7.2 Excision repair cross-complementation group 1 (*ERCC1*)

As a DNA repair protein, *ERCC1* has shown a potential role in predicting the sensitivity of oxaliplatin in colorectal cancer patients. Studies have shown that a high level of *ERCC1* correlates with poor response to oxaliplatin (Chua et al., 2009; P. Li et al., 2013). The oxaliplatin-induced nucleotide excision repair pathway may play a vital role in drug resistance. In the DNA repair process, a high *ERCC1* level in the nucleotide excision repair pathway could repair the DNA double-strand and protect the cell from oxaliplatin damage (Paré et al., 2008). Due to the limited number of studies investigating the clinical value of *ERCC1* as a biomarker of response to oxaliplatin, further investigation is required before it can be routinely used in the clinic.

1.3.7.3 Topoisomerase I (*TOP1*)

TOP1 is the target of irinotecan, and it is, therefore, reasonable to speculate that the level of expression of *TOP1* or its mutation profile would affect the irinotecan sensitivity. Some studies have described *TOP1* mutations contributing to camptothecin resistance, the first *TOP1* inhibitor isolated from the Camptotheca tree (LI et al., 1996; Gongora et al., 2011). Decreased level of *TOP1* after camptothecin treatment was observed in both preclinical (Giovanella et al., 1989) and clinical studies (Horisberger et al., 2009). Moreover, in a large cohort study involving 1313 patients, the irinotecan response rate positively correlated with TOP1 protein expression levels assessed by immunohistochemistry (Braun et al., 2008). However, this observation could not be confirmed in a validation patient cohort, which included 545 patients (Koopman et al., 2009). These inconsistent results indicate that further validation studies are warranted to confirm *TOP1* predictive power.

2. OBJECTIVES

The initiation and progression of colorectal cancer involves multiple genetic alterations. A variety of genes and protein changes have been found implicated in the process of colorectal cancer tumorigenesis. Better understanding of colorectal cancer molecular features should lead to the improvement of treatment efficacy. Currently, the efficacy of chemotherapeutic compounds as single agents is not greater than 30%. In other words, approximately 70% of patients are treated with one compound for which they have no clinical response and suffer the associated side effects. Treatment failure is caused by primary and acquired resistance to drug exposure. It has been shown that thoroughly validated molecular biomarkers can accurately predict treatment response. A good example available for patients with colorectal cancer is the mutational state of *KRAS* to predict the response to EGFR inhibitors. Although new promising biomarkers have emerged for colorectal cancer in recent years, unfortunately, there are currently no biomarkers capable of predicting the response to chemotherapeutic agents approved for patients with colorectal cancer.

Here, we hypothesize that colorectal cancer cell lines can be used as a tool to identify predictive biomarkers of drug response in colorectal cancer patients by incorporating the molecular features (DNA, mRNA and protein) and the drug sensitivity data of the cell line. In addition, these biomarkers identified from the cell line panel could be validated in the patients' samples and implemented in the clinical setting to guide the chemotherapeutic treatment of colorectal cancer patients.

Therefore, the specific aims of this study are:

1. Characterize a panel of cell lines' sensitivity to 5-FU, oxaliplatin and irinotecan.
2. Use “omics” to identify biomarkers of prediction of response to 5-FU, oxaliplatin and irinotecan.
3. Validate the predictive power of the biomarkers in a cohort of 548 patients treated with these agents.

3. MATERIALS AND METHODS

3.1 Materials

3.1.1 Cell lines

A total of 92 cell lines were used (see **Table 3** for all the cell lines' details): CACO2, COLO201, COLO205, COLO320DM, DLD1, HCT116, HCT15, HCT8, HT29, LOVO, LS174T, LS180, LS411, RKO, SKCO1, SW1116, SW1463, SW403, SW48, SW480, SW620, SW837, SW948, T84 and WIDR were purchased from ATCC (Manassas, VA, USA). HDC108, HDC111, HDC114, HDC133, HDC15, HDC54, HDC75, HDC8, HDC87 and HDC9 were kind gifts from Dr. Johannes Gebert (Institute of Pathology, University Hospital Heidelberg, Germany). CCK81, COLO678, CX1, GP2D, HCA46, HDC57, HDC90, HRA19, HT115, HT55, LIM1215, LIM1863, LIM1899, LIM2099, LIM2405, LIM2537, LIM2550, LIM2551, NCIH716, NCIH747, SNU175, SNU283, SNUC1, SNUC4, VACO10, VACO432 and VACO4S were a kind gift from Dr. John Mariadason (Olivia Newton-John Cancer Research Center, Australia). HCC2998, KM12, RW2982 and RW7213 were kindly provided by Dr. L. H. Augenlicht (Albert Einstein Cancer Center, USA). GP5D, HCA7, HUTU80, SNUC2B and VACO5 were kind gifts from Dr. L.A. Aaltonen (Biomedicum Helsinki, Finland). ALA, CO115, FET, ISRECO1, ISRECO2, ISRECO3, LS1034, LS513, TC71 and V9P were kind gifts from Dr. Richard Hamelin (INSERM U434 CEPH, France). C10, C106, C125, C135, C32, C80 and C99 were kindly provided by Dr. Walter Bodmer (Weatherall Institute of Molecular Medicine, USA). DIFI and GEO were kind gifts from Dr. Z Fan (M.D. Anderson Cancer Center, TX, USA). SW1222 was a kind gift from Prof. M. Herlyn (The Wistar Institute, PA, USA). TLA-HEK293-T was purchased from Open Biosystem and maintained with Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum. The other cell lines were cultured with Minimum Essential Media (MEM) (Life Technologies, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS), antibiotic/antimycotic (100 U/ml streptomycin, 100 U/ml penicillin, and 0.25 µg/ml amphotericin B), MEM non-essential amino acids solution and 10mM HEPES buffer solution (all from Life Technologies) at 37°C and 5% CO₂. The cell lines were cultured until they reached 70-80% confluence and were harvested

using 0.25% trypsin and 1 mM EDTA in phosphate-buffered saline (PBS). All cell lines were negative for mycoplasma contamination before the screening assay. The identity of the cell lines was confirmed by hierarchical clustering analysis comparing our internal transcriptional profiling and cell lines from other repositories or laboratories, such as the Cancer Cell Line Encyclopedia (CCLE) and Dr. Alberto Bardelli colorectal cancer cell line collection (Medico et al., 2015).

Table 3: Cell lines used in this study

Order	Cell line	MS status ^a	CIMP status ^b	CMS ^c	Cell seeding density for SRB assay (cells/cm ²)	ccRECIST irinotecan ^d	ccRECIST 5-FU	ccRECIST Oxaliplatin
1	ALA	MSS	-	-	5 x 10 ³	N	-	-
2	C10	MSS	CIMP-	CMS4	7.5 x 10 ³	R	N	N
3	C106	MSS	CIMP-	CMS2	1.25 x 10 ⁴	N	N	N
4	C125	MSS	CIMP-	CMS2	3.5 x 10 ⁴	N	N	R
5	C135	MSI	CIMP+	CMS1	1.5 x 10 ⁴	R	R	R
6	C32	MSS	CIMP-	CMS2	7 x 10 ³	N	N	N
7	C80	MSS	CIMP-	CMS2	2 x 10 ⁴	N	R	R
8	C99	MSS	CIMP-	CMS2	2.5 x 10 ⁴	R	R	R
9	CACO2	MSS	CIMP-	CMS2	2.5 x 10 ³	N	-	-
10	CCK81	MSI	CIMP-	CMS2	2 x 10 ⁴	N	R	R
11	CO115	MSI	CIMP+	CMS1	1 x 10 ⁴	R	R	R
12	COLO201	MSS	CIMP+	CMS1	1.5 x 10 ⁴	R	-	-
13	COLO205	MSS	CIMP+	CMS1	1.5 x 10 ⁴	R	-	-
14	COLO320DM	MSS	CIMP-	-	1 x 10 ⁴	N	-	-
15	COLO678	MSS	CIMP+	CMS4	1 x 10 ⁴	N	N	N
16	CX1	MSS	CIMP+	CMS3	1 x 10 ⁴	N	R	R
17	DIFI	MSS	-	CMS2	1 x 10 ⁴	N	N	R
18	DLD1	MSI	CIMP+	CMS1	2.5 x 10 ³	N	-	-
19	FET	MSS	-	CMS2	1 x 10 ⁴	N	N	N
20	GEO	MSI	-	CMS2	1 x 10 ⁴	R	R	N
21	GP2D	MSI	CIMP-	CMS2	1 x 10 ⁴	R	R	R
22	GP5D	MSI	CIMP-	CMS2	1 x 10 ⁴	N	R	R
23	HCA46	MSS	CIMP-	CMS2	4 x 10 ⁴	N	N	R
24	HCA7	MSI	CIMP+	CMS1	7 x 10 ³	N	N	N
25	HCC2998	MSS	CIMP-	-	2 x 10 ³	N	-	-
26	HCT116	MSI	CIMP+	CMS1	2 x 10 ³	R	-	-
27	HCT15	MSI	CIMP+	-	7 x 10 ³	N	-	-
28	HCT8	MSI	CIMP+	CMS2	5 x 10 ³	N	-	-
29	HDC108	MSI	-	-	1 x 10 ⁴	R	-	-
30	HDC111	MSS	-	-	1.5 x 10 ⁴	R	-	-
31	HDC114	MSS	-	CMS1	7.5 x 10 ³	N	N	N
32	HDC133	MSS	-	-	2.5 x 10 ⁴	N	-	-
33	HDC15	MSS	-	-	1 x 10 ⁴	N	-	-
34	HDC54	MSS	CIMP-	CMS2	1.5 x 10 ⁴	N	R	R
35	HDC57	MSS	-	-	1.5 x 10 ⁴	R	N	N
36	HDC75	MSS	-	-	7.5 x 10 ³	N	N	N
37	HDC8	MSS	-	CMS4	1 x 10 ⁴	R	N	N

38	HDC87	MSS	CIMP-	-		N	-	-
39	HDC9	MSI	-	CMS3	1.5×10^4	N	N	N
40	HDC90	MSS	CIMP-	-	1×10^4	N	N	N
41	HRA19	MSS	CIMP-	CMS2	3×10^4	-	-	-
42	HT115	MSS	CIMP-	CMS3	1×10^4	N	N	N
43	HT29	MSS	CIMP+	CMS3	4×10^3	R	-	-
44	HT55	MSS	CIMP-	CMS2	1×10^4	N	N	R
45	HUTU80	MSS	-	CMS4	2×10^3	R	N	R
46	ISRECO1	MSS	-	CMS1	4×10^3	R	N	N
47	ISRECO2	MSS	-	CMS2	5×10^3	N	N	N
48	ISRECO3	MSS	CIMP-	CMS2	1×10^4	N	N	N
49	KM12	MSI	CIMP+	CMS1	2.8×10^4	N	-	-
50	LIM1215	MSI	CIMP-	CMS2	1.5×10^4	R	-	-
51	LIM1863	MSS	-	CMS2	2×10^4	N	N	R
52	LIM1899	MSI	CIMP-	-	1×10^4	N	N	N
53	LIM2099	MSS	CIMP+	CMS4	2×10^3	N	N	N
54	LIM2405	MSI	CIMP+	CMS1	1×10^4	R	-	-
55	LIM2537	MSI	-	CMS1	2×10^4	N	N	N
56	LIM2550	MSI	-	-	1.5×10^4	R	N	N
57	LIM2551	MSI	CIMP-	CMS1	1.5×10^4	R	N	N
58	LOVO	MSI	CIMP-	CMS1	1×10^4	N	N	R
59	LS1034	MSS	-	CMS2	5×10^3	N	N	N
60	LS174T	MSI	CIMP-	CMS3	1.4×10^4	N	-	-
61	LS180	MSI	CIMP-	CMS3	1×10^4	N	N	N
62	LS411	MSI	CIMP+	CMS1	6×10^3	N	N	R
63	LS513	MSS	CIMP+	CMS3	1×10^4	N	N	N
64	NCIH716	MSS	CIMP-	CMS4	3×10^4	N	N	N
65	NCIH747	MSS	CIMP-	CMS1	7.5×10^3	R	R	N
66	RKO	MSI	CIMP+	CMS1	2.5×10^3	R	-	-
67	RW2982	MSS	CIMP-	-	1×10^4	N	-	-
68	RW7213	MSS	CIMP-	-	1×10^4	R	-	-
69	SKCO1	MSS	CIMP+	CMS3	5×10^3	R	-	-
70	SNU175	MSI	CIMP-	CMS3	1.5×10^4	R	N	R
71	SNU283	MSS	CIMP-	CMS2	1.2×10^4	R	N	N
72	SNUC1	MSS	CIMP-	CMS2	7.5×10^3	R	R	R
73	SNUC2B	MSI	CIMP-	CMS1	1×10^4	N	N	N
74	SNUC4	MSI	CIMP-	CMS3	7.5×10^3	R	R	R
75	SW1116	MSS	CIMP-	CMS2	2×10^4	N	-	-
76	SW1222	MSS	CIMP-	CMS2	2.8×10^4	N	N	R
77	SW1463	MSS	CIMP-	CMS2	1.5×10^4	N	N	R
78	SW403	MSS	CIMP-	CMS2	2.4×10^4	N	R	N
79	SW48	MSI	CIMP+	CMS1	2×10^4	R	-	-
80	SW480	MSS	CIMP-	CMS4	2×10^4	R	-	-
81	SW620	MSS	CIMP-	-	5×10^3	N	-	-
82	SW837	MSS	CIMP+	CMS4	1.5×10^4	R	-	-
83	SW948	MSS	CIMP-	CMS2	1.6×10^4	N	-	-
84	T84	MSS	CIMP-	CMS2	1.5×10^4	N	-	-
85	TC71	MSI	-	CMS4	7.5×10^3	R	-	-
86	V9P	MSS	CIMP-	CMS2	1×10^4	N	N	N
87	VACO10	MSS	CIMP-	CMS2	1.2×10^4	N	R	R
88	VACO432	MSI	-	-	2×10^4	R	N	N
89	VACO4S	MSS	CIMP-	CMS2	1.5×10^4	N	R	R
90	VACO5	MSI	-	CMS1	1.2×10^4	R	N	R
91	WIDR	MSS	CIMP+	-	5×10^3	N	-	-

Note: ^{a)} MS status: microsatellite status, MSS: microsatellite stable, MSI: microsatellite instability; ^{b)} CIMP status: CpG island methylator phenotype (+: positive, -: negative); ^{c)} CMS: consensus molecular subtype; ^{d)} cclRECIST: cancer cell line response evaluation criteria in solid tumors; R: responders (complete response and partial response), NR: non-responders (progressive disease and stable disease).

3.1.2 Tumor samples

Formalin-fixed, paraffin-embedded primary colorectal cancer tumor samples were collected at collaborating medical institutions in Spain and Germany. Written informed consent for genetic analysis of the tumor sample was obtained from each patient. Sample collection was approved by the Human Investigations and Ethical Committee in the related Institution. All patients received irinotecan-based adjuvant chemotherapy after potentially curative colectomy. Tumor samples were selected from paraffin blocks containing a high percentage of tumor cells to generate tissue microarrays (TMA). A Beecher Instrument tissue arrayer (Silver Spring, MD) was used to array triplicate 0.6-mm cores from each sample in a fresh paraffin block.

The TMA contains tumor samples collected at three different medical institutions in Barcelona (n = 161), Madrid (n = 172) and Germany (n = 215).

3.1.3 Antibodies

Primary antibodies: Anti-SLC29A1, rabbit monoclonal antibody (Catalog: SAB5500117), was purchased from Merck Life Science (Madrid, Spain). Anti-GUCY2C, rabbit polyclonal antibody (Catalog: HPA037655), Anti-Vinculin, mouse monoclonal antibody (Catalog: V4505) and Anti- β -Tubulin, mouse monoclonal antibody (Catalog: T4026) were purchased from Sigma Life Science (St. Louis, MO, United States). Anti-GAPDH mouse monoclonal antibody (Catalog: sc-32233) was purchased from Santa Cruz Biotechnology (California, United States).

Secondary antibodies: Goat anti-Mouse Ig-HRP antibody (Catalog: P0447); Swine anti-Rabbit Ig-HRP antibody (Catalog: P0217) were purchased from Dako/Agilent (California, United States).

3.1.4 Primers

Table 4: Primers used in this study.

Primers' name	Target	Application	Sequence 5' - 3'
pENTRs (R)	SLC29A1	Sequencing	GTAACATCAGAGATTTTGAGACAC
pINDUCER20 (F)	SLC29A1	Sequencing	ACCTCCATAGAAGACACC
pcDNA3-T7 (F)	GUCY2C	Sequencing	TAATACGACTCACTATAGGG

Note: F= forward, R=reverse.

3.1.5 Plasmids

SLC29A1 (Homo sapiens) in pDONR221 (pDONR221-SLC29A1, Catalog: FLH183309.01L) was purchased from DNASU Plasmid Repository (Arizona, USA). The Lentiviral expression vector of GUCY2C (pHAGE-GUCY2C, Catalog: 116749) was purchased from Addgene (Watertown, MA, USA) (Ng et al., 2018). Tet-inducible lentiviral vector pINDUCER20 was a kind gift from Dr. Steve Elledge (Harvard Medical School, USA). pDONR221-SLC29A1 was verified by Sanger sequencing using the primers listed before. pHAGE-GUCY2C was verified by Sanger sequencing and diagnostic restriction digest using SspI and XhoI restriction enzymes. We conducted Gateway cloning to transiently overexpress SLC29A1 in TLA-HEK293T (also called HEK293) cells. The pDONR221-SLC29A1 plasmid was isolated using Miniprep Kit (Sigma-Aldrich) and verified by sequencing. The SLC29A1 coding sequences were transferred into the destination vector pINDUCER20 flanked by two attB sites (LR reaction) to overexpress the target gene. pHAGE-GUCY2C was generated by Addgene via Gateway cloning to overexpress the target gene.

3.1.6 Therapeutic Agents

Irinotecan (stock concentration: 20mg/mL; from Fresenius), 5-FU (stock concentration: 50mg/mL; from Accord), and oxaliplatin (stock concentration: 5mg/mL; from Teva) approved

by AEMPS (Agencia Española de Medicamentos y Productos Sanitarios) were obtained from Vall d'Hebron University Hospital (Barcelona, Spain).

3.1.7 OMICS Data

mRNA expression (RNA sequencing and microarray), proteomics, and whole-exome sequencing (WES) experiments were performed, and data were preprocessed in collaboration with Dr. John Mariadason (Olivia Newton-John Cancer Research Center, Australia) and Oliver Sieber (The Walter and Eliza Hall Institute of Medical Research, Australia). These data are referred to as “internal data”. External data includes RNAseq from Cancer Cell Line Encyclopedia (CCLE, Broad Institute, MA, USA), mRNA expression microarray from Dr. Alberto Bardelli's laboratory (AB, Medico et al., 2015), and proteomics data from Dr. Choudhary's laboratory (TR, Roumeliotis et al., 2017).

3.1.7.1 Internal datasets

- 1) **RNA sequencing.** Total RNA was extracted using TRIzol Reagent (Life Technologies, Carlsbad, CA). RNA sequencing analysis was processed at the Australian Genome Research Facility (AGRF) using an Illumina HiSeq 2000, yielding a depth of > 100 million paired-end reads per sample in 57 colorectal cancer cell lines. Paired reads were aligned to the version hg19 of the genome (hg19) with TopHat (Mouradov et al., 2014). In total, 25,369 transcripts were analyzed for 57 cell lines. This dataset is referred to as “RNAseq-1”.
- 2) **Microarray.** RNA was extracted using TRIzol Reagent (Life Technologies, Carlsbad, CA). The quantification of mRNA was performed using Affymetrix HG-U133 Plus 2.0 chips (Arango et al., 2005) in a subset of 42 colorectal cancer cell lines. The levels of transcripts, including 19,971 well-characterized genes, were included. Raw data were normalized by RMA (Robust Multichip Average) (Irizarry et al., 2003). This dataset is referred to as “Microarray-1”.

- 3) **Proteomics.** Liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based shotgun proteomic analyses to quantify protein levels on colorectal cancer cell lines were performed in collaboration with Oliver Sieber laboratory. Specifically, proteins were extracted from the cells and trypsin was used to digest proteins into peptides. Then the tryptic peptides were fractionated using off-line basic reversed-phase high-pressure liquid chromatography (bRPLC). Thermo Orbitrap-Velos mass spectrometer was used for fraction analysis by reversed-phase HPLC. Raw data were processed and used for database and spectral library searching using three different search engines, Myrimatch, Peptide and MS-GF+ (J. Wang et al., 2017). In total, 5,233 proteins in 47 colorectal cancer lines were identified. Relative protein abundance across the cell lines was used for further analysis (scaled intensities range: 0–1,000). This dataset is referred to as “Proteomics-1”.
- 4) **WES.** TruSeq Exome Enrichment Kit (Illumina) was used to perform whole-exome sequencing. One hundred bp paired-end read sequencing was performed on an Illumina HiSeq 2000 at AGRF. We detected variants using a modified GATK Best Practice protocol. Variants were filtered against known germline variations and detected in 114 in-house normal colorectal tissues. Regions of known germline chromosomal segmental duplications were also excluded. We analyzed 12,154 genes, and > 20-fold coverage, > 80% of targeted exons was achieved. The sequencing was done for 63 colorectal cancer lines (Mouradov et al., 2014). The mutational profile of each cell line was labeled as nominal variables for each gene: wild-type (WT) and mutant (MUT). A cell line was considered mutant for a gene when it presented at least one non-synonymous mutation, including insertion or deletion and single-nucleotide variant (SNV).

3.1.7.2 External datasets

- 5) **CCLC RNA sequencing.** External RNA sequencing data of 35 cell lines were obtained from the CCLC repository (<http://www.broadinstitute.org/cclc>). RNAseq analysis was performed on an Illumina HiSeq 2000 or HiSeq 2500, with coverage of 100 million paired reads per sample. The data were analyzed according to the iRAP pipeline. Quality-filtered reads were aligned to the reference human genome from Ensembl via

TopHat2. Raw counts (number of mapped reads summarized and aggregated over each gene) were generated using the htseq-count package in Python. The average expression of different probes mapping to the same gene was obtained (Barretina et al., 2012a). This dataset is referred to as “RNAseq-2”.

- 6) **AB microarray.** The external mRNA microarray data of 70 cell lines were obtained from Dr. Alberto Bardelli's laboratory (Candiolo Cancer Institute IRCCS, Candiolo, Torino). RNA was extracted using miRNeasy Mini Kit (Qiagen). The RNA quantification and quality analysis were performed on a Bioanalyzer 2100 (Agilent), using RNA 6000 nano Kit (Agilent). The synthesis of biotinylated cRNA was performed using the IlluminaTotalPrep RNA Amplification Kit (Ambion), using 500 ng of total RNA. Hybridization of cRNAs (750 ng) was carried out using Illumina Human 48 k gene chips (Human HT-12 V4 BeadChip). Array washing was performed using Illumina High Temp Wash Buffer for 10' at 55 °C, then staining using streptavidin-Cy3 dyes (Amersham Biosciences). Probe intensity data were obtained using the Illumina Genome Studio software (Genome Studio V2011.1). Raw data were analyzed after Loess normalization with the Lumi R package (Medico et al., 2015). This dataset is referred to as “Microarray-2”.
- 7) **TR proteomics.** The external proteomics data were obtained from Dr. Choudhary's laboratory (Wellcome Trust Sanger Institute, Cambridge, UK; PRIDE: PXD005235). Briefly, peptides were labeled with Tandem Mass Tag (TMT)10plex reagents and fractionated with high-pH C18 high-performance liquid chromatography (HPLC). LC-MS analysis was performed on Dionex Ultimate 3000 system with Orbitrap Fusion Mass Spectrometer. The quantification of 9,410 proteins was obtained for 50 cell lines by the sum of column-normalized TMT spectrum intensities and the row-mean scaling (Roumeliotis et al., 2017). This dataset is referred to as “Proteomics-2”.

3.1.8 Tissue microarrays (TMAs)

Sections of TMAs containing formalin-fixed paraffin-embedded tissue (FFPE) pellets of a panel of 77 colorectal cancer cell lines (all in triplicate) were stained with anti-SLC29A1 and anti-GUCY2C antibodies as described below. Relative protein expression from mass spectrometry shotgun proteomics for 59 of the cell lines in the TMAs was available for

comparison with the IHC expression levels. Finally, once the specificity of the antibodies was confirmed, sections of the TMAs containing triplicate tumor samples from a total of 548 patients treated with irinotecan were stained with the SLC29A1 and GUCY2C antibodies.

3.2 Methods

3.2.1 Growth Inhibition Assay

All colorectal cancer cell lines were maintained in their original culture medium without antibiotics/antimycotic according to the supplier guidelines after thawing out. When cells reached 90% - 100% of confluence, the supernatant was collected to test for mycoplasma contamination by PCR. Mycoplasma-positive cells received an anti-mycoplasma treatment using Plasmocin (25µg/mL, Invivogen, CA, USA) until they were negative. Mycoplasma negative cells were cultured with Minimum Essential Media (MEM) supplemented with 10% fetal bovine serum (FBS), antibiotic/antimycotic (100 U/ml streptomycin, 100 U/ml penicillin, and 0.25 µg/ml amphotericin B), MEM non-essential amino acids solution and 10mM HEPES buffer solution at 37°C and 5% CO₂. In the context of a previous project, we did a seeding test to determine the number of cells to be seeded in 96 well plates to ensure a cell density of approximately 80% confluence after the 96h experimental setting. Details of the number of cells seeded can be found in **Table 3**.

Sulforhodamine B (SRB) colorimetric assay was performed to characterize cell line sensitivity to chemotherapeutic reagents. SRB is an aminoxanthene dye that can bind to amino-acids (Horobin & Kiernan, 2020). On day 0, cells suspended in 100 µL of MEM medium were seeded in a 96-well plate. Twenty-four hours after plating, each compound dissolved in 100 µL MEM medium was added to the 96 well plates. The final concentrations used were as follows: irinotecan (0, 0.1, 0.5, 1, 2.5, 5, 10, 15, 20, 40, 80, and 100 µM), 5-FU (0, 0.01, 0.1, 0.5, 1, 2.5, 5, 10, 25, 50, 100, and 500 µM), and oxaliplatin (0, 0.05, 0.1, 0.5, 1, 5, 10, 25, 50, 100, 200, and 500 µM).

Besides, we set aside one plate containing 12 wells with only MEM media and 12 wells with cells in 100 μL MEM on day 0 (-24h). 24h after seeding, this plate was fixed with 50 μL trichloroacetic acid solution (TCA, final concentration 10%) for one hour at room temperature as control. Ninety-six hours after plating, all remaining plates were fixed with TCA. The plates were washed six times with slow-running tap water, and the excess water was removed using paper towels. Then, we allowed the plates to air-dry at room temperature. Wells were then stained with 50 μL SRB (0.4% in 1% acetic acid) for 30 minutes at room temperature, and then the plates were rinsed four times with 1% acetic acid to remove unbound dye. The plates were air-dried at room temperature. To measure the absorbance of the protein-bound dye, we added 200 μL of 10 mM Tris base solution (pH 10.5) to solubilize the SRB. The Optical Density at 510 nm (OD_{510}) was measured using a plate spectrophotometer (**Figure 19**) (Vichai & Kirtikara, 2006).

We calculated five parameters to build i dose-response curves and extract drug sensitivity metrics:

$T_i = \text{OD}_{510}$ of cells after 72h of treatment (i represents each drug concentration)

$T_0 = \text{OD}_{510}$ of cells before the beginning of drug treatment (24h after the seeding)

$C = \text{OD}_{510}$ of cells after 72h without treatment

$B_0 = \text{OD}_{510}$ of the medium before the treatment (24h after seeding); $B_i = \text{OD}_{510}$ of the medium 96h after the seeding day.

According to these measurements, we calculated the growth rate (GR) for each technical replicate, which is the ratio of growth rates with and without drug treatment.

After having the GR values, we normalized them as follows: 1) If GR was higher than 0 (cytostatic action), we calculated the percentage of viable cells compared to the non-treated cells after 72h of drug treatment (average of C); 2) If the GR was lower than 0 (cytotoxic action), the percentage of cell killing was calculated by comparing end-point cell density to that just before the treatment ($T_0 - B_0$), as shown below.

$GR_{(i)} > 0,$

$$GR_{(i)} = \frac{T_i - B_i - (T_0 - B_0)}{C - B_i - (T_0 - B_0)} / [C - B_i - (T_0 - B_0)]$$

$GR_{(i)} < 0$,

$$GR_{(i)} = \frac{T_i - B_i - (T_0 - B_0)}{C - B_i - (T_0 - B_0)} / (T_0 - B_0)$$

Dose-response curves were generated for each cell line, and the parameters IC₅₀, EC₅₀, Emax, AUC, and Hill slope (**Figure 20**), were extracted using GraphPad Prism 9 (GraphPad, CA, USA). These experiments were carried out at least three times to obtain a mean IC₅₀ value with a variance (SD/Mean) between experimental replicates below 25%.

3.2.2 Classification of cell lines according to drug response

In 2000, the Response Evaluation Criteria in Solid Tumors (RECIST) working group published the RECIST criteria, mainly using computed tomography (CT) and magnetic resonance imaging (MRI) to measure the size of target lesions selected for response assessment (Therasse et al., 2000). In the latest RECIST guidelines (version 1.1), revised in 2009, patients were stratified into four categories: progressive disease (PD), stable disease (SD), partial response (PR), and complete response (CR) depending on the specific tumor growth (Eisenhauer et al., 2009).

Based on RECIST 1.1, we stratified the cell lines according to criteria adapted from RECIST. We called this new cell line adapted response evaluation criteria (cclRECIST), and it was done by calculating the ratio of cell density on each dose and T0 cell density (T₀ - B₀). All cell lines were classified into four groups according to the following criteria:

Complete Response (CR): Total lack of cells at the endpoint (T_i).

Partial response (PR): At least a 30% decrease in the cell density compared to the treatment time point (T₀).

Stable Disease (SD): Decrease in the cell density compared to the treatment time point (T_0) lower than 30% or increase in the cell density compared to time point 0 (T_0) lower than 20%.

Progression Disease (PD): at least a 20% increase in cell density compared to time point 0 (T_0).

CR and PR cell lines were regarded as responders, while SD and PD classes were grouped and classified as non-responders for the subsequent analysis.

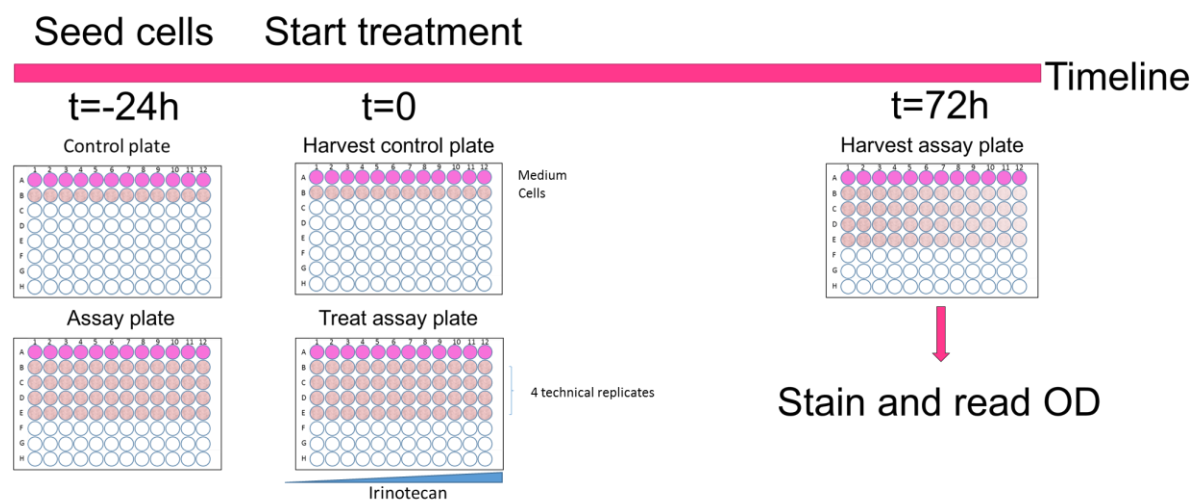


Figure 19: Schedule of the SRB method to characterize the drug sensitivity of cancer cell lines (irinotecan as the example).

Each colorectal cancer cell line was seeded in 96 well plates at the density specified in table 3. Twenty-four hours after plating, control cells were fixed with TCA, while the remaining cells were treated with the three compounds independently. Seventy-two hours after treatment, the cells were fixed with TCA. All plates were stained with SRB and solubilized with Tris, followed by OD measurement.

3.2.3 Functional group enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 was used to investigate whether there were gene sets with significant enrichment in the number of genes and proteins correlated with drug response. A Fisher's exact test was used to identify significantly enriched categories of genes associated with drug response rate. The results were deemed statistically significant if the multiple test corrected (FDR) p-value was less than 0.05 (D. W. Huang et al., 2009).

3.2.4 Removing batch effects in high-throughput experiments

Omics data produced by different studies are confounded by systematic error “batch effects”. ComBat is an algorithm based on the Empirical Bayes method to homogenize varied datasets (W. E. Johnson et al., 2007). To remove the batch effect between internal and external mRNA and proteomics data datasets, we performed ComBat correction using R language through the Gene Pattern (version 3.1) online platform (Reich et al., 2006).

3.2.5 Principle component analysis (PCA)

PCA is a statistical technique for reducing the dimensionality of a dataset while preserving the maximum amount of information. We used PCA via ClustVis (Metsalu & Vilo, 2015) (<https://biit.cs.ut.ee/clustvis/>) to visualize transcriptomics and proteomics, which contain a high number of dimensions. The first principal component (PC1) is the line that represents the maximum variance direction in the data. The second principal component (PC2) is oriented to reflect the second largest source of variation in the data while being orthogonal to the first PC.

3.2.6 Clustering analysis

Hierarchical clustering analysis is a general approach to cluster analysis in which the objects in a matrix (*e.g.*, a matrix containing mRNA or protein expression of a panel of cell lines) can be grouped together that have the shortest distance measures between objects ("close") to one another. The outcome is represented graphically as a dendrogram (S. C. Johnson, 1967).

Transcripts from 285 cell lines from four datasets were used for the transcriptomic analysis. Specifically, 25,369 transcripts of 57 cell lines were from in-house RNA sequencing, 19,971 transcripts across 42 cell lines were from internal microarray, 56,318 transcripts of 58 lines were from CCLE RNA sequencing, and 47,323 transcripts of 155 lines were publicly available and obtained using microarray platform (Medico et al., 2015).

In the combined dataset, 64 out of the 285 colorectal cancer cell lines were represented more than once. Detailed inspection of the hierarchical clustering analysis after normalization

revealed that 60 out of the 64 cell lines that were represented more than once clustered with at least one identical cell line from an independent gene expression database. Specifically, the outliers were SKCO1, CX1, SW948, RW2982, SW403, LIM2550, LIM2551, SNU175 from RNAseq-1, RW7213, SW1116, SW48, RW2982 from microarray-1, COLO201, DLD1, SW480, COLO320, HUTU80, RKO, NCIH747, SW1463 from RNAseq-2, and HUTU80, LIM1899, HCT116, and SW1463 from microarray-2.

Similar to the transcriptomics data processing, we merged liquid chromatography mass spectrometry-based proteomics from two datasets identified as Proteomics-1 and Proteomics-2. First, we conducted ComBat to remove the batch effect. The variability of PC1 and PC2 were 11% and 6% after conducting ComBat, and the cell lines were no longer separated based on the dataset of origin. Regarding proteomics data transformation, log10 outperformed log2 as 17 of 20 identical cell lines aligned well. Cell line outliers were T84, SKCO1, and SW948. After excluding these three cell lines, we re-transformed the proteomics data into the linear scale.

3.2.7 Transfection

TLA-HEK293T cells were seeded at the density of 1.0×10^6 /ml in the 150mm Petri dish to achieve approximately 75% confluence at the time of transfection. After 24 hours, cells were transfected with 25 μ g of plasmid mixture by incubating with 4 μ g of polyethylenimine (PEI, Polyscience, PA, USA) and 2,385 μ l of 150 mM NaCl for 15 minutes at room temperature. The plasmid mixture was composed of 22.5 μ g plasmid encoding the gene of interest and 2.5 μ g plasmid encoding the reporter protein EGFP (enhanced green fluorescent protein, pINDUCER20-EGFP). The later was used to determine transfection efficiency. After 48 hours of transfection, cells were imaged with a fluorescence microscope and harvested for Western blotting (WB) and immunohistochemistry (IHC), as described below. Doxycycline (25 μ g) was used to promote SLC29A1 expression in Tet-inducible pINDUCER20-SLC29A1.

3.2.8 Western blotting

For protein extractions, cell pellets were resuspended in RIPA buffer (0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, and protease inhibitors [Complete ULTRA Tablets, Mini, EDTA-free, Roche, Mannheim, Germany]). Cell lysates were incubated in orbital rotation for 10 minutes at 4 °C and then sonicated for 10 seconds. Samples were centrifuged for 15 minutes at 15,000 rpm at 4 °C, and supernatants were collected.

Western blotting was carried out using total protein extracts. BCA Protein Assay Kit (Catalog: 23227, ThermoFisher) was used for protein quantification. Twenty-five µg of total protein from cell lysate was loaded into the wells of the SDS-PAGE gel, along with a molecular weight marker, and a 120 V electric field was applied to allow the separation of proteins. Next, proteins were electrophoretically transferred from the gel to PVDF filters (Bio-Rad, Hercules, CA) for 2 hours at 100V and 4°C. Membranes were probed with the appropriate dilutions of primary and secondary antibodies. The antibodies concentration used in the study were: rabbit monoclonal anti-SLC29A1 antibody (1:1000), rabbit polyclonal anti-GUCY2C antibody (1/1,000), mouse monoclonal anti-β-tubulin antibody (1:10,000); mouse monoclonal anti-GAPDH antibody (1:5,000); mouse monoclonal anti-vinculin antibody (1:5000); Goat anti-Mouse Ig-HRP antibody (1:5000); Swine anti-Rabbit Ig-HRP antibody (1:10000). Bound antibodies were detected by chemiluminescence.

3.2.9 Immunohistochemistry

Immunohistochemical detection was used to determine the relative levels of protein expression in the tumor samples. First, to confirm the specificity and the sensitivity of the SLC29A1 and GUCY2C antibodies in FFPE samples, sections of control HEK293 cells and HEK293 cells with ectopic overexpression of the proteins of interest were stained with the antibodies as described below.

Cell pellets were harvested and fixed by 10% formalin overnight at room temperature. Then the cells in 10% formalin were centrifuged, and cell pellets were resuspended using the 6% low melting point agarose and transferred into a cylinder. When the agarose was solid, cylinders containing the cells were fixed with 10% formalin for 24 additional hours and then transferred

to 70% ethanol. Samples were then dehydrated in a tissue processor device and finally embedded in paraffin.

All paraffin blocks and TMAs were sectioned (4 µm) onto poly-L-Lysine-coated glass slides, and the slides were incubated at 80 °C for 1 hour, deparaffinized by immersion in xylene twice (10 minutes and 5 minutes) and hydrated through sequential immersion in decreasing concentrations of ethanol [100 % ethanol (2x5 minutes), 96% ethanol (5 minutes), 70% ethanol (5 minutes), 50% ethanol (5 minutes), and distilled water (10 minutes)]. Citrate buffer (10 mM, pH 6.0) was used for antigen retrieval. The slides were immersed in citrate buffer (pH 6, 10 mM) at 180°C for 4 minutes in a pressure cooker. After blocking with Leica blocking solution and washing with PBS to inactivate endogenous peroxidase twice, samples were incubated with primary antibodies overnight at 4°C (anti-SLC29A: 1/1000, 1/500, 1/100; anti-GUCY2C: 1/2000, 1/1000, 1/500, 1:100 were used for testing). Then, samples were incubated with the secondary antibody (Novolink polymer) for 30 min at room temperature. Then, the samples were washed with PBS and incubated with Novolink polymer (Leica) for 30 minutes at room temperature. Chromogen DAB (3,3-diaminobenzidine)/substrate reagent (Leica) was added to the slides to detect the secondary antibodies. Finally, slides were counterstained by Harry's hematoxylin (Quimica Clinica Aplicada S.A), dehydrated with ethanol and xylene, and mounted with DPX (dibutyl phthalate in xylene).

3.2.10 Quantification of the cell line tissue microarrays (TMAs)

QuPath bioimage analysis software was used to quantify the immunohistochemistry staining of cell line TMAs and compare it with LC-MS proteomic expression data across the cell line panel (Bankhead et al., 2017).

3.2.11 Patients grouping

We considered the period ranging from the treatment start to the death of patients as the overall survival (OS), while the length of time from treatment to patient progression as progression-free survival (PFS). Regarding the response to irinotecan, all patients from the three cohorts

were classified as progressive disease (PD), stable disease (SD), partial response (PR), and complete response (CR) according to RECIST 1.1 criteria. PD and SD were considered non-responders, while PR and CR were classified as responders.

3.2.12 Normalization of immunohistochemistry in patients' samples staining

To remove the batch effect associated with tissue collection and processing between the three patient cohorts, we normalized the immunohistochemistry staining score by dividing the score of each sample by a correction factor to remove the batch effect among the three cohorts. The correction factor was calculated by dividing the average immunohistochemistry staining score of each cohort by the average score of three cohorts.

3.2.13 Biomarker cutoff determination

To investigate the possible association between protein biomarker expression and the overall survival (OS) or progression-free survival (PFS) after irinotecan-based treatment for metastatic colorectal cancer, patients were dichotomized as having high or low tumor levels of protein expression using the expression value that maximized the overall survival differences between the two groups (lowest log-rank test P-value).

An online tool, the Cutoff finder (https://molpathoheidelberg.shinyapps.io/CutoffFinder_v1/), was used to optimize the cutoff of protein levels to dichotomize the patients into high-level (IHC+) and low-level (IHC-) groups that maximized the overall survival differences between the two groups (lowest log-rank test P-value) (Budczies et al., 2012).

3.2.14 Statistical analysis

Shapiro–Wilk test was used to test the normal distribution of samples.

Mann-Whitney U test was used to compare differences between two independent groups when variables were not normally distributed.

T-test was used to compare differences between two independent groups when variables were normally distributed. Welch's t-test was used when variances were not equal, and the Student t-test was used when variances were equal. F Test was used to compare two variances.

Fisher's exact test was used to determine where there were nonrandom associations between two categorical variables in 2×2 contingency tables. In contrast, the Chi-squared test (χ^2 test) was used to compare a greater number of variables in contingency tables.

Spearman's rank correlation coefficient was used to examine whether two non-normal distributed variables were correlated. Pearson correlation coefficient was used to examine whether two normally distributed variables were correlated.

Log-rank test was used to compare differences between the populations in the probability of an event (death or disease progression) at any time point.

Cox regression (or proportional hazards regression) was used for controlling for multiple confounders and evaluating the effect of several variables upon the time a specified event (death or disease progression) takes to happen.

Logistic regression was used to control for multiple confounders and evaluate the influence of several variables on patients outcome (RECIST).

Cox regression and logistic regression were conducted using SPSS 26. The other statistical analyses were conducted with GraphPad Prism 9.

4 RESULTS

4.1 Characterization of the sensitivity of colorectal cancer cell lines to chemotherapeutic agents

Human cancer cell lines are representative models for pre-clinical biomedical research. Indeed, colorectal cancer cell lines have been demonstrated to faithfully recapitulate the main subtypes of primary tumors at the genomic, transcriptional, and translational levels and have extensively been used to investigate colorectal cancer biology and drug response (Mouradov et al., 2014; J. Wang et al., 2017). Moreover, dose-response analyses of large cell line collections have been widely used to screen drug efficacy to identify novel biomarkers (Basu et al., 2013; Rees et al., 2016; Seashore-Ludlow et al., 2015).

The half maximal inhibitory concentration (IC_{50}) is the most common measure derived from dose-response curves of cell-to-cell and compound-to-compound (Fallahi-Sichani et al., 2013). Given that curves give additional information on sensitivity, four more parameters, EC_{50} , E_{max} , AUC, and Hill slope, can also be extracted from the dose-response curves. EC_{50} , half-maximal effective concentration, is the concentration that induces a response halfway between the baseline and maximum effect after the specified exposure time to the drug (Neubig et al., 2003). This parameter is highly dependent on the drug concentration range used in the experimental setting. IC_{50} is the drug concentration required to reduce cell growth by 50%. The only difference between IC_{50} and EC_{50} is how to define the 0%. For IC_{50} , 0% refers to the control value, the cell density before the treatment. In EC_{50} , 0% refers to the bottom plateau of the curve (**Figure 20a**). E_{max} represents the concentration at which the maximum effect of the drug is achieved. The area under the curve (AUC) measures the whole area underneath the dose-response curve. The Hill slope reflects the steepness of the slope observed in the exponential part of the curve (**Figure 20b**). Significant variability has been observed in the EC_{50} , IC_{50} , E_{max} , AUC, and Hill slope among cell lines.

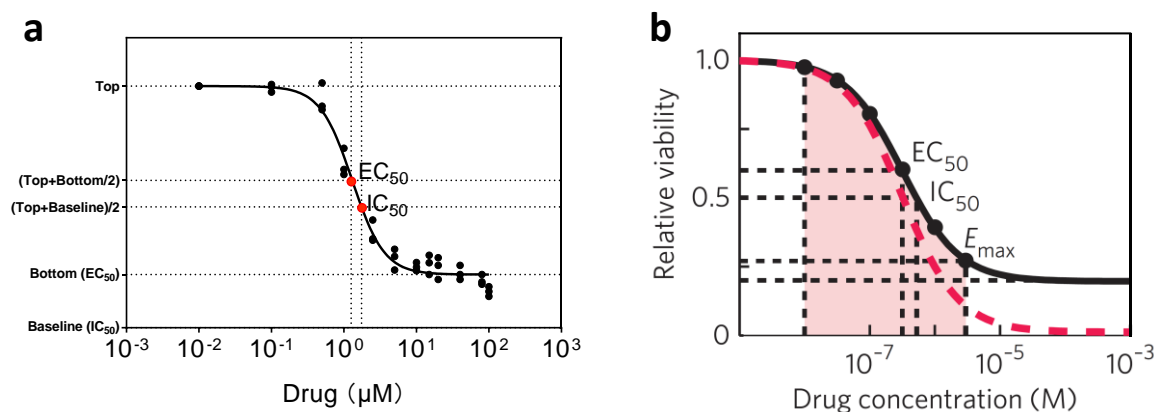


Figure 20: Dose-response curve and drug sensitivity parameters.

a) The graph compares the difference between IC_{50} and EC_{50} . **b)** Schematic of IC_{50} , EC_{50} , E_{max} , Hill slope, and AUC (pink area) extracted from the dose-response curve. The red dashed line indicates the case of $E_{max} = 0$, $EC_{50} = IC_{50}$ and Hill slope = 1. Figure modified from (Fallahi-Sichani et al., 2013).

Existing cell line drug response repositories. It has been reported that the drug sensitivity measurements between different high throughput cell line drug sensitivity screenings are not consistent (Haibe-Kains et al., 2013). The widely used cell line datasets are Cancer Cell Line Encyclopedia (CCLE), Cancer Therapeutics Response Portal (CTRP), and Genomics of Drug Sensitivity in Cancer (GDSC). The CCLE project provides a dataset to conduct a genetic characterization and response data of 138 anti-cancer drugs against a large panel of 727 human cancer cell lines (Barretina et al., 2012) (<https://sites.broadinstitute.org/ccle/>). The CTRP dataset measured 481 small-molecules sensitivity data to 224 cancer cell lines (Basu et al., 2013) (<https://portals.broadinstitute.org/ctrp.v2.1/>). The GDSC repository integrated molecular data of 11,289 tumors and 1,001 cell lines and measured the response profiling of 1,001 cancer cell lines to 265 anti-cancer drugs (Iorio et al., 2016) (<https://www.cancerrxgene.org/>). Of note, CCLE does not include the sensitivity data of 5-FU and oxaliplatin, and irinotecan is not available in CTRP. CCLE dataset has IC_{50} , EC_{50} , E_{max} , and AUC, while CTRP used E_{max} , EC_{50} , AUC, and Hill slope, and GDSC only measured IC_{50} and AUC. Therefore, the combination of comparison was different across the three drugs (**Table 5, Table 6 and Table 7**). To verify whether the sensitivity of 5-FU, oxaliplatin, and irinotecan in colorectal cancer cell lines across different datasets was consistent, we performed Pearson's (normal distribution) or Spearman's (non-normal distribution) correlation between the cell line datasets CCLE, CTRP, GDSC.

Table 5: Comparison of drug response measures from different datasets for 5-FU.

Sources	IC ₅₀	EC ₅₀	E _{max}	AUC	Hill Slope
Internal	√	√	√	√	√
CCLE	NA	NA	NA	NA	NA
CTRP	NA	√	√	√	√
GDSC	√	NA	NA	√	NA

NA: not applicable

Table 6: Comparison of drug response measures from different datasets for oxaliplatin.

Sources	IC ₅₀	EC ₅₀	E _{max}	AUC	Hill Slope
Internal	√	√	√	√	√
CCLE	NA	NA	NA	NA	NA
CTRP	NA	√	√	√	√
GDSC	√	NA	NA	√	NA

NA: not applicable

Table 7: Comparison of drug response measures from different datasets for irinotecan.

Sources	IC ₅₀	EC ₅₀	Emax	AUC	Hill Slope
Internal	√	√	√	√	√
CCLE	√	√	√	√	NA
CTRP	NA	NA	NA	NA	NA
GDSC	√	NA	NA	√	NA

NA: not applicable

For 5-FU, AUC from CTRP and GDSC showed a significant correlation ($R = 0.37$, $P\text{-value} = 0.027$) based on the 36 overlapping cell lines (**Figure 21**). No other measures were available from these two datasets.

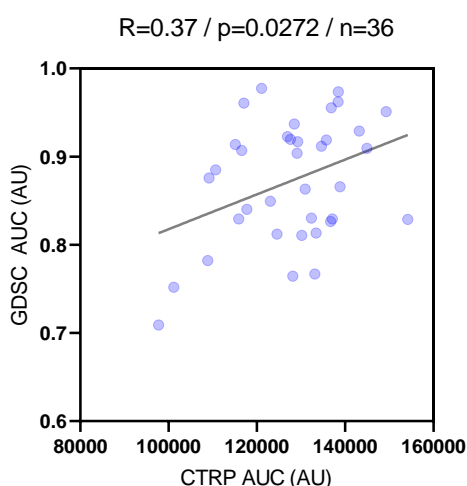


Figure 21: Comparison of 5-FU sensitivity measured in overlapping cell lines between the pharmacologic datasets.

CTRP: Cancer Therapeutics Response Portal. GDSC: Genomics of Drug Sensitivity in Cancer. R: correlation coefficient. P: P-value. n: the number of overlapping cell lines across the cell line datasets compared.

Then we conducted the correlation of oxaliplatin sensitivity from different datasets. Nevertheless, the AUC of GDSC and CTRP did not show a positive correlation ($R = 0.17$, $P\text{-value} = 0.3234$) (**Figure 22**) across the 37 overlapping cell lines.

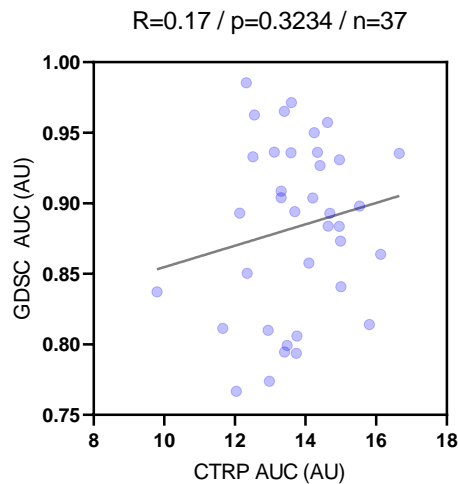


Figure 22: Comparison of oxaliplatin sensitivity measured in overlapping cell lines between the pharmacologic datasets.

CTRP: Cancer Therapeutics Response Portal. GDSC: Genomics of Drug Sensitivity in Cancer. R: correlation coefficient. P: P-value. n: the number of overlapping cell lines across the cell line datasets compared.

Next, we conducted the correlation of irinotecan sensitivity data among different public datasets. The IC_{50} and AUC from GDSC and CCLE did not show a significant association for the 9 overlapping cell lines (IC_{50} : R = 0.67, P-value = 0.0589; AUC: R = -0.66, P-value = 0.0533) (**Figure 23**).

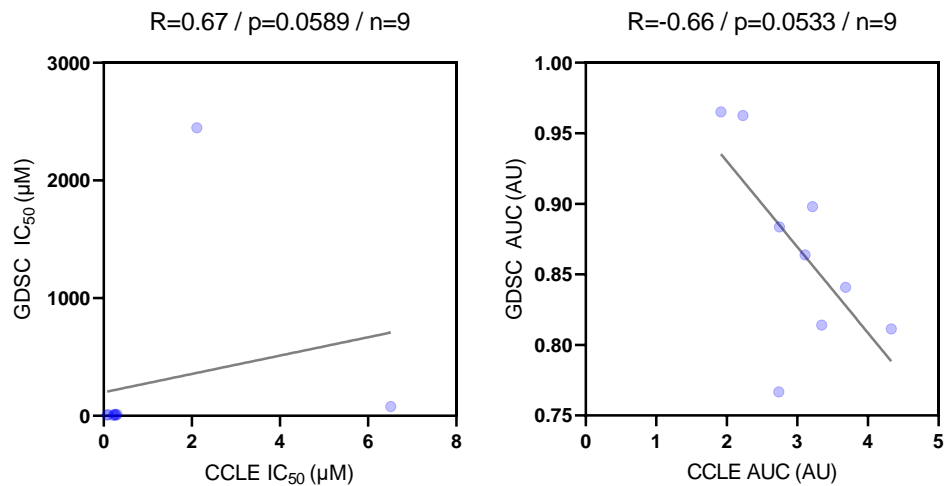


Figure 23: Comparison of irinotecan sensitivity measured in overlapping cell lines between the pharmacologic datasets.

GDSC: Genomics of Drug Sensitivity in Cancer. CCLE: Cancer Cell Line Encyclopedia. R: correlation coefficient. P: P-value. n: the number of overlapping cell lines across the cell line datasets compared.

Collectively, based on the sensitivity of 5-FU, oxaliplatin, and irinotecan, we did not find a high correlation between the different datasets interrogated. Given that the repositories used different strategies and a range of concentrations in the dose-response screening, it is not easy to have a consistent result across datasets. In addition, since the common cell line repositories were not targeted at a specific disease, they all have a limited number of cell lines on a particular subtype of cancer. Therefore, we decided to conduct the drug sensitivity screening in a large panel of colorectal cancer cell lines using a rigorous method of dose-response exposure and analysis.

Here, we used a panel of colorectal cancer cell lines ($n = 90$) to screen their sensitivity to the main chemotherapeutic agents used for colorectal cancer patients. The sensitivity data of 57 cell lines to 5-FU and oxaliplatin and 90 cell lines for irinotecan was determined. Sulforhodamine B (SRB) colorimetric assay was performed to characterize cell line sensitivity to chemotherapeutic reagents. This method is widely used in drug screens using cell lines and uses the protein content (*i.e.*, SRB staining) as a surrogate marker for cell number (Shoemaker, 2006; Vichai & Kirtikara, 2006). The cytotoxic effect was assessed using 12 different concentrations of these three drugs in 96-well plates. Twenty-four hours after seeding, we started to treat the cells were treated for 72 hours and determination of protein content amounts as a surrogate measurement of cell numbers. As explained in more detail in the “Materials and Methods” section, the relative sensitivity of the different cell lines can be summarized using different measurements from the resulting dose-response sigmoid curves.

4.1.1 - Sensitivity of colorectal cancer cell lines to 5-FU

For 5-FU, 57 colorectal cancer cell lines were included in the sensitivity screening. The range and fold change of IC_{50} , EC_{50} , E_{max} , AUC and Hill slope was 0.50 to 146.59 μ M (293.18-fold), 0.04 to 148.74 μ M (3718.50-fold), -91.43 to 29.39 %, 542.9 to 943.9 AU (arbitrary units) (1.74-fold), and -0.22 to -2.37 AU (10.77-fold), respectively (**Figure 24**).

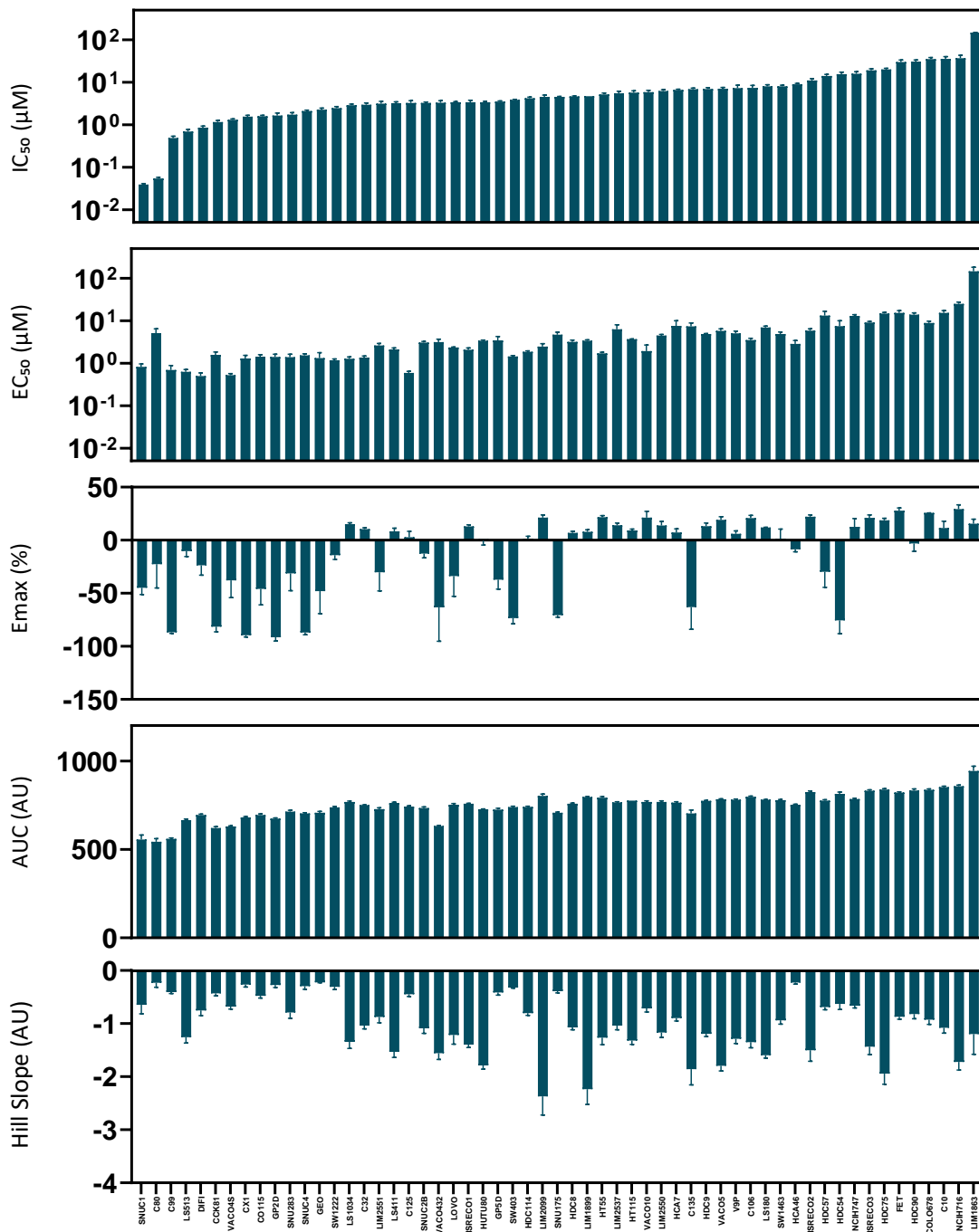


Figure 24: 5-FU sensitivity metrics for the colorectal cancer cell lines panel.

Histograms showing the values of IC₅₀, EC₅₀, E_{max}, AUC, and Hill slope for each cell line. AU, arbitrary units. Columns show mean \pm SEM from at least three independent experiments.

4.1.2 - Sensitivity of colorectal cancer cell lines to oxaliplatin

For oxaliplatin, we used 57 colorectal cancer cell lines for the screening. The range and fold change of IC_{50} , EC_{50} , E_{max} , AUC and Hill slope was 0.01 to 8.03 μM (803.00-fold), 0.02 to 28.38 μM (1419.00-fold), 2.94 to -97.19 %, 273.1 to 569.8 AU (2.09-fold), and -0.42 to -3.2 AU (7.62-fold), respectively (**Figure 25**).

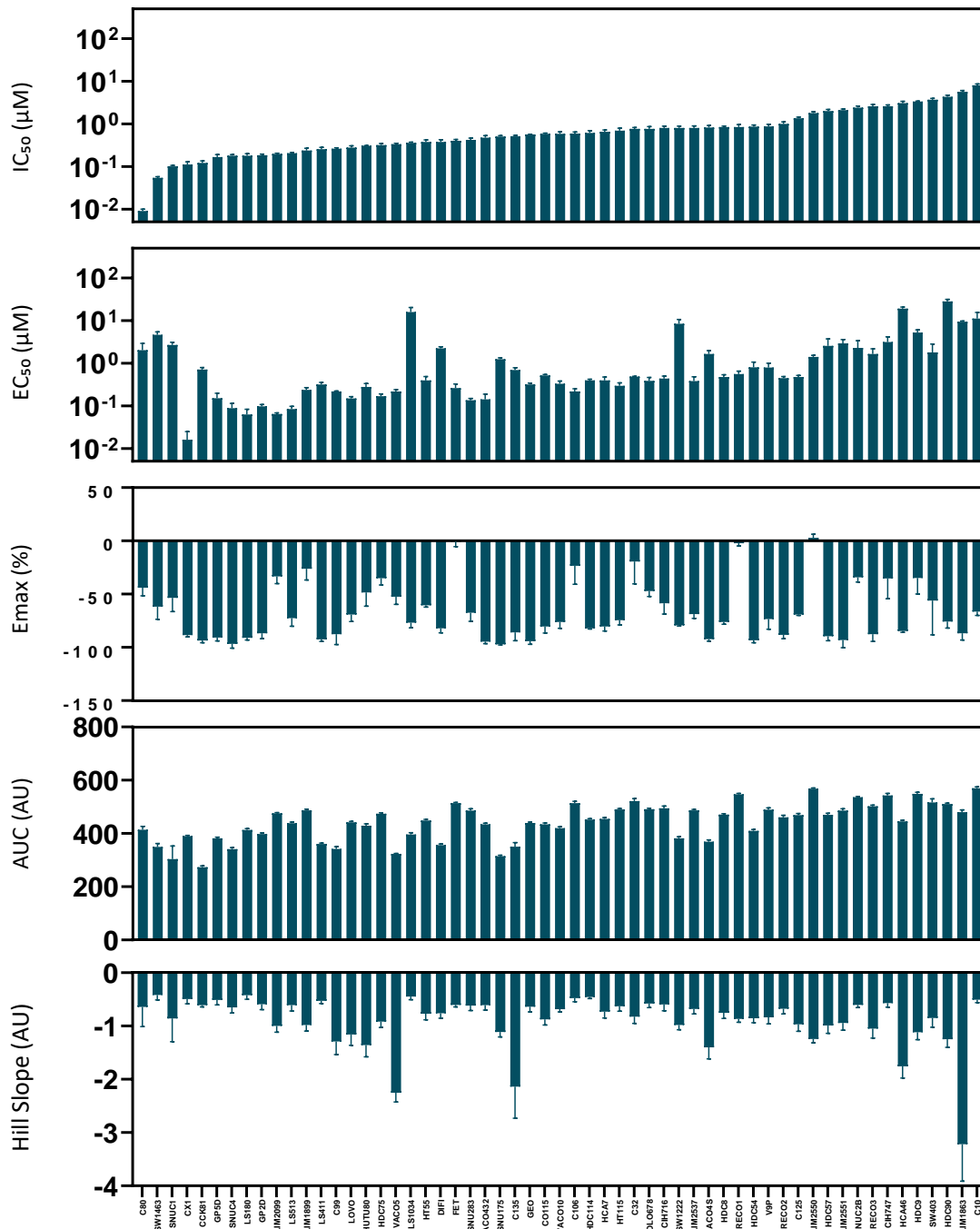


Figure 25: Oxaliplatin sensitivity metrics for the colorectal cancer cell lines panel.

Histograms showing the values of IC₅₀, EC₅₀, Emax, AUC, and Hill slope for each cell line. Columns show mean +/- SEM from at least 3 independent experiments.

4.1.3 - Sensitivity of colorectal cancer cell lines to irinotecan

Regarding irinotecan, 90 colorectal cancer cell lines were included in the screening. The EC₅₀ varied from 0.20 to 121.80 μ M, with a 614-fold change. IC₅₀ ranged from 0.01 to 73.16 μ M, giving a difference of 7316-fold. Emax varied from -103.31 to 29.60 %. AUC had values ranging from 247.9 to 585.5 AU. Hill slope showed a variance from -3.85 to -0.15 AU (**Figure 26**).

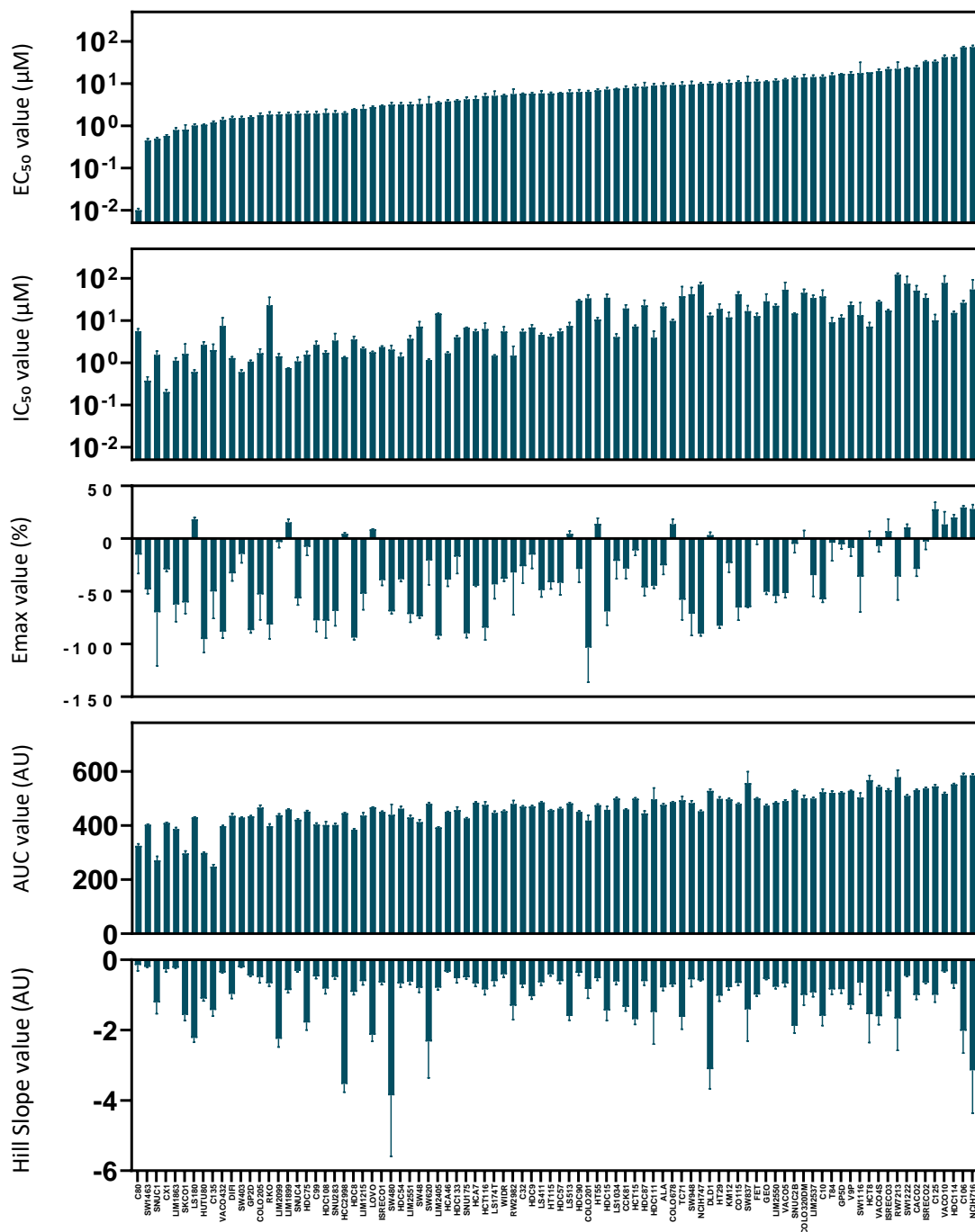


Figure 26: Irinotecan sensitivity metrics for the panel of colorectal cancer cell lines. Histograms showing the values of EC₅₀, IC₅₀, Emax, AUC, and Hill slope for each cell line. Columns show mean \pm SEM from at least independent experiments.

Comparison of the sensitivity data generated with the existing data in public repositories.

Then we carried out a comparison of the sensitivity data generated (internal) and the sensitivity data available in different high throughput public repositories described before (external datasets). For 5-FU, AUC of internal, CTRP and GDSC showed a significant correlation

between them. IC₅₀ from internal and GDSC showed a high correlation. Other comparisons did not show a significant positive correlation (**Figure 27**).

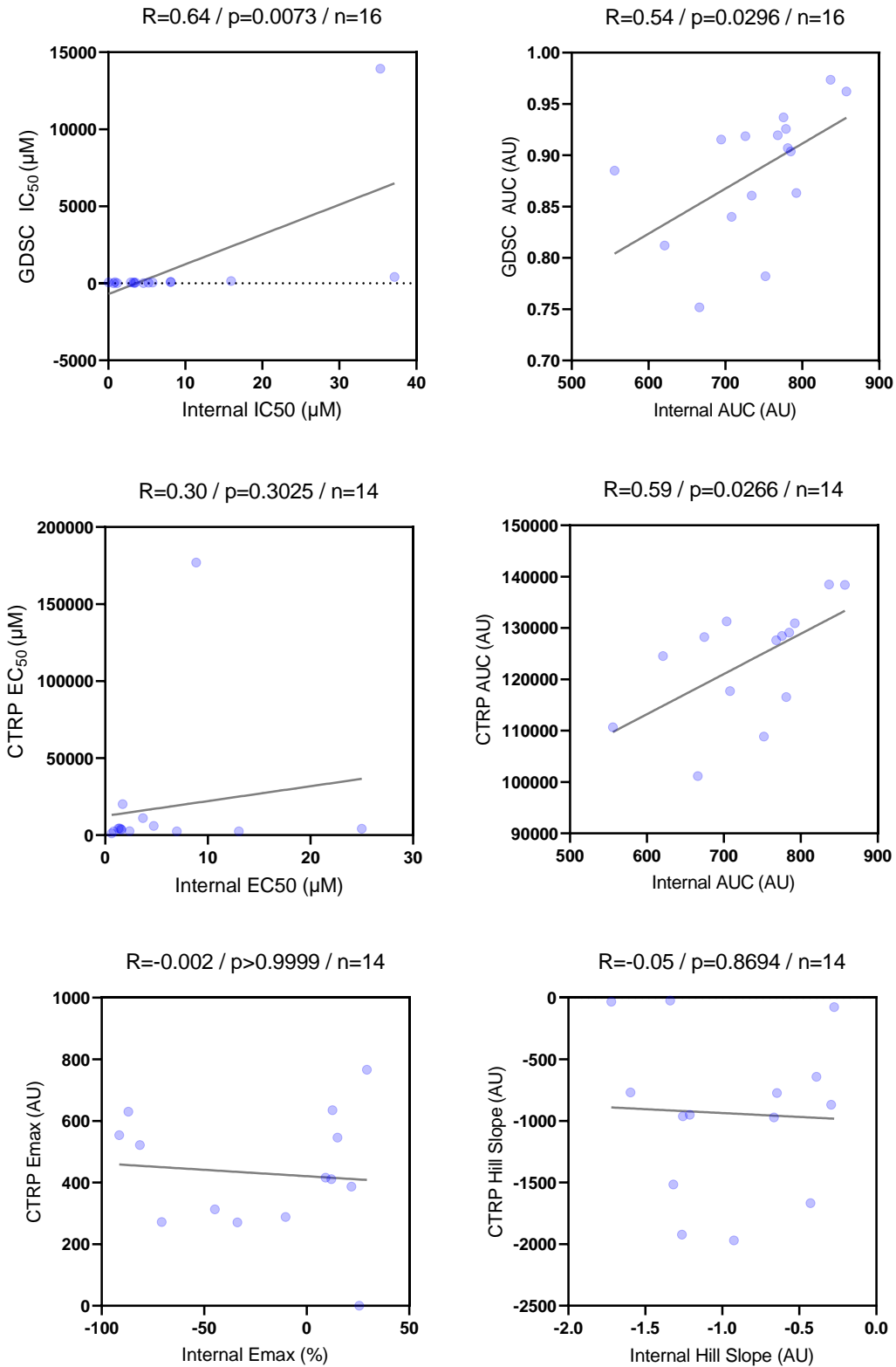


Figure 27: Comparison of 5-FU sensitivity measured in overlapping cell lines between different datasets. GDSC: Genomics of Drug Sensitivity in Cancer. CTRP: Cancer Therapeutics Response Portal. R: correlation coefficient. P: P-value. n: number of overlapping cell lines across cell line datasets.

For oxaliplatin, the only significant correlation observed was the IC₅₀ between internal and GDSC. Other comparisons did not show a positive correlation (**Figure 28**).

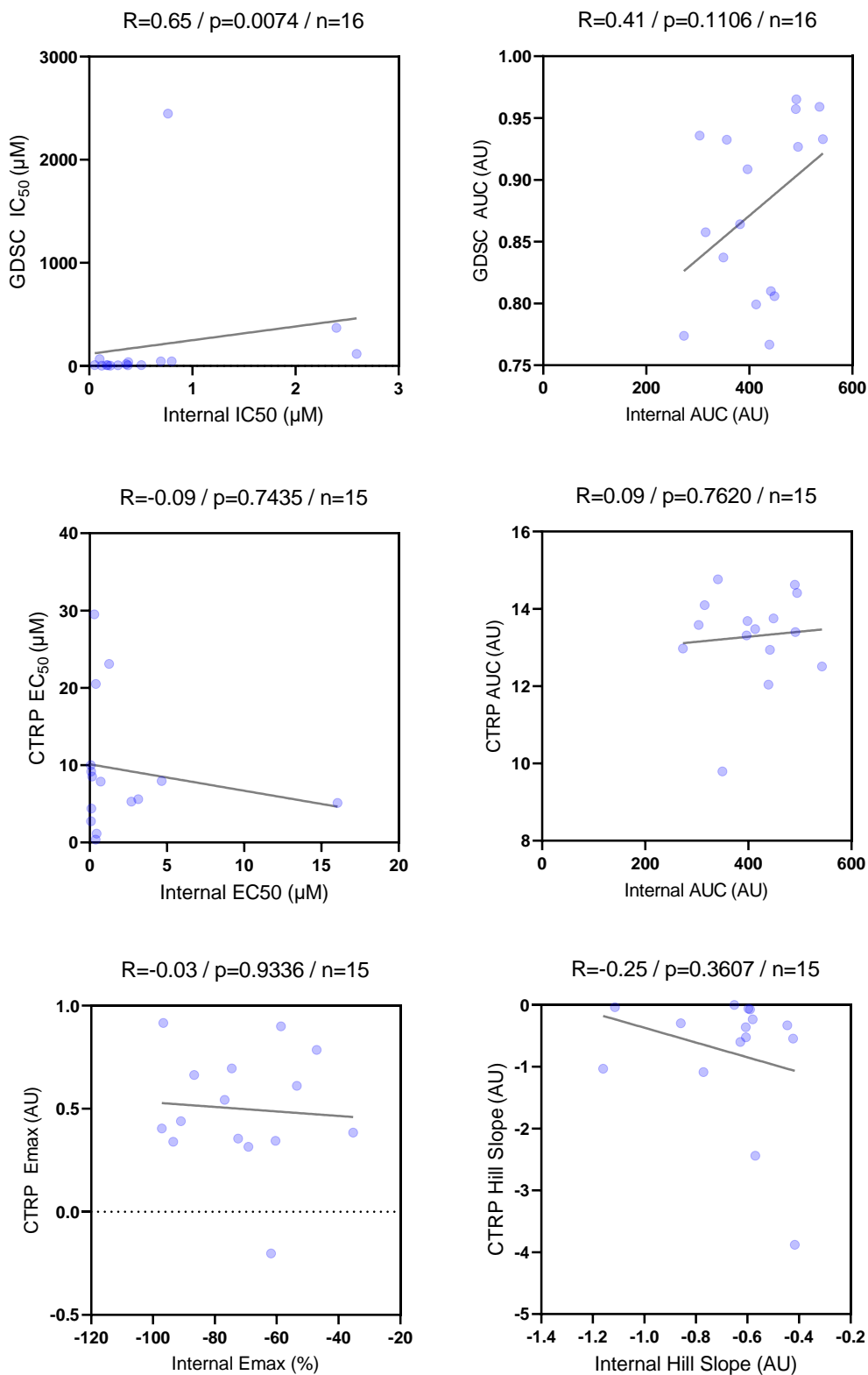


Figure 28: Comparison of oxaliplatin sensitivity measured in overlapping cell lines between the different datasets.

GDSC: Genomics of Drug Sensitivity in Cancer. CTRP: Cancer Therapeutics Response Portal. R: correlation coefficient. P: P-value. n: number of overlapping cell lines across cell line datasets.

For irinotecan, most measures did not show a significant correlation between the internal data and the public sensitivity data. Only IC_{50} from internal screening and GDSC and Emax of internal and CCLE showed a significant positive correlation (**Figure 29**).

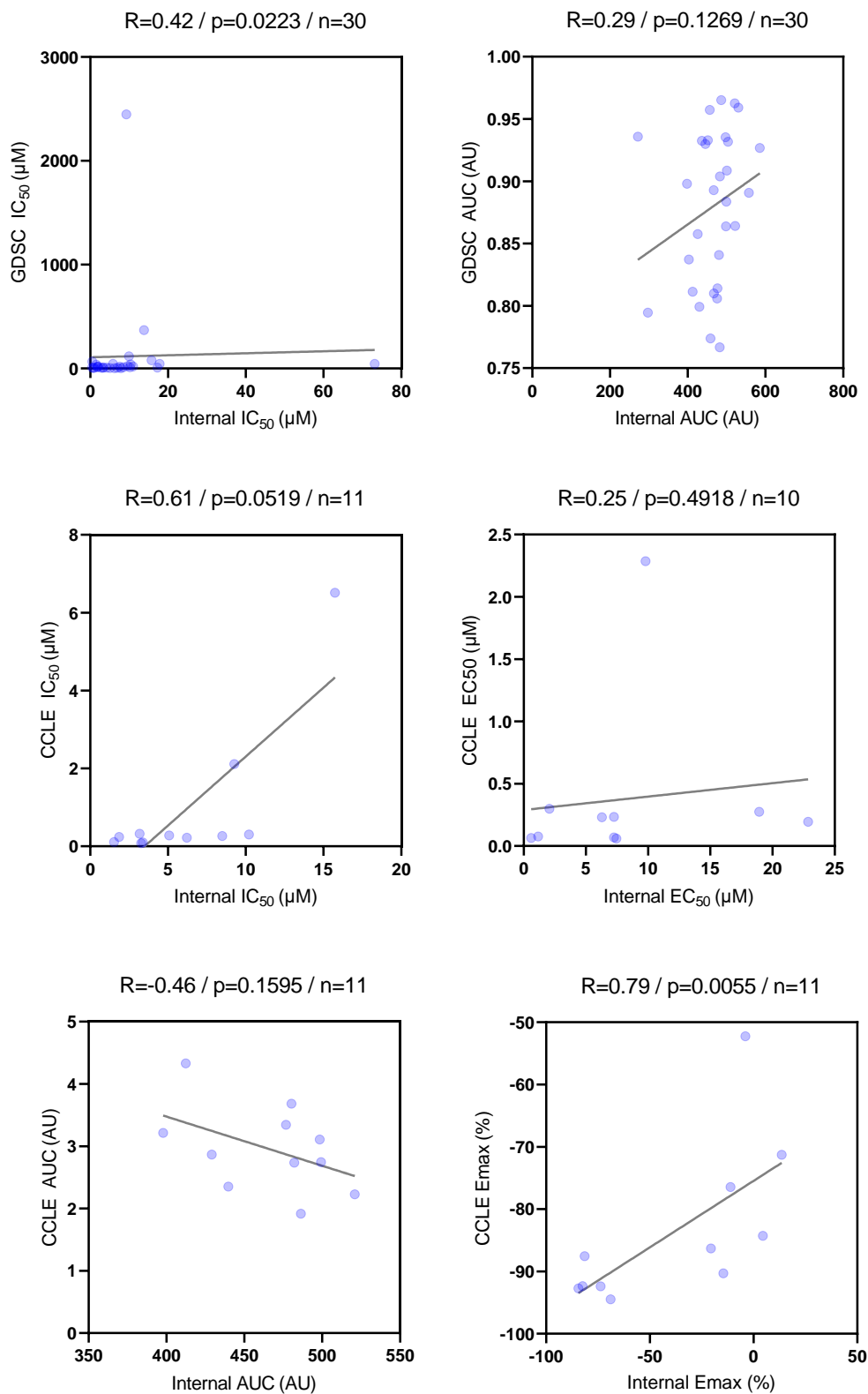


Figure 29: Comparison of irinotecan sensitivity measured in overlapping cell lines between the different datasets.

GDSC: Genomics of Drug Sensitivity in Cancer. CCLLE: Cancer Cell Line Encyclopedia. R: correlation coefficient. P: P-value. n: number of overlapping cell lines across cell line datasets.

From the above results, we showed that although some measures between our data and public datasets were statistically associated, inconsistency was common, which hampers to combining the sensitivity data from different sources.

4.1.4 - Comparison of the different metrics of drug sensitivity

To assess the consistency between the different metrics summarizing the sensitivity of the cell lines to the agents of interest, we next conducted a Spearman's correlation between these five parameters extracted from the dose-response curves, EC_{50} , IC_{50} , E_{max} , AUC and Hill slope. For 5-FU, all metrics used showed a significant correlation with each other. For oxaliplatin, AUC showed a significant correlation with IC_{50} and E_{max} , IC_{50} displayed a significant correlation with EC_{50} , and Hill slope showed a significant correlation with IC_{50} . Other combinations did not show a significant correlation. Regarding irinotecan, except for E_{max} and EC_{50} , Hill slope and EC_{50} , all metrics significantly correlated with each other (**Figure 30**). This observation was in agreement with previous studies, which indicate different metrics convey various and complementary drug response information (Fallahi-Sichani et al., 2013).

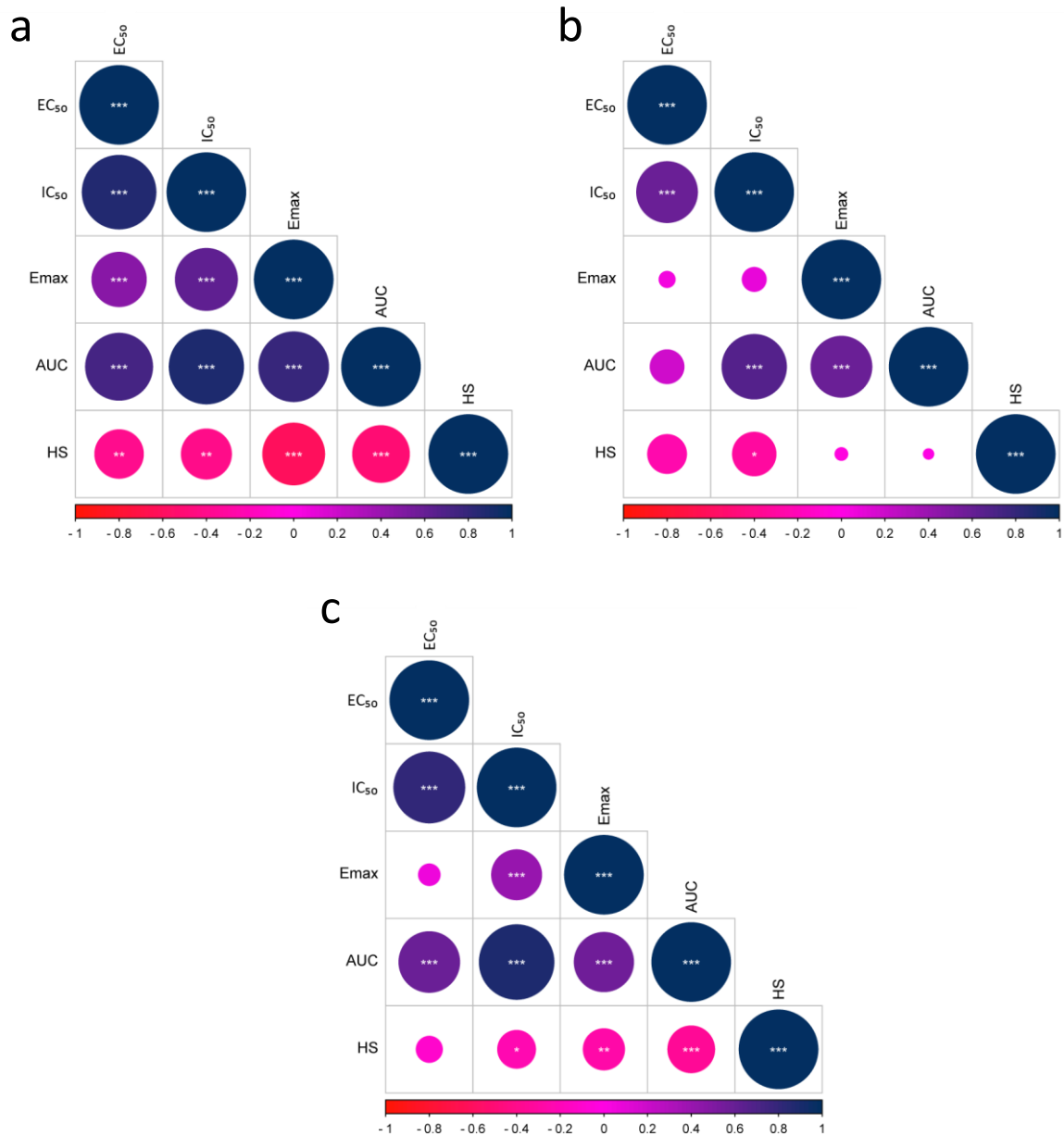


Figure 30: Spearman's correlation coefficients of the five sensitivity metrics.

Spearman's correlation coefficients were calculated between the EC₅₀, IC₅₀, Emax, AUC, and Hill slope (HS) for 5-FU (a), oxaliplatin (b), and irinotecan (c). Spearman's correlation coefficients are shown. ns, P > 0.05; * P ≤ 0.05; ** P ≤ 0.01; *** P ≤ 0.001; **** P ≤ 0.0001.

4.1.5 Classification of colon cancer cell lines into two clinically relevant categories: responders and non-responders

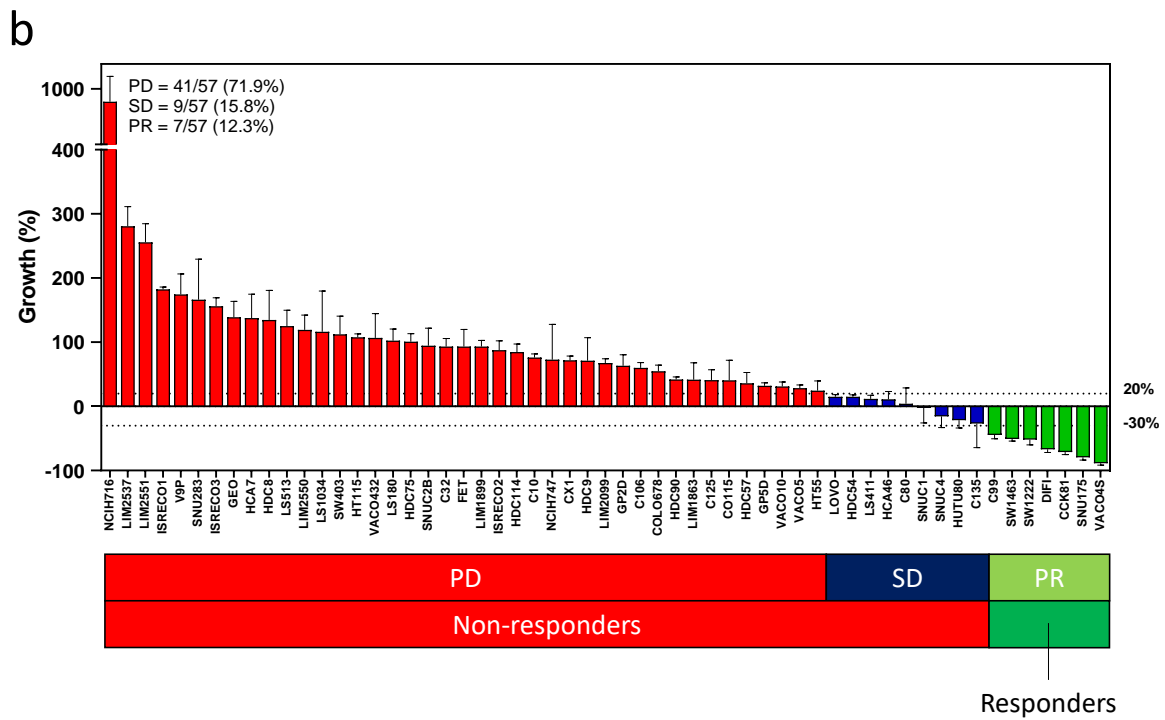
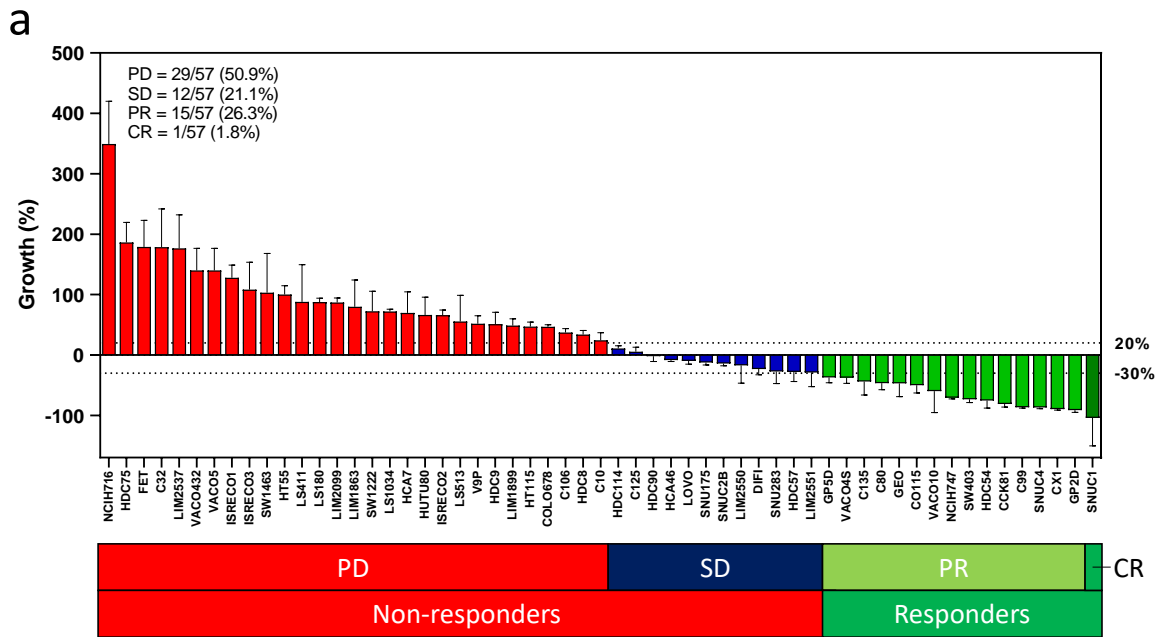
The response of primary colorectal tumors to a chemotherapeutic agent is often used as a surrogate for treatment decision-making. In 2000, the Response Evaluation Criteria in Solid

Tumors (RECIST) working group published the RECIST criteria, mainly using computed tomography (CT) and magnetic resonance imaging (MRI) to measure the size of target lesions selected for response assessment (Therasse et al., 2000). In the latest revised RECIST guidelines (version 1.1) revised in 2009, patients are stratified into four categories: progressive disease (PD), stable disease (SD), partial response (PR), and complete response (CR) depending on the specific tumor growth (Eisenhauer et al., 2009).

The primary goal of our study was to find biomarkers to predict the chemotherapy response of primary tumors in a clinical setting, and thus we aim to provide information regarding which group of patients can benefit from the specific treatment and who will not. In addition, we proposed to apply cell line models and bioinformatic approaches to identify biomarkers. Therefore, it was important to find a way to categorize the cell lines into responders and non-responders. Considering that there are five sensitivity metrics, the discrepancies between them and that these metrics are all continuous variables, it was difficult to objectively choose the best cutoff to dichotomize the cell lines into responders and non-responders. We proposed a method adapted from the clinic to stratify the cell lines: cancer cell line adapted Response Evaluation Criteria in Solid Tumors (ccIRECIST), the standard clinical criteria to evaluate the response to the treatment.

Unlike patients, which are given an optimal dose of the compound, cells were treated with 11 different concentrations of each drug. Therefore, we aimed to select the doses that more closely resemble the response rates observed clinically. The response rates of 5-FU, oxaliplatin, and irinotecan observed in the treatment of metastatic colorectal tumors are about 16-30% (Cunningham et al., 2009), 10-20% (Raymond, Chaney, et al., 1998), and 16-41% (Rougier et al., 1997), respectively. After comparing the proportion of responder and non-responder cell lines among the 11 concentrations in the cell line panel, we selected 500 μ M for 5-FU, 10 μ M for oxaliplatin, and 80 μ M for irinotecan to stratify cell lines into clinically relevant categories (**Figure 31 Appendix 1**).

The objective response rate (ORR), defined as the proportion of responders (CR and PR) out of all treated patients, is an essential criterion for evaluating the therapeutic effect in the clinic (Eisenhauer et al., 2009; Oxnard et al., 2016). Therefore, PR and CR cell lines were classified as responders and PD and SD as non-responders. In the following analysis, all results are based on comparing non-responder and responder cell lines.



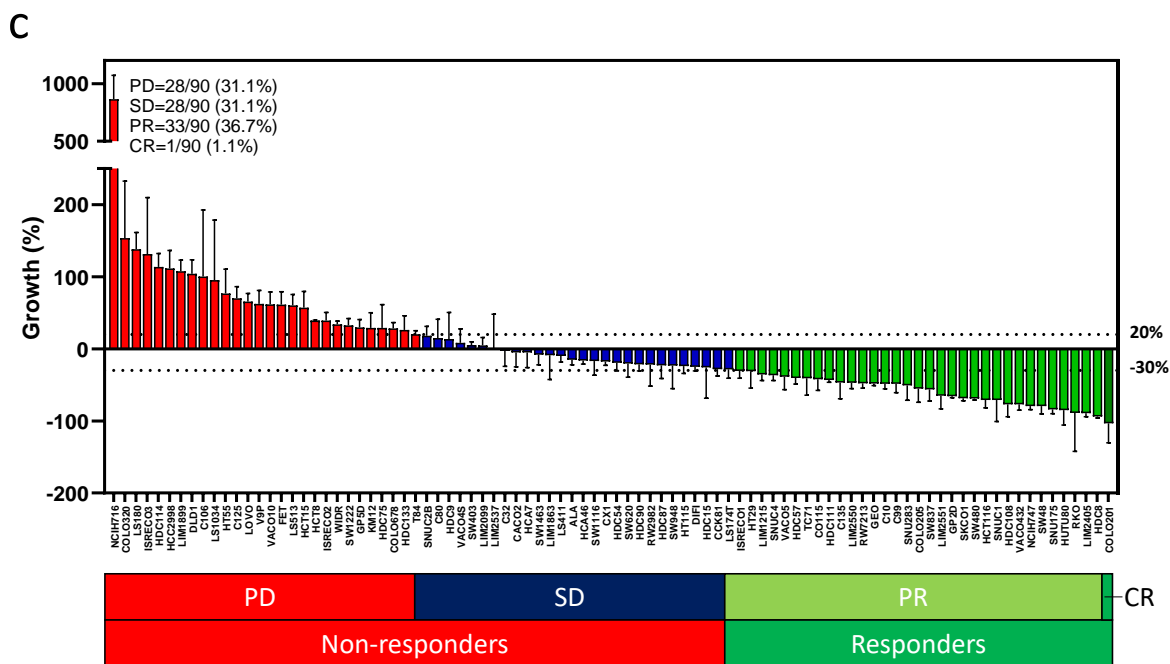
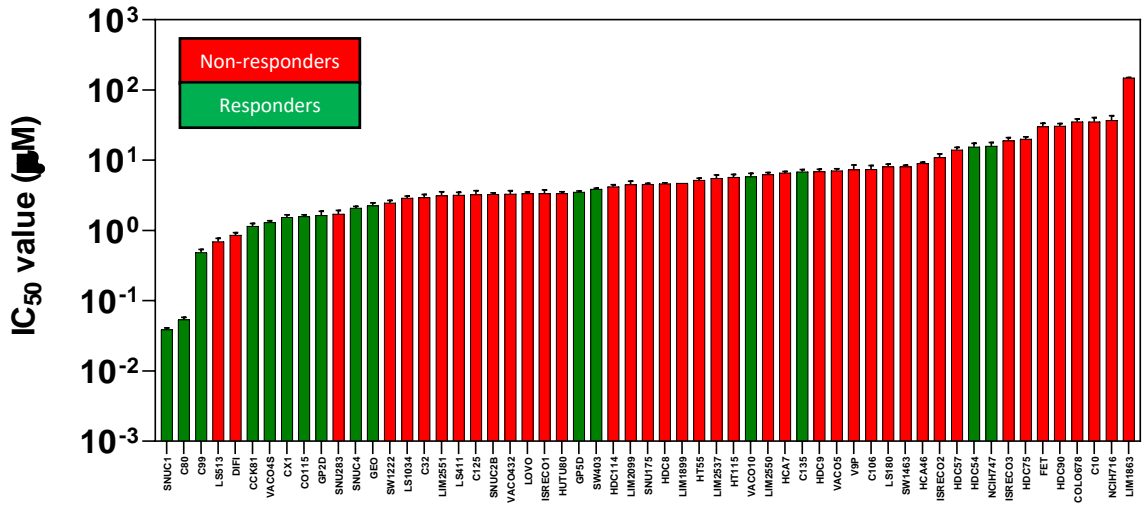


Figure 31: Classification of the sensitivity of colorectal cancer lines to 5-FU, oxaliplatin, and irinotecan according to cancer cell line Response Evaluation Criteria in Solid Tumors (ccRECIST).

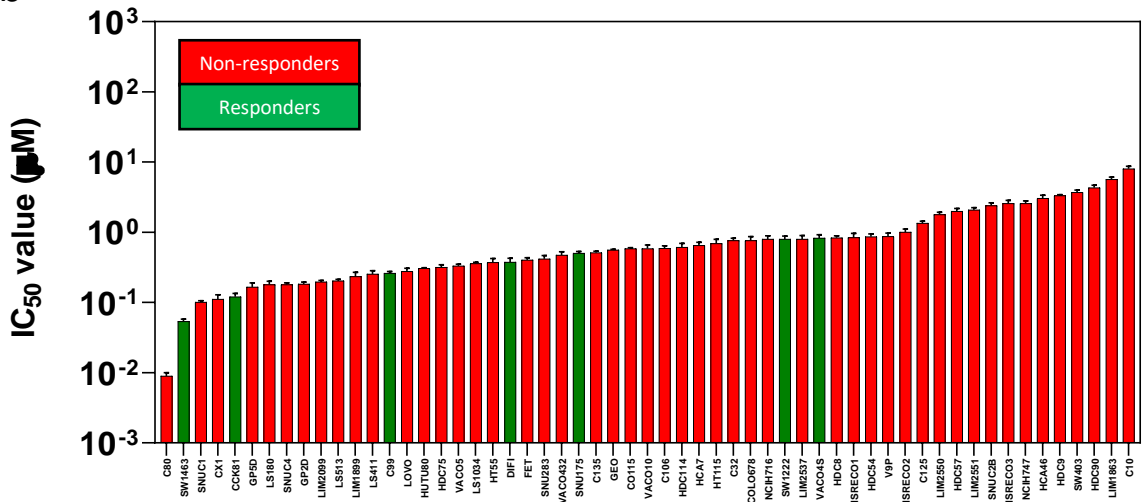
The histogram shows the percentage of growth of the cell lines upon 72-hour treatment with 5-FU (a) at 500 μ M, oxaliplatin (b) at 10 μ M and irinotecan (c) at 80 μ M. The stratification in different classes was done with an adaptation of RECIST criteria. The color boxes below the graph represent the ccRECIST categories as well as the percentage of cell lines classified into responders and non-responders. Columns show the mean (\pm SEM) from at least three independent experiments. PD, progressive disease, SD, stable disease, PR, partial response, CR, complete response.

After classifying all cell lines into responders and non-responders by ccRECIST, we evaluated how the classification was associated with the sensitivity to the drug summarized as the IC_{50} , one of the most common sensitivity measures (**Figure 32**).

a



b



C

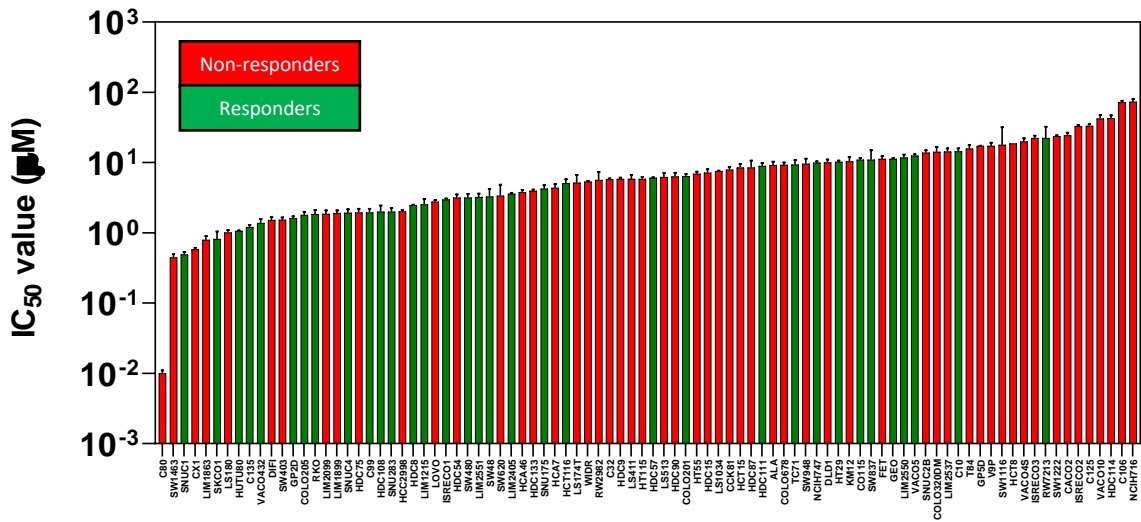


Figure 32: Distribution of cell line categorized by cancer cell line Response Evaluation Criteria (ccIRECIST) and IC₅₀.

a) 5-FU, b) oxaliplatin, c) irinotecan. Cell lines were ranked according to the IC₅₀ values. Green bars indicate the cell lines defined as responders by ccIRECIST, while red bars illustrate non-responders.

ccIRECIST was significantly associated with IC₅₀ for 5-FU (P-value = 0.002) and irinotecan (P-value = 0.016). Non-responder cell lines had a higher IC₅₀ compared with responder cell lines. For oxaliplatin, Non-responders also tended to have higher IC₅₀ values than responder cell lines, but the difference was not significant (P-value = 0.150) (**Figure 33**).

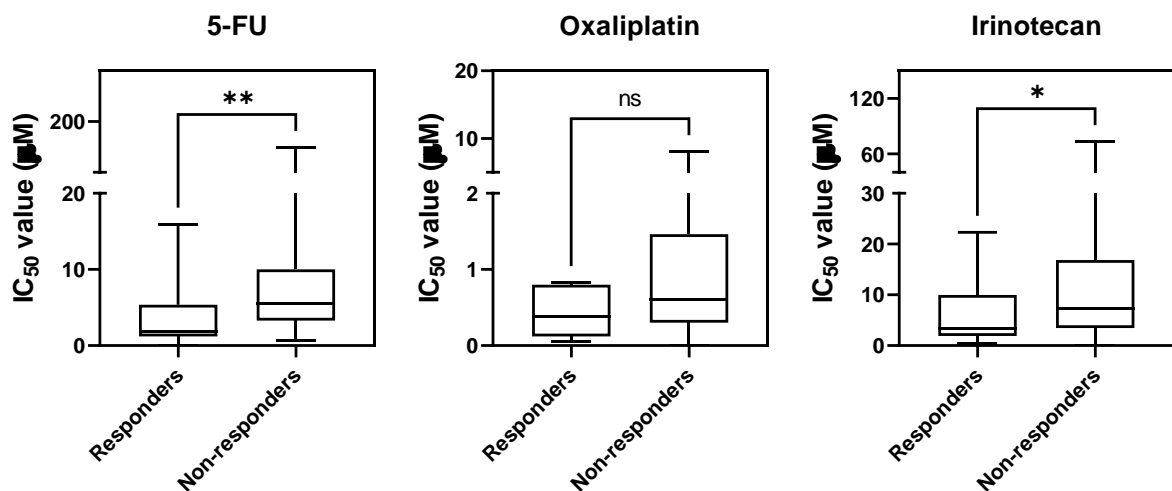


Figure 33: Comparison of cancer cell line Response Evaluation Criteria (ccIRECIST) and IC₅₀.

a) 5-FU, b) oxaliplatin, c) irinotecan. ns, P > 0.05; * P ≤ 0.05; ** P ≤ 0.01.

4.2 Association of drug sensitivity with colorectal cancer molecular subtypes

After stratifying all cell lines into responders and non-responders, we investigated the association of sensitivity with known molecular subtypes of colorectal cancer, including CpG island methylator phenotype (CIMP) status, microsatellite status (MSI and MSS), and Consensus Molecular Subtypes (CMS), as previously explained in the introduction.

4.2.1 - Colorectal cancer molecular subtypes and response to 5-FU

CpG island methylator phenotype (CIMP) status was available for 40 of the 57 (70.2%) cell lines with 5-FU sensitivity data. CIMP refers to the hypermethylation of the CpG dinucleotides in the CpG islands present in the promoter of genes, a critical event in the early stages of the development of a subset of colorectal tumors. Fisher's exact test showed that there was no significant difference in response rate between CIMP+ and CIMP- colorectal cancer cell lines (**Figure 34a**).

Microsatellite instability status was available for all 57 (100.0%) cell lines with 5-FU sensitivity data. Microsatellites are short, tandemly repeated sequences of DNA that change in length (show instability) when mismatch repair mechanisms are deficient, thus more prone to accumulate mutations and drive the tumorigenic process in a subset of colorectal tumors. Cell lines with microsatellite instability (MSI) tended to have a higher response rate compared to microsatellite stable (MSS) cell lines, although this difference did not reach statistical significance ($P = 0.054$; **Figure 34b**).

Furthermore, Guinney and his collaborators identified a gene expression-based colorectal cancer classification Consensus molecular subgroups (CMS) (Guinney et al., 2015) from patient samples, which includes four subtypes: CMS1 (microsatellite instability immune), CMS2 (canonical), CMS3 (metabolic), and CMS4 (mesenchymal). CMS classification was available for 51 cell lines (89.5%). There was no significant difference in response rate among CMS subtypes (**Figure 34c**).

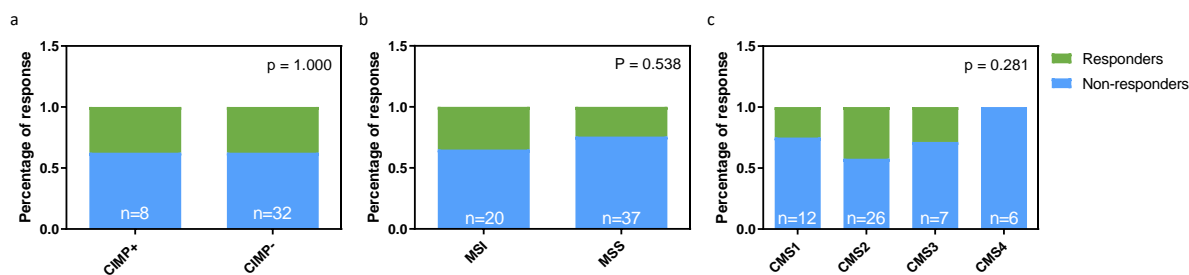


Figure 34: Association of 5-FU response with known colorectal cancer molecular subtypes.

Histograms show the percentage of cell lines belonging to the responder or non-responder group among the lines characterized as either CIMP– or CIMP+ (a), MSI or MSS (b), and CMS subtypes (c). N is the number of cell lines included in each category. Each graph shows the P-value from Fisher’s exact test analysis of the distribution of responders and non-responders groups. CIMP, CpG island methylator phenotype; MSI, microsatellite instable; MSS, microsatellite stable; CMS, consensus molecular subtypes.

4.2.2 - Colorectal cancer molecular subtypes and response to oxaliplatin

To interrogate the association of oxaliplatin sensitivity with molecular subtypes, we used 40 (70.2%) cell lines for which we had both, the CIMP status information and drug sensitivity data. Fisher’s exact test showed no significant difference in response to irinotecan between cell lines with CIMP+ and CIMP- (**Figure 35a**). We next explored the association of oxaliplatin sensitivity and microsatellite status in 57 (100.0%) cell lines. No significant difference in the response rate to the agent was found between cell lines with MSI and MSS (**Figure 35b**). The comparison of CMS subtypes and oxaliplatin sensitivity in 51 (89.5%) cell lines resulted in no significant difference in the percentage of response rate to the drug between the four independent CMS subtypes (**Figure 35c**).

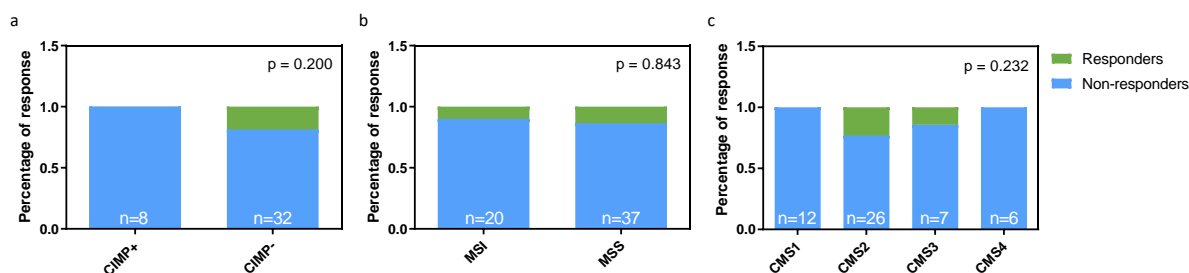


Figure 35: Association of oxaliplatin response with known colorectal cancer molecular subtypes.

Histograms show the percentage of cell lines belonging to the responder or non-responder group among the lines characterized as either CIMP– or CIMP+ (a), MSI or MSS (b), and CMS subtypes (c). N is the number of cell lines included in each category. Each graph shows the P-value from Fisher’s exact test analysis of the distribution of responders and non-responders groups. CIMP, CpG island methylator phenotype; MSI, microsatellite instable; MSS, microsatellite stable; CMS, consensus molecular subtypes.

4.2.3 - Colorectal cancer molecular subtypes and response to irinotecan

CIMP status information was available for 67 of the 91 (73.6%) cell lines with irinotecan sensitivity data. In cell lines with CIMP+, responders and non-responders accounted for 50%, respectively, while responders and non-responders accounted for 27% and 73% in the cell lines with CIMP-. The proportion of responders was higher in CIMP+ compared to CIMP- cell lines, but the difference was not significant (P-value = 0.099). This finding was similar to a published cohort study, including 615 colorectal cancer patients (Shiovitz et al., 2014). Shiovitz et al. showed that CIMP+ colorectal cancer patients had a trend towards improved overall survival when treated with adjuvant IFL regimen (irinotecan, 5-FU, and leucovorin) compared to FU/LV (5-FU and leucovorin; hazard ratio = 0.62; P-value = 0.07). However, this observation was not found in CIMP- colorectal cancer patients (hazard ratio = 1.38; P-value = 0.049; **Figure 36a**).

Regarding microsatellite status in the 89 (97.8%) available cell lines with irinotecan sensitivity, the proportion of responders was significantly higher in the MSI subgroup (53.1%) than in MSS (29.3%) (P-value = 0.040; **Figure 36b**). This finding was consistent with the results of a large cohort study of 702 colorectal cancer patients, where MSI was found to predict improved 5-year disease-free survival (DFS) to irinotecan plus fluorouracil/leucovorin regimen (IFL) (hazard ratio = 0.53; P-value = 0.03) compared with MSS. In addition, consistent with the results presented in the previous section for 5-FU, MSI status did not predict response in patients treated with fluorouracil and leucovorin regimen (FU/LV) (hazard ratio = 1.07; P-value = 0.80) (Bertagnolli et al., 2009).

Although the above clinical studies used survival rate while we used response rate to conduct the analysis, the response rate closely mirrors the survival rate (Blumenthal et al., 2015). The consistent findings between our cell line models and the primary tumor samples indicate that: 1) cell lines used in this study are good models to reflect the clinical settings; 2) the adapted cclRECIST is a promising method to stratify cell lines according to their drug sensitivity.

We compared the responders and non-responders frequency among different subtypes of CMS in the 71 (78.0%) available cell lines. The proportion of responders in CMS1, CMS2, CMS3 and CMS4 were 60.0%, 18.8%, 40.0%, and 66.7%, respectively (**Figure 36c**). Consistent with

the finding that MSI tumors and cell lines have a higher sensitivity to irinotecan, CMS1 (microsatellite instability immune) showed a significantly higher proportion of sensitive lines compared to CMS2 (canonical) (P-value = 0.003). CMS2 showed a significantly higher proportion of resistant lines compared to the CMS4 group (P value = 0.012). There was no significant difference in response rate between other CMS subtypes.

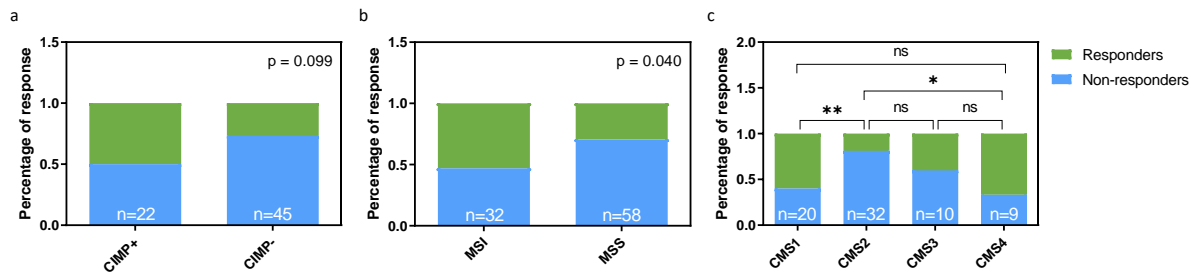


Figure 36: Association of irinotecan response with known colorectal cancer molecular subtypes.

Histograms show the percentage of cell lines belonging to the responder or non-responder group among the lines characterized as either CIMP– or CIMP+ (a), MSI or MSS (b), and CMS subtypes (d). N is the number of cell lines included in each category. Each graph shows the P-value from Fisher’s exact test analysis of the distribution of responders and non-responders groups. CIMP, CpG island methylator phenotype; MSI, microsatellite instable; MSS, microsatellite stable; CMS, consensus molecular subtypes.

4.3 Association between drug sensitivity and mutations in “driver” genes

As explained before, cancer driver genes are those whose mutations significantly contribute to the tumorigenic process (J et al., 2016). Here, the possible association between drug response and mutations in six widely recognized colorectal cancer driver genes (*APC*, *KRAS*, *BRAF*, *PIK3CA*, *SMAD4* and *TP53*) was investigated (D. Huang et al., 2018; Smit et al., 2020).

4.3.1 - Driver gene mutations and sensitivity to 5-FU

We first explored the association of driver gene mutations with the sensitivity to 5-FU. No difference was observed in the sensitivity of colorectal cancer cell lines with or without mutations in *APC*, *KRAS*, *BRAF*, *PIK3CA*, *SMAD4* and *TP53* (**Figure 37**).

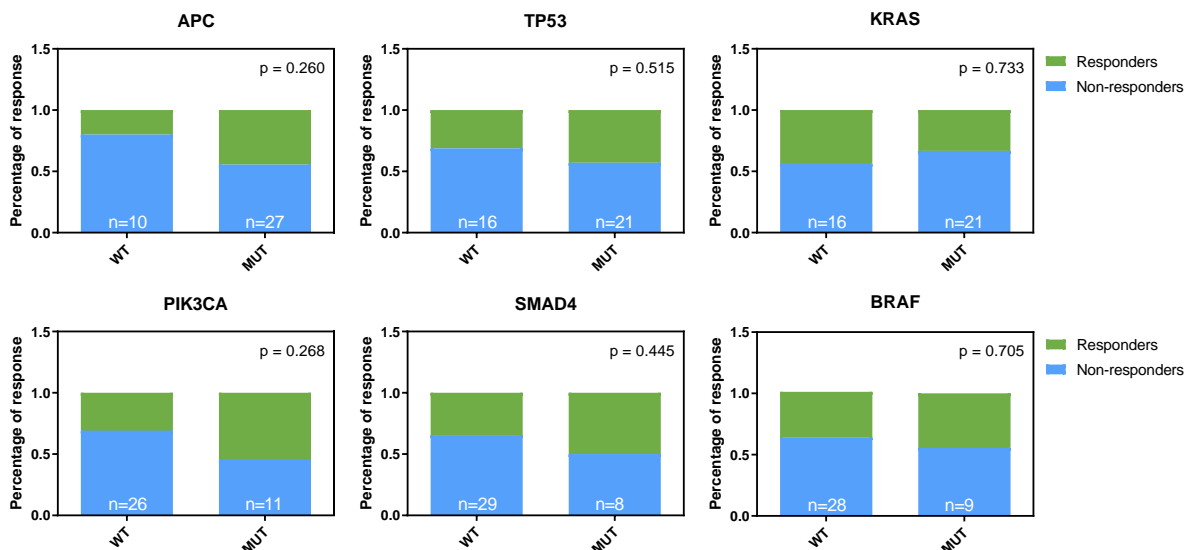


Figure 37: Association of mutations in colorectal cancer driver genes with the 5-FU response. Graphs show the proportion of responder or non-responder among the cell lines without (wild-type, WT) or with (mutant, MUT) mutations in *APC*, *TP53*, *KRAS*, *PIK3CA*, *SMAD4*, and *BRAF*. Each graph shows the P-value from Fisher's exact test analysis.

4.3.2 - Driver gene mutations and sensitivity to oxaliplatin

For oxaliplatin, there was no significant difference in response rate between cell lines that are wild-type or have mutations in *APC*, *TP53*, *KRAS*, *PIK3CA*, *SMAD4*, and *BRAF* (**Figure 38**).

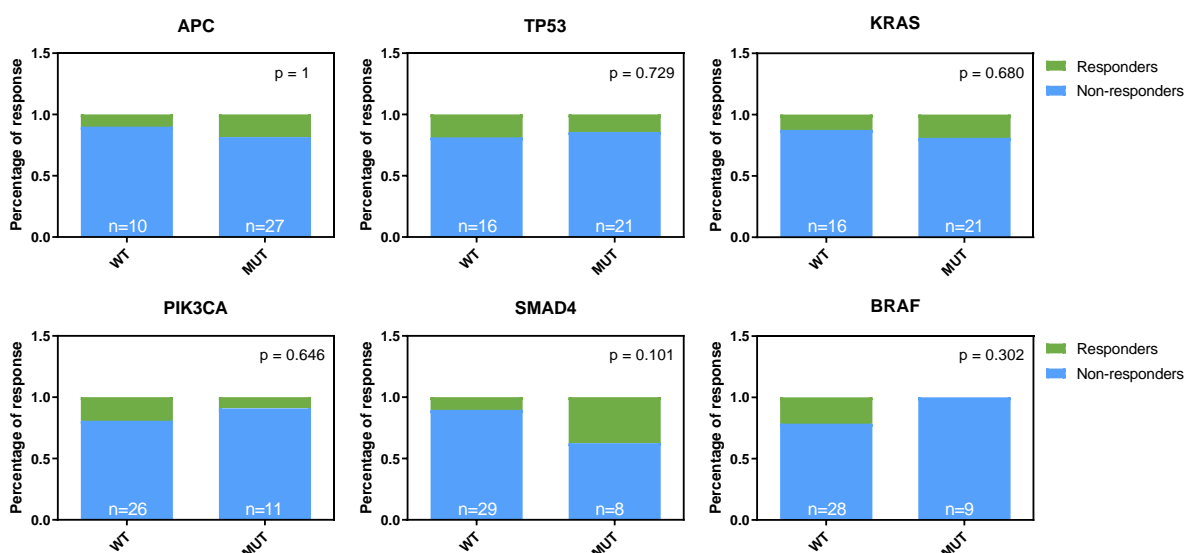


Figure 38: Association of mutations in colorectal cancer driver genes with the oxaliplatin response. Graphs show the proportion of responder or non-responder group among the cell lines of WT or MUT for *APC*, *TP53*, *KRAS*, *PIK3CA*, *SMAD4*, and *BRAF*. Each graph shows the P-value from Fisher's exact test analysis.

4.3.3 - Driver gene mutations and sensitivity to irinotecan

Next, the association between irinotecan sensitivity and mutations in these driver genes was investigated. The proportion of responders was significantly higher in *KRAS* WT than in *KRAS* MUT cell lines (P value = 0.015). In addition, the frequency of responders was higher in *TP53* WT than in *TP53* MUT, but this difference did not reach statistical significance (P-value = 0.071). No statistical difference in the response rates was observed between WT and MUT subgroups of *APC*, *PIK3CA*, *SMAD4* and *BRAF* (**Figure 39**). An association between *KRAS* mutations and the irinotecan response rates has not been observed before in patients with colorectal cancer. However, several studies have shown that irinotecan treatment activates RTK-RAS pathways (Mahli et al., 2018; Mazard et al., 2013), in which *KRAS* plays an important role, thus providing a potential explanation for the association between *KRAS* status and irinotecan sensitivity observed in cell lines treated with irinotecan as a single agent in highly controlled conditions.

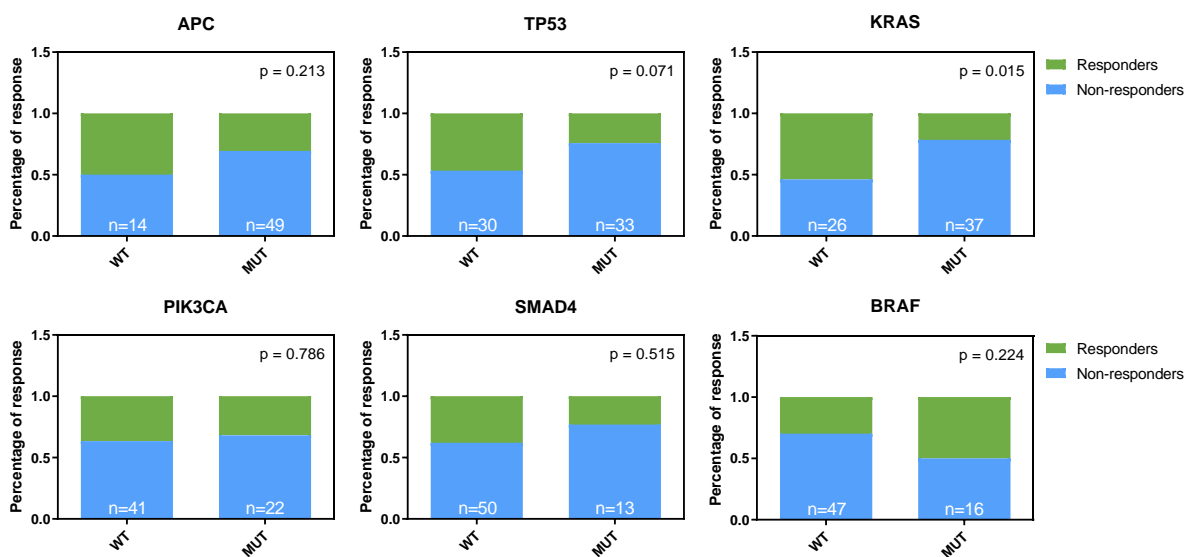


Figure 39: Association of mutations in colorectal cancer driver genes with the irinotecan response. Graphs show the proportion of responder or non-responder group among the cell lines of wild-type (WT) or mutant (MUT) for *APC*, *TP53*, *KRAS*, *PIK3CA*, *SMAD4*, and *BRAF*. Each graph shows the P-value from Fisher's exact test analysis.

4.4 Genomic, transcriptomic and proteomic datasets of a large panel of colorectal cancer cell lines

An increasing number of molecular features of the tumors have been shown to be associated with the response to chemotherapeutic drugs and can potentially be used to personalize cancer treatment. Over the past years, our group and others have characterized the genome, the transcriptome and the proteome of different overlapping cohorts of colon cancer cell lines. Here, the data generated in-house and available in public repositories for the different OMICs investigated was assembled in an attempt to maximize the number of colorectal cancer cell lines that could subsequently be used in the search for novel response biomarkers.

4.4.1 – Genomic data of colorectal cancer cell lines

We analyzed 12,154 genes, and > 20-fold coverage, > 80% of targeted exons was achieved. The sequencing was done in collaboration with Dr. John Mariadason and Dr. Oliver Sieber for 70 colorectal cancer lines (Mouradov et al., 2014). The mutational profile of each cell line was labeled as nominal variables for each gene: wild-type (WT) and mutant (MUT). A cell line was considered mutant for a gene when it presented at least one non-synonymous mutation, including insertion or deletion and single-nucleotide variant (SNV). Considering rarely mutated biomarkers may not be easy to be validated and implemented in the clinical setting, we only included the genes that fulfill: 1) mutation frequency is higher than 5% in primary colorectal tumors (according to TCGA data); 2) are expressed in colorectal tumors (average expression across the cell line panel >1); and 3) have mutations in at least three colorectal cancer cell lines in the panel. Therefore, 309 genes were included in the whole exome sequencing analysis.

4.4.2 – Transcriptomic data of colorectal cancer cell lines

We integrated the RNAseq profile of 57 cell lines in collaboration with Dr. John Mariadason (Olivia Newton-John Cancer Research Institute, Melbourne, Australia) using Illumina HiSeq

2000 (RNAseq-1) (Mouradov et al., 2014), RNAseq data of 35 cell lines from the CCLE repository using Illumina HiSeq 2000 or HiSeq 2500 (RNAseq-2) (Barretina et al., 2012), microarray of 42 cell lines done in our lab using Affymetrix HG-U133 Plus 2.0 chips (Microarray-1) (Arango et al., 2005), and microarray of 70 cell lines obtained from the study of Dr. Medico et al. using Human HT-12 V4 BeadChip (Microarray-2) (Medico et al., 2015). High-throughput data such as RNAseq and microarray provide an approach to investigating disease and gene relationships. However, due mostly to the cost of expression profiling techniques, a single study's sample number is often limited, and direct cross-comparison of the results obtained in different studies cannot be done directly (Lander, 1999).

Transcripts from 285 cell lines from four datasets were used for the transcriptomic analysis. All four datasets used had common transcriptomic profiles from 16,479 unambiguous genes, and 64 lines had transcriptomic profiles available from more than one dataset. Principal component analysis (PCA) (**Figure 40**) and clustering analysis (**Figure 41**) confirmed a strong batch effect. PCA analysis showed that principal component 1 (PC1) and principal component 2 (PC2) explain 94% and 1% of the total variance (**Figure 40a**).

ComBat is a well-known approach based on an empirical Bayes framework to correct batch effect (W. E. Johnson et al., 2007) in both transcriptomic and proteomic data (A. H. Lee et al., 2019a). ComBat normalization (GenePattern platform) was therefore performed to eliminate the batch effect from the raw expression data of the four distinct datasets used in this study. After normalization, the variability of PC1 and PC2 was 11% and 6% (**Figure 40b**) and cell lines no longer separated mostly based on the dataset of origin, indicating that this normalization approach effectively removed the batch effect.

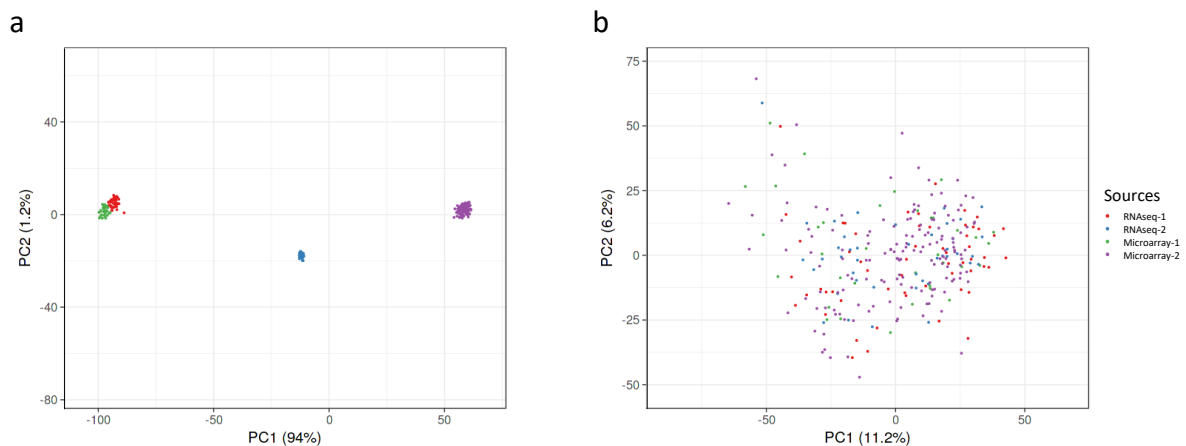


Figure 40: Principal component analysis (PCA) plots showed mRNA expression variability across 285 cell lines from four datasets (RNAseq-1, RNAseq-2, microarray-1, microarray-2) before and after conducting ComBat.

Before (a) and after (b) conducting ComBat. The X and Y axes show principal component 1 and principal component 2, respectively, which explains the total variance.

To investigate whether identical cell lines from different datasets exhibited to have a similar transcriptome, we conducted a hierarchical clustering analysis after Combat. The results showed that 60 out of the 64 cell lines represented more than once clustered with at least one identical cell line from other datasets. There were, however, several outliers that did not cluster together with the same cell lines from a different dataset, possibly due to mislabeling of the samples or cross-contamination with other cell lines (Rojas et al., 2008). We excluded these 24 outliers (described in the Methods) from the combined dataset, and the expression of the remaining cell lines represented more than once was averaged for all subsequent analyses.

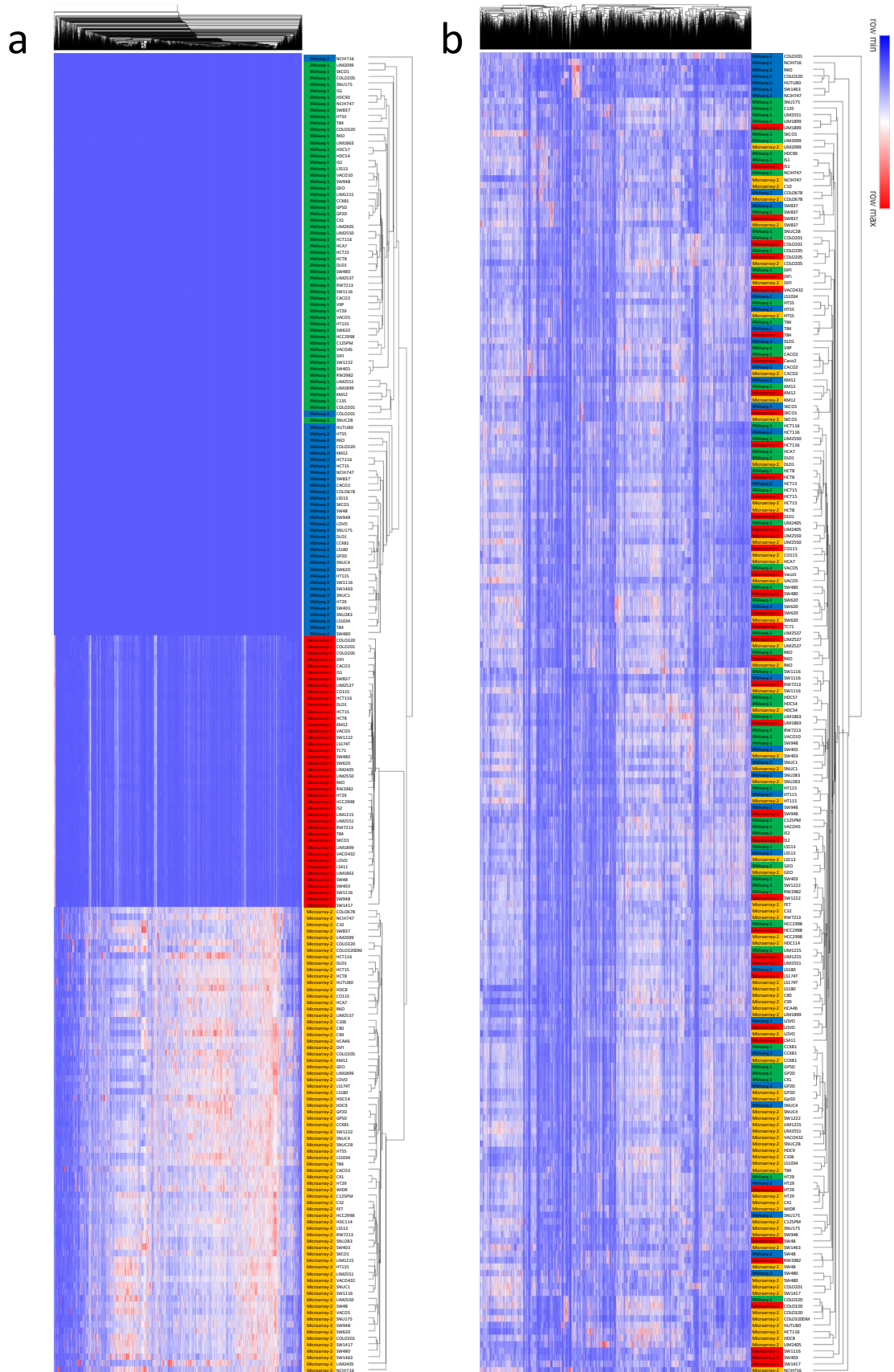


Figure 41: Hierarchical clustering of transcriptomes based on 5000 genes with the highest variations.

a) Before ComBat; **b)** After ComBat; The rows are ordered based on individual cell lines from different sources. The colored matrix indicates the gene expression. Relative gene expression levels above or below the mean were shown in red and blue. Cell lines from the same datasets were colored in the right column.

4.4.3 – Proteomic data of colorectal cancer cell lines.

Similar to the transcriptomics data processing, we merged Liquid Chromatography Mass Spectrometry-Based Proteomics from two datasets identified as Proteomics-1 (J. Wang et al., 2017) and Proteomics-2 (Roumeliotis et al., 2017). First, we conducted ComBat to remove the batch effect. The variability of PC1 and PC2 decreased from 37% and 8% PC2 before conducting ComBat to 11% and 8% after conducting ComBat, and the cell lines were no longer separated by the dataset of origin (**Figure 42**), suggesting that Combat removed the batch effect of proteomics from the two different datasets used. We then conducted a clustering analysis to identify the outliers, possibly due to mislabeling of the samples or cross-contamination with other cell lines. After excluding these cell lines, we re-transformed the proteomics data into the linear scale.

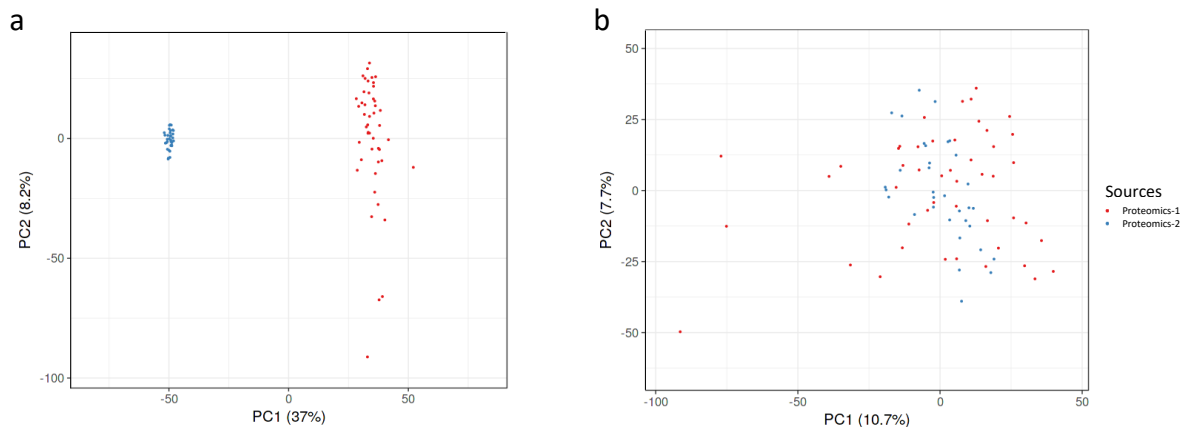


Figure 42: The similarity between Proteomics-1 and Proteomics-2 shotgun proteomics data and batch correction.

Principle component analysis (PCA) plots show the clustering of cell lines belonging to data set 1 (Proteomics-1), and data set 2 (Proteomics-2) batches before (a) and after (b) batch correction using the ComBat algorithm.

Clustering analysis showed that every identical cell line from different sources formed clusters in the hierarchical analysis. In contrast, no identical cell line from the two datasets formed clusters before conducting ComBat (**Figure 43**). If the proteome data were derived from more than one database, the average was used for this cell’s proteome data.

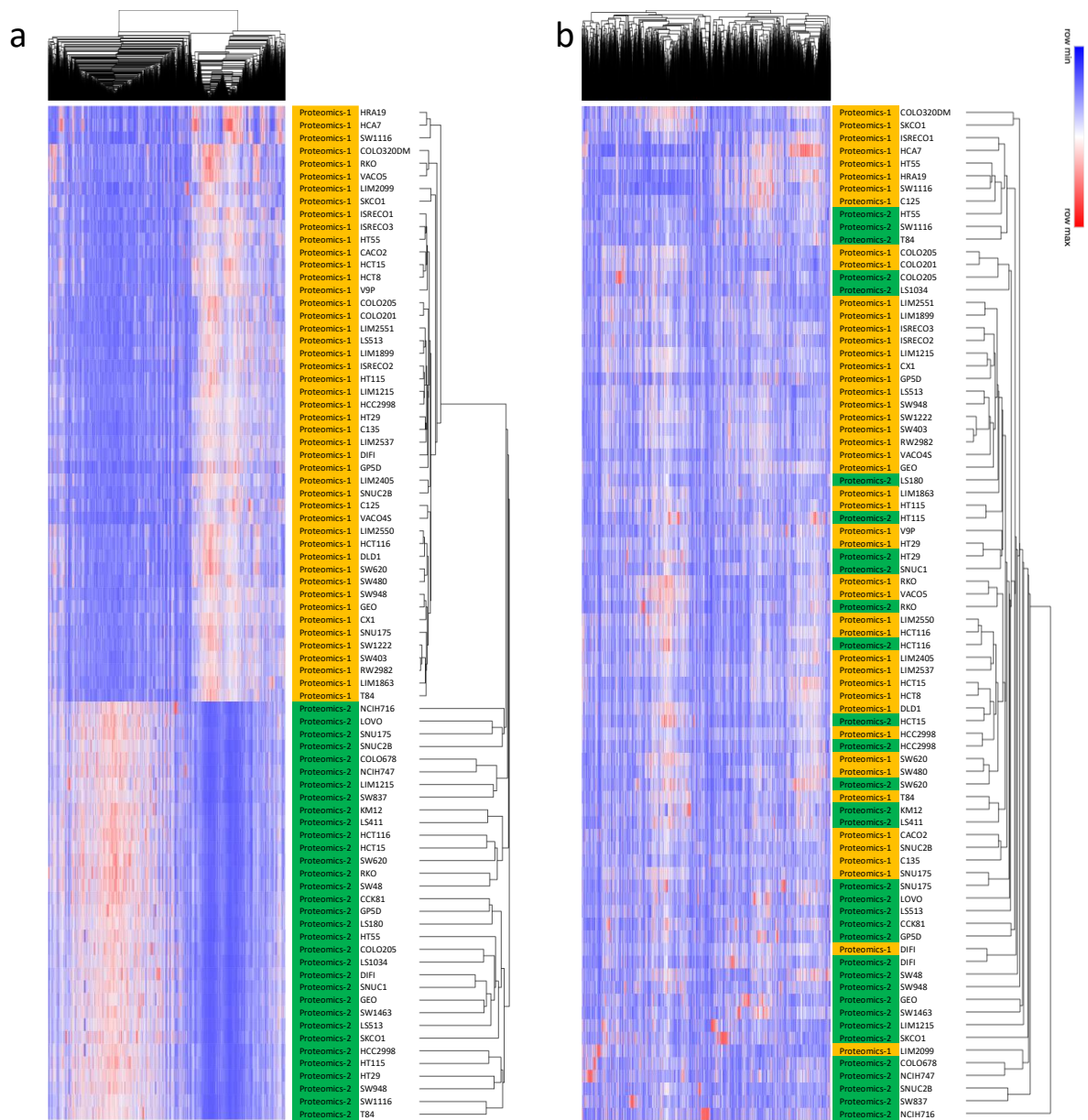


Figure 43: Hierarchical clustering analysis of proteome from two datasets after ComBat merging.

The rows are ordered based on individual cell lines from different sources. The colored matrix indicates the gene expression. Cell lines from the same datasets were highlighted by the same color in the right column. Relative gene expression levels above or below the mean were shown in red and blue. P1: protein dataset 1; P2: protein dataset 2.

4.5 Association between OMICs data and drug response

Currently, there are two major approaches to exploring biomarkers: mechanism-based strategies and unbiased high-throughput Screening. The association of *KRAS* mutations and

clinical resistance to epidermal growth factor receptor (EGFR)-specific antibody therapy is an example of biomarkers developed by mechanism-based strategies (de Roock et al., 2011; Zhu et al., 2021). However, many potential predictive biomarkers have been discovered by using high-throughput screening. This strategy enables the investigation of novel biomarkers that have yet to be well studied at the mechanistic level. For example, Oncotype DX (21-gene recurrence score), based on the expression of 21 genes, was developed by gene-expression profiling and accurately predicts endocrine and chemotherapy response in breast cancer patients (Andre et al., 2022).

Therefore, genomic, transcriptomic and proteomic profiling was used to identify new candidate biomarkers of response to 5-FU, oxaliplatin and irinotecan in colorectal cancer patients.

4.5.1 - Associations between OMICs data and sensitivity to 5-FU

Although many new drugs for colorectal cancer have emerged in the past few decades, 5-FU remains the backbone of the first-line treatment for colorectal cancer patients, both in the adjuvant and palliative settings. Nevertheless, the average response rate of 5-FU as monotherapy is only about 16-30% (Douillard et al., 2000; Giacchetti et al., 2000; Johnston & Kaye, 2001), and there is currently no biomarker approved to predict the response of 5-FU-based treatment.

Candidate DNA biomarkers of response to 5-FU. Analyses of whole-exome sequencing of colorectal tumors have previously revealed specific genetic alterations with clinical implications. For example, mutations in *KRAS* predict the poor response of EGFR inhibitors. 309 genes of 70 colorectal cancer cell lines were available and fulfilled the following criteria: 1) mutation frequency >5% in primary tumors (TCGA); 2) expressed in colorectal tumors (average mRNA >1); and 3) at least three cell lines with mutations in the panel analyzed, but 37 out of 70 cell lines had the 5-FU response data. Therefore, whole exon sequencing of a panel of 37 lines was used to identify possible coding mutations in expressed genes which could be associated with response to 5-FU. Of the 309 genes included in this analysis, 6 were significantly associated with the response to 5-FU (Fisher's exact test $P < 0.05$). However, after

multiple test correction, none of these genes showed statistically significant differences (FDR < 0.2) in their mutation frequency between responder and non-responder cell lines (**Table 8**).

Table 8: Association of mutations with responders and non-responders to 5-FU.

	WT		MUT		P-value	FDR
	Responders	Non-responders	Responders	Non-responders		
BRWD3	9	22	5	1	0.021	1
COL3A1	14	16	0	7	0.031	1
ROBO1	8	21	6	2	0.035	1
CYLD	11	23	3	0	0.047	1
CEP135	11	23	3	0	0.047	1
PTPRM	11	23	3	0	0.047	1

Note: P-values from Fisher’s exact test analysis are shown.

Candidate mRNA biomarkers of response to 5-FU. Over the last two decades, genome-wide microarray and RNAseq analysis have revolutionized the capacity to identify and search for mRNA molecules whose expression may be associated with drug response. Here, transcriptomic profiling was available for 53 out of 57 cell lines to identify novel candidate biomarkers of response to 5-FU. Regarding the transcriptomic analysis of 5-FU, of the 16,479 genes included in this analysis, 921 (5.6%) genes showed a significant difference (Mann-Whitney test $P < 0.05$) between responder and non-responder cell lines. Of these 921 genes, 457 genes (49.6%) showed higher levels in non-responder cell lines, while the remaining 464 (50.4%) genes showed higher expression in the responder cell lines compared to the non-responder cell lines. Importantly, since 16,479 individual analyses were conducted, a P-value of 0.05 would be expected to yield 5% of false positive results (*i.e.*, 824 genes). Therefore, we used the Benjamini-Hochberg (BH) method for multiple test correction. No significant differences in expression at the mRNA level were observed between responder and non-responder cell lines after BH correction (FDR < 0.2) for multiple testing (**Figure 44a**).

Candidate protein biomarkers of response to 5-FU. Next, we investigated whether gene expression at the protein level could be used to identify novel biomarkers of response to 5FU. Of the 4,770 proteins included in this analysis, 198 (4.2%) showed a significant difference in their expression between responder and non-responder cell lines. Of these 198 proteins, 108

(54.5%) exhibited higher expression in the responder group, while 90 (45.5%) proteins showed higher expression in non-responder cell lines compared to the responder group. However, none of these proteins showed statistically different expression between responder and non-responder cell lines after multiple test correction ($FDR < 0.2$), as shown in **Figure 44b**.

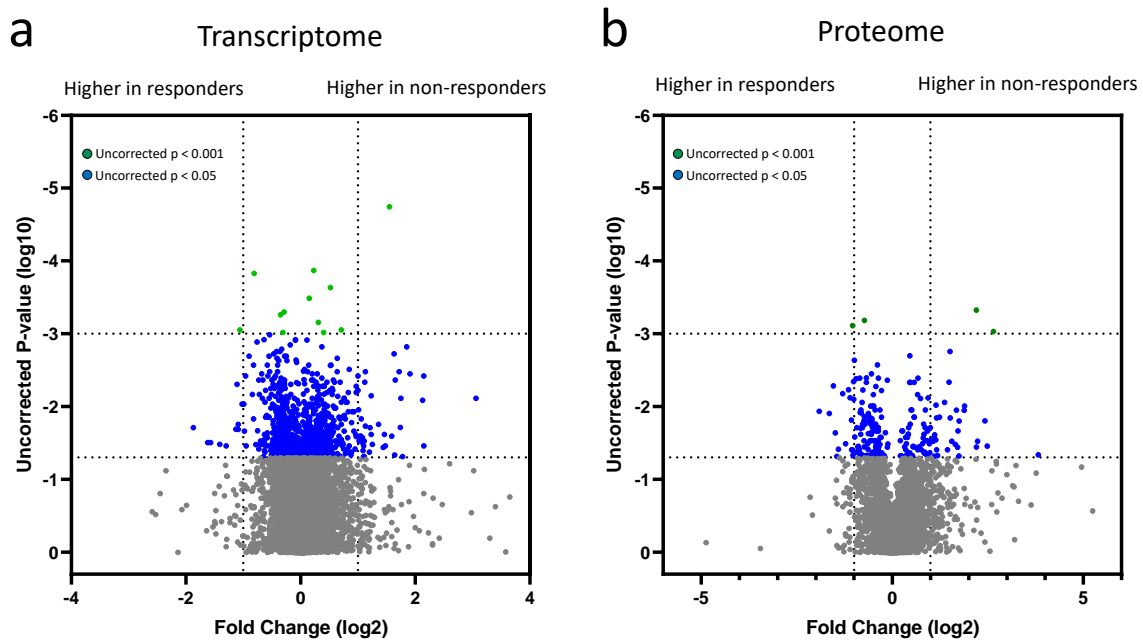


Figure 44: Association of transcriptome and proteome expression with responder or non-responder colorectal cancer cell lines to 5-FU.

a-b) Volcano plots show genes whose expression at the mRNA (transcriptome), **a)** and protein level (proteome); **b)** are different between responders and non-responders. The graphs also indicate if the expression is higher in responders (dots in the left part) or non-responders (dots in the right part). The color of the dots indicates significance levels.

4.5.2 - Associations between OMICs data and sensitivity to oxaliplatin

Oxaliplatin as a single agent has demonstrated a 10-20% response rate in colorectal cancer patients (Raymond, Chaney, et al., 1998). However, there is currently no approved biomarker to predict the response rate to oxaliplatin.

Candidate DNA biomarkers of response to oxaliplatin. Exome sequencing data and oxaliplatin sensitivity were available for 37 cell lines, and 309 genes fulfilled the selected criteria described before. Four genes were significantly associated with the response to

oxaliplatin (uncorrected $P < 0.05$), but none of them passed the multiple test correction (FDR < 0.2) (**Table 9**).

Table 9: Association of mutations with responders and non-responders to oxaliplatin.

	WT		MUT		P-value	FDR
	Responders	Non-responders	Responders	Non-responders		
SOX9	2	27	4	4	0.013	1
THBS2	4	31	2	0	0.023	1
NOTCH2	3	28	3	3	0.042	1
GTF3C1	3	28	3	3	0.042	1

Note: P-values from Fisher’s exact test analysis are shown.

Candidate mRNA biomarkers of response to oxaliplatin. A significant difference (Mann-Whitney test $P < 0.05$) between responder and non-responder cell lines was observed for 1,171 (7.1%) transcripts. Of these 1,171 mRNAs, 583 (49.8%) displayed a higher expression in the responder group, while 588 (50.2%) genes were higher expressed in non-responder cell lines compared to the responder group. Nine transcripts, namely *ARNTL*, *LINC01549*, *BCL9L*, *BTBD10*, *IRF7*, *YAP1*, *TRPM4*, *TAP2*, and *DIAPH1*, showed significant differences in the responder and non-responder groups after multiple test correction (FDR < 0.2 ; **Figure 45a**).

Candidate protein biomarkers of response to oxaliplatin. A total of 275 out of 4,770 proteins (5.8%) showed a significant difference in their expression between responder and non-responder cell lines. Of these 275 proteins, 156 (56.7%) displayed higher expression in the responder group, while 119 (43.3%) proteins had higher expressed in non-responders cell lines compared to the responder group. After multiple test correction, HNF4G showed a significant expression difference (FDR < 0.2) between the responder and non-responder groups, as shown in **Figure 45b**.

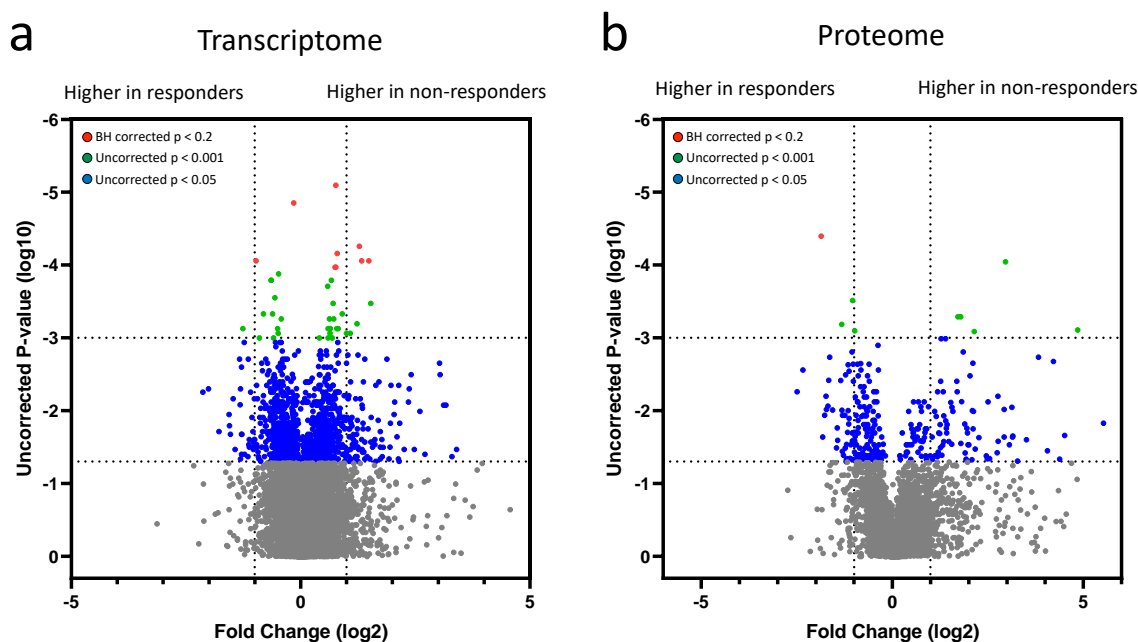


Figure 45: Association of transcriptome and proteome expression with responder or non-responder colorectal cancer cell lines to oxaliplatin.

a-b) Volcano plots show transcriptome (a) and proteome (b) whose expressions are different between responders and non-responders. The graphs also indicate if the expression is higher in responders (dots in the left half part) or non-responders (dots in the right half part). The color-coding dots indicate significance levels.

4.5.3 - Associations between OMICs data and sensitivity to irinotecan

As a key drug for treating metastatic colorectal cancer, irinotecan monotherapy demonstrated a 16-41% response rate (Rougier et al., 1997). However, there is currently no approved biomarker to predict the sensitivity of irinotecan.

Candidate DNA biomarkers of response to irinotecan. Exome sequencing data and irinotecan sensitivity were available for 63 cell lines, and 309 genes fulfilled the selected criteria described before. Six genes were significantly associated with the response to irinotecan (uncorrected $P < 0.05$), but none of these genes passed the $FDR < 0.2$ multiple-test correction (Table 10).

Table 10.

Table 10: Association of mutations with irinotecan responder and non-responder cell lines.

	WT		MUT		P-value	FDR
	Responders	Non-responders	Responders	Non-responders		
KRAS	14	12	8	29	0.015	1
FGFR2	16	39	6	2	0.018	1
CNOT1	22	33	0	8	0.042	1
RANBP2	22	33	0	8	0.042	1
KALRN	17	39	5	2	0.045	1
THBS2	18	40	4	1	0.046	1

Note: P-values from Fisher's exact test analysis are shown.

Candidate mRNA biomarkers of response to irinotecan. The expression of 16,479 genes at the transcript level in a panel of 79 responder and non-responder cell lines was compared using a Mann-Whitney test. A significant difference (uncorrected $P < 0.05$) between non-responder and responder lines was observed for 1,331 (7.1%) transcripts. Of these 1,331 mRNAs, 855 (64.2%) showed higher expression in the responder group, while 476 (35.8%) transcripts were expressed at a higher level in non-responder cell lines compared to responder lines. However, none of these transcripts was differentially expressed between responders and non-responders using an FDR of 0.2, as shown in the volcano plot in **Figure 46a**.

Candidate protein biomarkers of response to irinotecan. Protein expression data for 4,770 genes in 56 cell lines with irinotecan sensitivity data was available. A total of 448 proteins (9.4%) showed a significantly different expression between responder and non-responder cell lines (uncorrected $P < 0.05$). Of these 448 proteins, 230 (51.3%) proteins showed higher expression in the responder cell lines, while 218 (48.7%) proteins showed higher expression in the non-responder group. Importantly, 40 of these proteins had significant differences in their expression in responder and non-responder groups after multiple-test correction (FDR < 0.2) (**Figure 46b**).

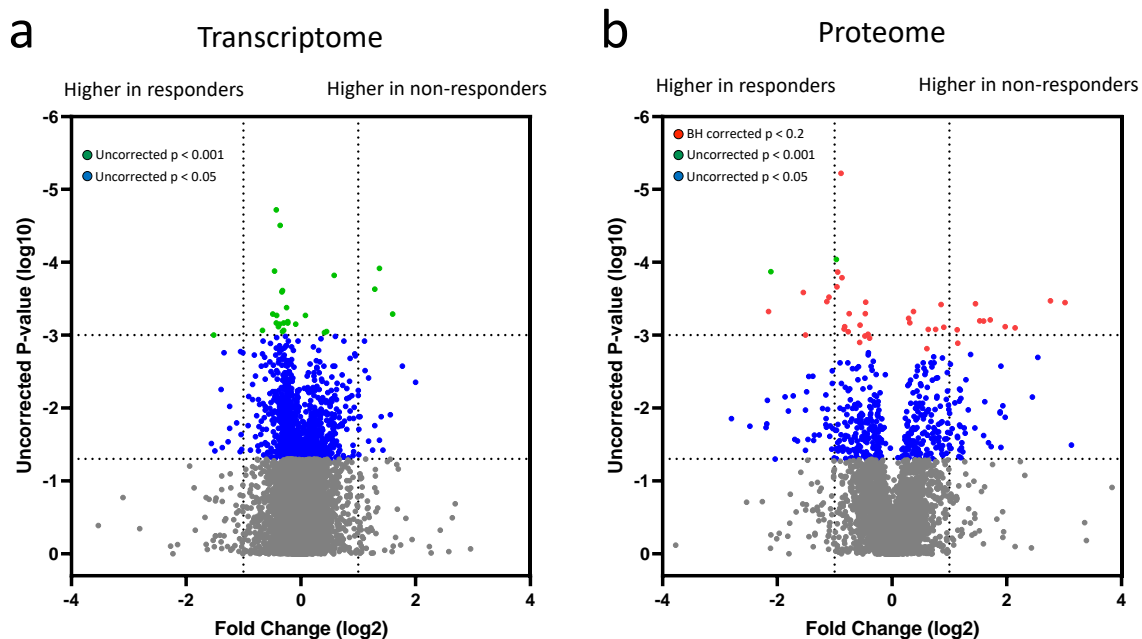


Figure 46: Association of transcriptome and proteome expression with responder or non-responder colorectal cancer cell lines to irinotecan.

a-b) Volcano plots show transcriptome (a) and proteome (b) whose expressions are different between responders and non-responders. The graphs also indicate if the expression is higher in responders (dots in the left half part) or non-responders (dots in the right half part). The color-coding dots indicate significance levels.

4.6 Functional group enrichment analysis

To explore and understand the biological function and regulatory pathways of differentially expressed genes between responder and non-responder cell lines, we performed Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the differentially expressed ($P < 0.05$) genes and proteins between responders and non-responders to 5-FU, oxaliplatin and irinotecan. Gene ontology annotations were obtained from the DAVID database (<https://david.ncifcrf.gov/home.jsp>), which included the biological process (BP), cellular component (CC), and molecular function (MF). The results were deemed statistically significant if the multiple test corrected (FDR) P-value was less than 0.05.

4.6.1 Functional group enrichment analysis of genes differentially expressed in responder and non-responder cell lines to 5-FU

For 5-FU, in the enrichment analysis of the transcriptome, the most enriched GO terms were “mitochondrial matrix” and “mitochondrion” in cellular component and “protein binding” in molecular function. No differences were observed in any annotations for biological process and KEGG pathways.

In the proteomic analysis, the most enriched GO terms were “oligosaccharyltransferase complex”, “mitochondrial inner membrane”, and “mitochondrial matrix” in cellular component, and “RNA binding” and “protein binding” in molecular function. The most enriched KEGG pathways were “Metabolic pathways”. All the GO terms and KEGG pathways that were significantly enriched (FDR < 0.05) were listed in the enrichment dot bubble plot (**Figure 47**) and tables (**Appendix 1**)

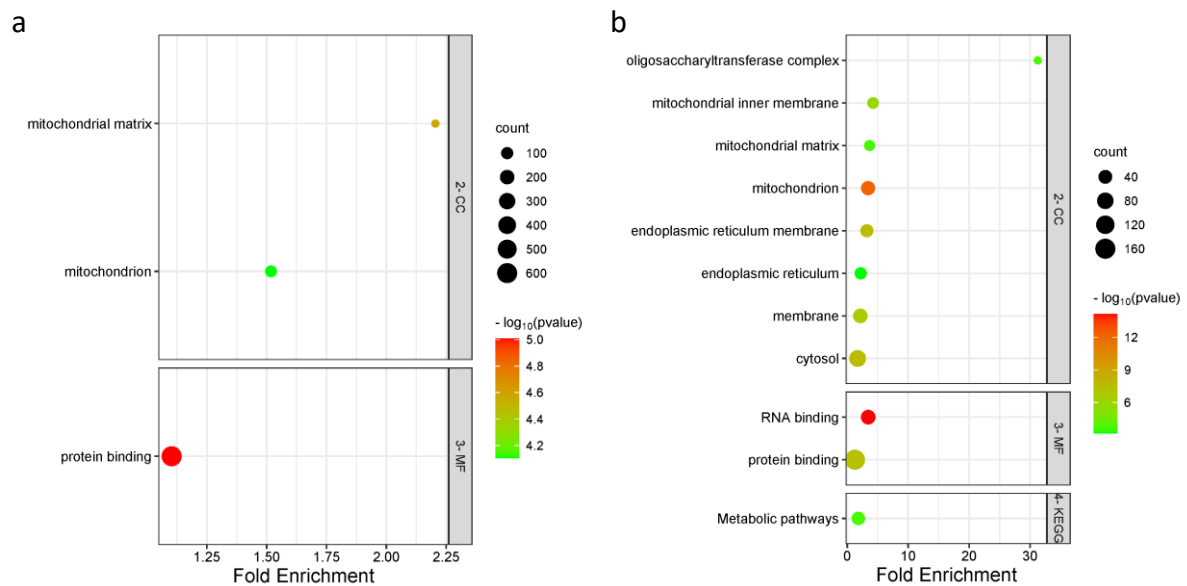


Figure 47: Functional group enrichment analysis of genes differentially expressed in responder and non-responder cell lines to 5-FU.

GO terms and KEGG pathway enrichment plot using transcriptomic (a) and proteomic (b) data between responder and non-responder cell lines to 5-FU. The vertical axis indicates the cellular component (CC), molecular function (MF), and KEGG pathways (KEGG) name, and the horizontal axis represents the enrichment score. The color of the dots represents the size of the P-value. The smaller the P-value, the closer the color is to the red color. The dots size expresses the number of genes in each GO term or pathways. The enriched GO terms and pathways (FDR < 0.05) are represented in the plot.

4.6.2 Functional group enrichment analysis of genes differentially expressed in responder and non-responder cell lines to oxaliplatin

In the enrichment analysis of the transcriptome, the most enriched GO terms were “ruffle membrane”, “cell junction”, and “focal adhesion” in cellular component and “protein binding” in molecular function.

In the enrichment analysis using proteomics data, the most enriched GO terms were “postsynaptic cytosol”, “ruffle membrane”, and “peroxisome” in cellular component, and “cadherin binding involved in cell-cell adhesion”, “cadherin binding”, and “actin binding” in molecular function. The most enriched KEGG pathways were “Metabolic pathways” (**Figure 48**).

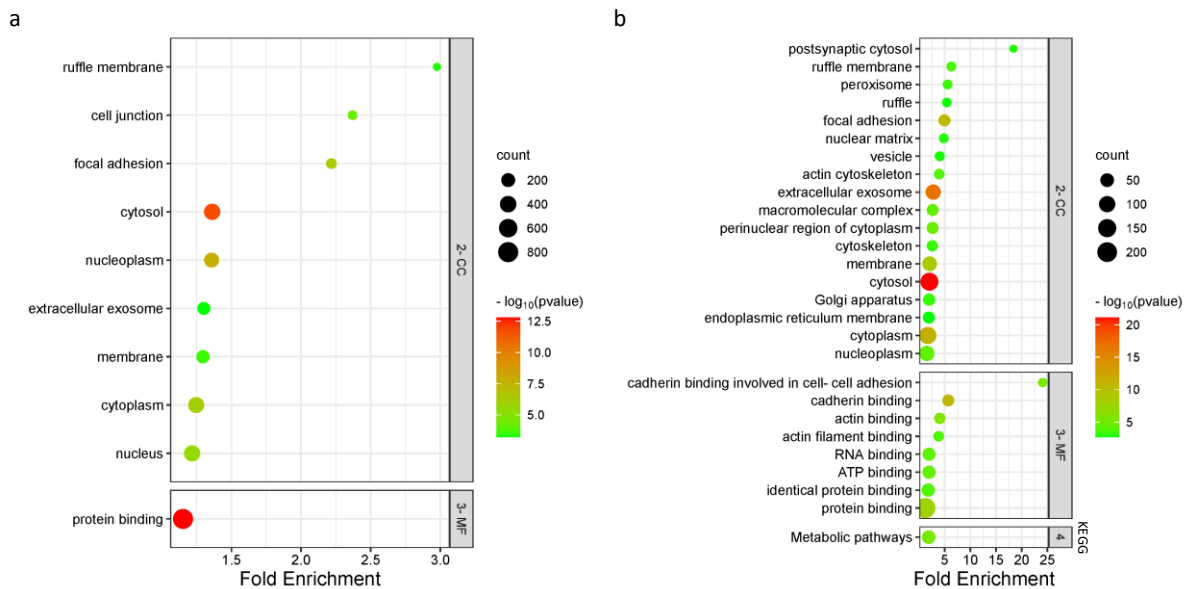


Figure 48: GO terms and KEGG pathway enrichment plot using proteomes between responder and non-responder cell lines to irinotecan.

The vertical axis indicates the cellular component (CC), molecular function (MF), and KEGG pathways (KEGG) name, and the horizontal axis represents the enrichment score. The color of the dots represents the size of the P-value. The smaller the P-value, the closer the color is to the red color. The dots size expresses the number of genes in each GO term or pathways. The enriched GO terms and pathways (FDR < 0.05) are represented in the plot.

4.6.3 Functional group enrichment analysis of genes differentially expressed in responder and non-responder cell lines to irinotecan

In the enrichment analysis of transcriptomics data, the most enriched GO terms were “centriole”, “basolateral plasma membrane”, and “centrosome” in cellular component, and “ubiquitin-protein transferase activity”, “ATPase activity”, and “protein serine/threonine kinase activity” in molecular function (**Figure 49**).

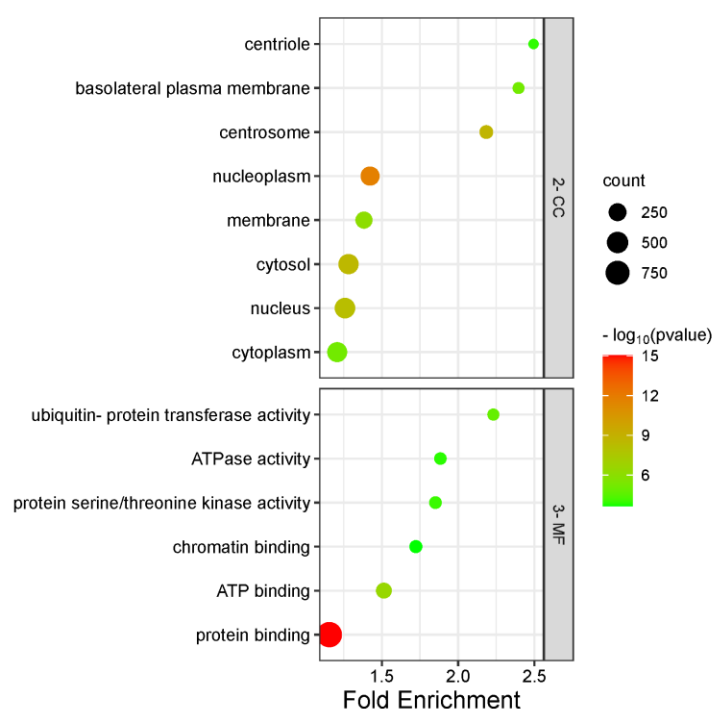


Figure 49: GO terms and KEGG pathway enrichment plot using proteomes between responder and non-responder cell lines to irinotecan.

The vertical axis indicates the cellular component (CC) and molecular function (MF) name, and the horizontal axis represents the enrichment score. The color of the dots represents the size of the P-value. The smaller the P-value, the closer the color is to the red color. The dots size expresses the number of genes in each GO term or pathways. The enriched GO terms and pathways ($FDR < 0.05$) are represented in the plot.

In the enrichment analysis of proteomics data, the most enriched GO terms were “formation of cytoplasmic translation initiation complex”, “protein transport”, “negative regulation of amyloid fibril formation” in biological process, and “cytosol”, “extracellular exosome”, and “mitochondrion” in cellular component, and “protein binding”, “cadherin binding”, and “RNA binding” in molecular function. The most enriched KEGG pathways were “Metabolic

pathways”, “Valine, leucine and isoleucine degradation”, and “Carbon metabolism” (Figure 50 and tables Appendix 4).



Figure 50: GO terms and KEGG pathway enrichment plot using proteomes between responder and non-responder cell lines to irinotecan.

The vertical axis indicates the biological process (BP), cellular component (CC), molecular function (MF), and KEGG pathways (KEGG) name, and the horizontal axis represents the enrichment score. The color of the dots represents the size of the P-value. The smaller the P-value, the closer the color is to the red color. The dots size expresses the number of genes in each GO term or pathways. The enriched GO terms and pathways (FDR < 0.05) are represented in the plot.

4.6.4 Common enriched functional group

We also explored the common enriched functional group between 5-FU, oxaliplatin, and irinotecan analysis. Although our objective is identifying biomarkers of response in monotherapy, given that these drugs are combined in the clinical setting, investigating the genes whose expression co-variates might be valuable. Furthermore, additional biomarkers could be identified in this way. For the transcriptome analysis, “protein binding” was the overlapping gene sets enriched in three drugs analysis. Regarding proteomics analysis, “cytosol”, “membrane”, “RNA binding”, “protein binding”, and “metabolic pathways” were shared in the study of the three drugs (**Figure 51**).

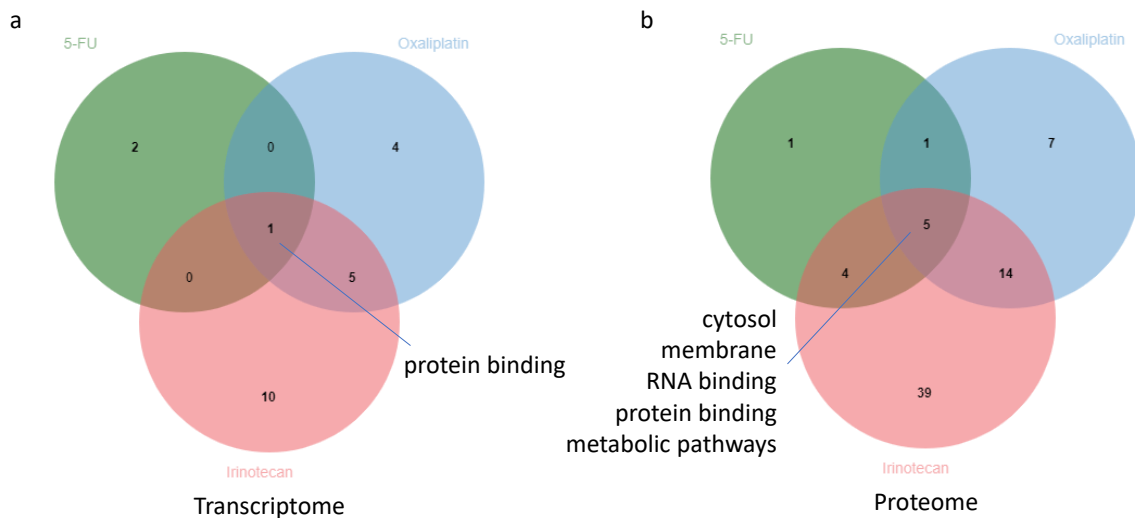


Figure 51: Intersection between the 5-FU, oxaliplatin, and irinotecan functional group analysis.

a) Venn diagram reporting the number of enriched gene sets shared between 5-FU, oxaliplatin, and irinotecan functional group analysis on transcriptomic data of responder and non-responder cell lines; **b)** Venn diagram reporting the number of enriched gene sets shared between 5-FU, oxaliplatin, and irinotecan functional group analysis on proteomic data of responder and non-responder cell lines.

4.7 Potential biomarkers of drug response


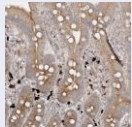
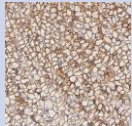
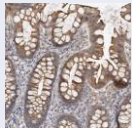
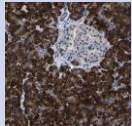
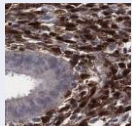

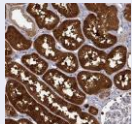
To assess the potential of the candidate biomarkers identified in the panel of colorectal cancer cell lines in the clinical setting, we next aimed at the validation of some of these biomarkers on a large cohort of colorectal cancer patients with advanced colorectal cancer that were treated with the chemotherapeutic agents studied. Since the sensitivity data of irinotecan was available for 90 cell lines, while there were only 57 cell lines with 5-FU and oxaliplatin drug response data, the work of this thesis focused on the validation of the predictive power of candidate biomarkers of response to irinotecan. Because no significant association ($FDR < 0.2$) was found between gene mutations or mRNA expression in responder and non-responder cell lines to irinotecan, we focused the validation studies on the protein expression biomarkers found in the cell lines. The potential predictive biomarkers of 5-FU and oxaliplatin will be investigated in future studies.

As described above, 40 proteins were found to have a significantly different expression in responder and non-responder cell lines based on proteome analysis and correction for multiple testing ($FDR < 0.2$; red dots in **Figure 46b**). To prioritize the candidate biomarkers to be further validated in a cohort of samples from 548 colorectal cancer patients treated with irinotecan, first, we focused on the 16 genes showing a significant difference in the average protein expression between responder and non-responder groups that was greater than 2-fold: candidate biomarker 4, candidate biomarker 6, candidate biomarker 7, candidate biomarker 9, candidate biomarker 10, candidate biomarker 11, candidate biomarker 12, candidate biomarker 13, candidate biomarker 14 and candidate biomarker 16 expressed higher in non-responders; candidate biomarker 1, candidate biomarker 2, candidate biomarker 3, candidate biomarker 5, candidate biomarker 8 and candidate biomarker 15 expressed higher in responders. Next, to preliminary assess the potential of these 16 new candidate biomarkers in the clinical setting and further prioritize the genes to be validated in our cohort of 548 patients samples, we used an external database of 221 metastatic colorectal cancer patients treated with irinotecan for which the response to treatment is publicly available through the online analysis tool ROC Plotter (<https://www.rocplot.org>) (Fekete & Györffy, 2019). However, because ROC Plotter uses mRNA (rather than protein) expression information to assess associations with the response to irinotecan, we first assessed whether the expression at the mRNA and protein levels was significantly correlated in the 16 candidate biomarkers. Importantly, a significant

correlation (Spearman's correlation P-value < 0.05) for all 16 genes was found (**Appendix 4**), and initial validation using ROC Plotter could therefore be carried out. Next, ROC Plotter was used to assess whether the transcript expression of these 16 genes was different in the responder and non-responder groups in irinotecan-treated patient samples and found that ten of them had a significant difference (Mann-Whitney test P-value < 0.05): Candidate biomarker 4, candidate biomarker 6, candidate biomarker 7, candidate biomarker 8 candidate biomarker 9, candidate biomarker 11, candidate biomarker 15 and candidate biomarker 16 showed higher expression in non-responders, while candidate biomarker 1 and candidate biomarker 5 had higher expression in the responder group compared to the non-responder group. Eight out of ten genes, candidate biomarker 1, candidate biomarker 4, candidate biomarker 5, candidate biomarker 6, candidate biomarker 7, candidate biomarker 9, candidate biomarker 11 and candidate biomarker 16, showed the same trend as our findings in the cell line panel (**Appendix 5**).

Immunohistochemistry is a robust, widely used technique that can accurately determine the levels of expression of proteins in histological sections of formalin-fixed, paraffin-embedded tumor samples using specific antibodies. The validation of candidate biomarkers was carried out by immunohistochemistry on tissue microarray (TMA) sections containing a total of 1419 tumor samples from 548 patients (2-3 independent samples/patient). Therefore, it was important to have commercially available antibodies that could be used on formalin-fixed, paraffin-embedded samples. We conducted a thorough analysis of the commercial antibodies available for each protein and found reliable antibodies for its use in immunohistochemistry on formalin-fixed, paraffin-embedded samples for all 8 candidate biomarkers (**Table 11**) (<https://www.proteinatlas.org/>). After a profound review of the literature, SLC29A1 and GUCY2C were selected to pursue further validation of the clinical samples using immunohistochemistry.

Table 11: Representative antibodies recognizing the potential biomarkers in immunohistochemistry.

Proteins	Host species	Concentration	Antigen retrieval buffer	Representative staining
SLC29A1	Rabbit	0.2386 mg/ml	HIER pH6	
GUCY2C	Rabbit	0.2386 mg/ml	HIER pH6	
ABCB1	Rabbit	0.145 mg/ml	HIER pH6	
Candidate biomarker 11	Rabbit	0.085 mg/ml	HIER pH6	
Candidate biomarker 16	Rabbit	0.93 mg/ml	HIER pH6	
Candidate biomarker 1	Rabbit	0.0247 mg/ml	HIER pH6	
Candidate biomarker 5	Rabbit	1.152 mg/ml	HIER pH6	
Candidate biomarker 4	Rabbit	0.03 mg/ml	HIER pH6	

HIER: heat-induced epitope retrieval.

4.8 Validation of the potential of SLC29A1 protein expression to predict response to irinotecan-based treatment for colorectal cancer patients

Solute Carrier Family 29 Member 1 (SLC29A1) is an equilibrative nucleoside transporter localized to the plasma and mitochondrial membranes and mediates the cellular uptake of

nucleosides. SLC29A1 is required to uptake nucleotide synthesis in the cells that depend on the salvage pathway nucleotide synthesis. Since previous studies showed that SLC29A1 plays a vital role in transporting nucleotides and impacts cell viability (Y. Huang, Cate, et al., 2004), it may associate with anti-cancer drugs.

4.8.1 Validation of SLC29A1 mRNA levels as a biomarker of response to irinotecan in an external cohort of 221 metastatic colorectal cancer patients.

As described previously, we first confirmed that the expression of SLC29A1 shows a significant correlation at the mRNA and protein levels (**Figure 52a**) and used for validation an external cohort of 221 metastatic colorectal cancer patients that received irinotecan-based treatment and for which response and mRNA expression data is publicly available (ROC Plotter; <https://rocplot.org/>). Consistent with the findings in the panel of 90 cell lines, high levels of SLC29A1 mRNA expression were significantly associated with lower response rates to irinotecan (P-value < 0.001, AUC = 0.65; **Figure 52b-c**).

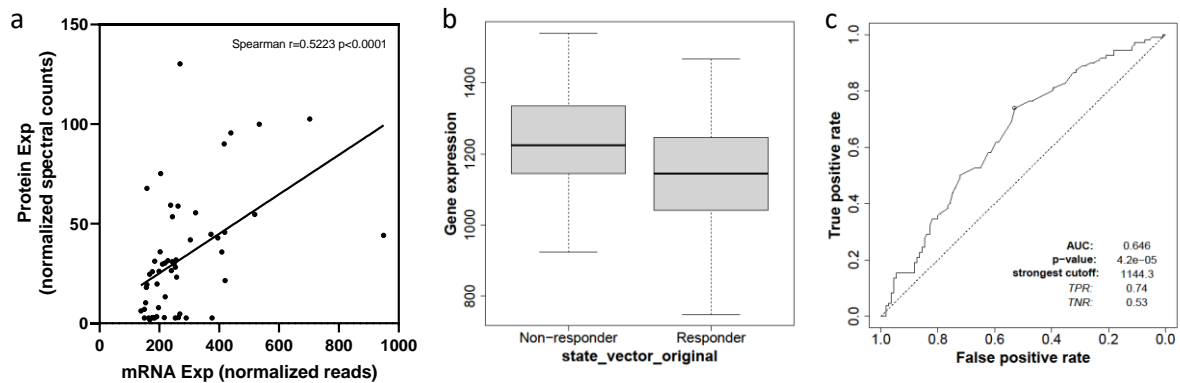


Figure 52: Validation of SLC29A1 predictive power in an external cohort of 221 metastatic colorectal cancer patients treated with irinotecan.

a) Correlation between SLC29A1 mRNA and protein expression in the panel of colorectal cancer cell lines (Spearman correlation). **b)** The *SLC29A1* mRNA expression in irinotecan responder and non-responder patients in an external cohort of 221 metastatic colorectal patients (Mann-Whitney test P-value < 0.001). **c)** ROC curve (receiver operating characteristic curve) reveals the ability of this model to classify the patients in this cohort as responders and non-responders at all expression thresholds that can be used to dichotomize patients as responders and non-responders.

4.8.2 Validation of SLC29A1 protein levels as a biomarker of response to irinotecan in a large cohort of metastatic colorectal cancer patients

To validate the clinical potential of SLC29A1 protein levels as a predictive biomarker to irinotecan treatment in formalin-fixed, paraffin-embed (FFPE) tumors from a large cohort of metastatic colorectal cancer patients by immunohistochemistry, first the specificity and the sensitivity of the commercial antibody listed in Table 11 was thoroughly validated (**Figure 53**).

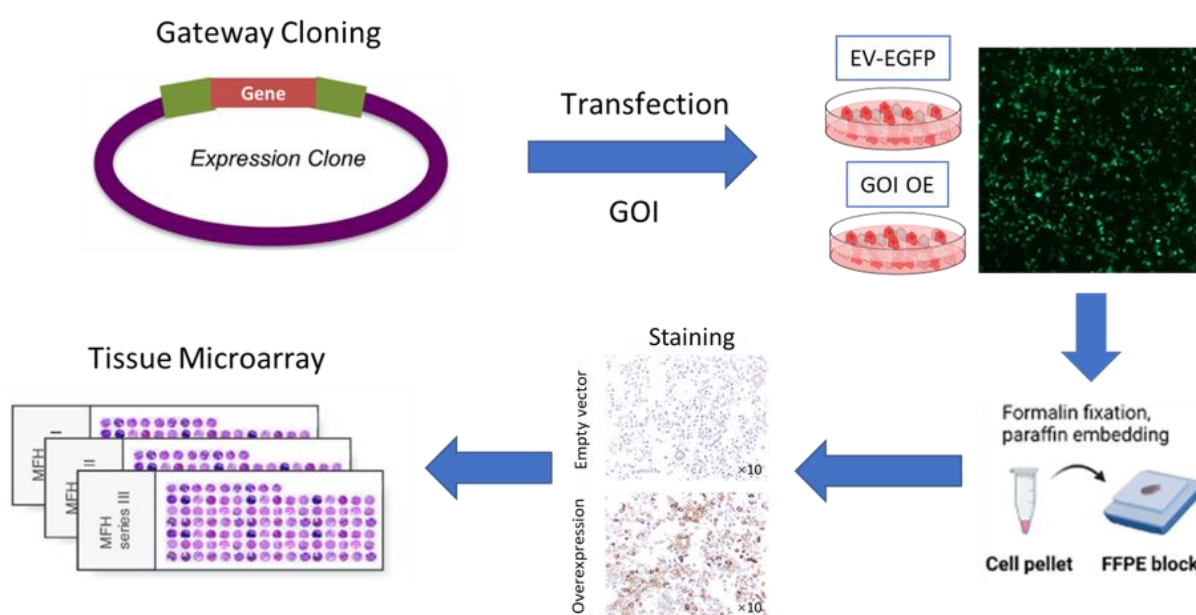


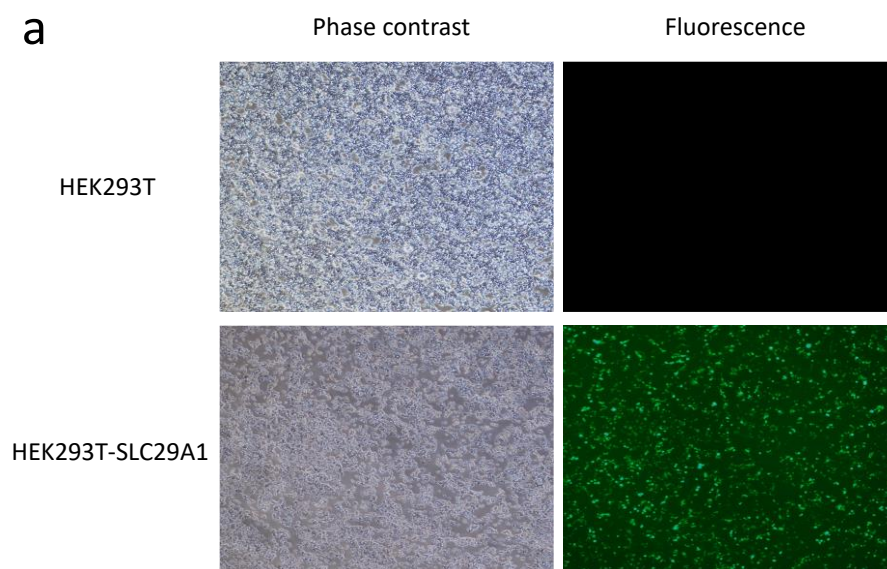
Figure 53: Biomarker validation workflow.

The protein of interest and control EGFP protein was overexpressed in HEK293T cells using doxycycline-inducible lentiviral vectors. Fluorescence microscopy was conducted 48h after transfection to determine the transfection efficiency. Cell pellets were collected, and formalin-fixed paraffin-embedded (FFPE) to test the specificity of the selected antibodies by immunohistochemistry staining. Then, the potential biomarker was validated on a tissue microarray (TMA) of colorectal cancer cell lines by immunohistochemistry. We confirmed the consistency of the TMA results and LC/MS proteomics data by performing a Pearson correlation analysis.

4.8.2.1 Confirmation of the specificity of the anti-SLC29A1 antibody

First, a cell line system with forced transient overexpression SLC29A1 was engineered. HEK293T cells were transfected with pINDUCER20-SLC29A1 together with pINDUCER20-

EGFP (**Figure 54a**). Then, cell pellets were processed for both, protein extraction and Western blotting and formalin-fixed and paraffin-embedded as routinely conducted for primary tumor samples in the pathology department of clinical institutions worldwide. Then the overexpression of SLC29A1 and the specificity of the antibody were tested by Western blotting in control cells and cells transfected with SLC29A1 together with EGFP. A robust overexpression of the protein of interest was confirmed by the presence of a band of 50 KDa in the SLC29A1 overexpressing cells (**Figure 54b**). Next, histological sections of the paraffin blocks containing cell pellets were used to assess the capacity of the antibody to detect SLC29A1 in FFPE samples. Significantly stronger immunostaining was observed in HEK293T cells overexpressing SLC29A1 compared to control HEK293 cells (**Figure 54c**). Finally, to further confirm the specificity of the antibody to detect SLC29A1 on FFPE samples, a tissue microarray (TMA) containing 54 colorectal cancer cell lines for which shotgun LC-MS proteomic data was available was used. TMA sections were immunostained with the anti-SLC29A1 antibody, and the relative levels of expression were quantified using QuPath software (Bankhead et al., 2017). Immunohistochemistry staining was significantly correlated with LC-MS proteomic SLC29A1 expression data across the cell line panel (Spearman's correlation $P=0.003$; **Figure 55c**). Collectively, these results confirmed that the anti-SLC29A1 antibody used has high sensitivity and specificity to determine SLC29A1 protein levels in FFPE tumor samples.



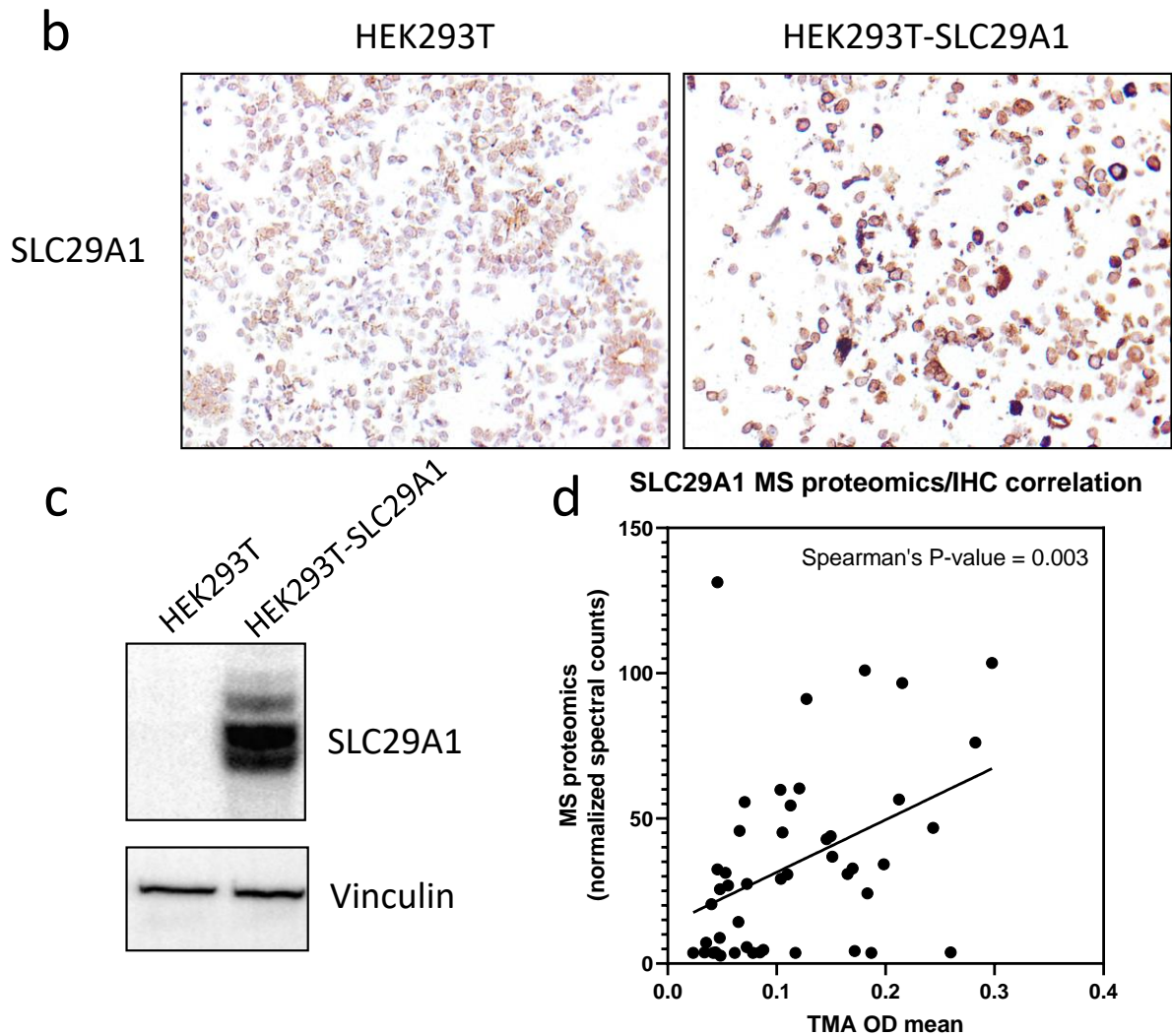


Figure 54: Validation of SLC29A1.

a) Transient transfection of the SLC29A1 overexpression model. Fluorescence microscopy images show the expression of GFP on HEK293T-SLC29A1 cells 48 h after the transfection with overexpression plasmid (pINDUCER20-SLC29A1) and pINDUCER20-EGFP. **b, c)** Immunohistochemistry (**b**) and Western blot (**c**) of parental and SLC29A1-overexpressing HEK293T cells. **d)** Spearman's correlation and the corresponding P-value of SLC29A1 expression between mass spectrometry-based proteomics and immunohistochemistry. IHC, immunohistochemistry; MS, mass spectrometry-based proteomics; TMA, tissue microarray; OD, Optical Density; DAB, 3,3'-Diaminobenzidine, used in immunohistochemical staining as a chromogen.

4.8.2.2 Validation of SLC29A1 protein expression as a biomarker of response to irinotecan-based treatment using a tissue microarray of 548 metastatic colorectal cancer patients

To further investigate the potential of SLC29A1 as a biomarker of response to irinotecan in metastatic colorectal cancer patients, a TMA containing duplicate/triplicate tumor samples from a total of 548 patients treated with irinotecan-based treatment was used. The TMA contained tumor samples collected at three different medical institutions in Madrid (n = 172), Barcelona (n = 161), and Germany (n = 215). The clinicopathological features of the patients included in the study can be found in **Table 12**.

Table 12: Baseline patients and tumors clinicopathological features from three cohorts of samples analyzed.

Characteristic variables	Cohort				P value
	Madrid (n=172)	Barcelona (n=161)	Germany (n=215)	Total (n=548)	
Gender					0.1530
Male	101 (58.7%)	98 (60.9%)	145 (67.4%)	344 (62.8%)	
Female	71 (41.3%)	60 (37.3%)	68 (31.6%)	199 (36.3%)	
Not available	0 (0.0%)	3 (1.9%)	2 (0.9%)	5 (0.9%)	
Median (range) Age (years)	66.0 (32.0-84.0)	66.7 (34.2-84.2)	/	66.6 (32.0-84.2)	0.9962
Not available	1 (0.6%)	10 (6.2%)	215 (100.0%)	226 (41.2%)	
Grade					0.0008*
Low grade	126 (73.3%)	104 (64.6%)	170 (79.1%)	400 (73.0%)	
High grade	17 (9.9%)	44 (27.3%)	43 (20.0%)	104 (19.0%)	
Not available	29 (16.9%)	13 (8.1%)	2 (0.9%)	44 (8.0%)	
T of first diagnosis					0.1817
T1+T2	7 (4.1%)	13 (8.1%)	16 (7.4%)	36 (6.6%)	
T3+T4	164 (95.3%)	133 (82.6%)	181 (84.2%)	478 (87.2%)	
Not available	1 (0.6%)	15 (9.3%)	18 (8.4%)	34 (6.2%)	
N of first diagnosis					0.3594
N0+N1	92 (53.5%)	81 (50.3%)	97 (45.1%)	270 (49.2%)	
N2+N3	75 (43.6%)	64 (39.8%)	101 (47.0%)	240 (43.9%)	
Not available	5 (2.9%)	16 (9.9%)	17 (7.9%)	38 (6.9%)	
M of first diagnosis					<0.0001*
M0	35 (20.3%)	37 (23.0%)	4 (1.9%)	76 (13.9%)	
M1	102 (59.3%)	73 (45.3%)	199 (92.6%)	374 (68.2%)	
Not available	35 (20.3%)	51 (31.7%)	12 (5.6%)	98 (17.9%)	
Regimen					<0.0001*
1	39 (22.7%)	62 (38.5%)	58 (27.0%)	159 (30.0%)	
2	37 (21.5%)	46 (28.6%)	5 (2.3%)	88 (16.1%)	
3	96 (55.8%)	45 (28.0%)	150 (69.8%)	291 (53.1%)	
Not available	0 (0.0%)	8 (5.0%)	2 (0.9%)	10 (1.8%)	
Line					0.3015
1	51 (29.7%)	34 (21.1%)	0 (0.0%)	85 (15.5%)	
2	102 (59.3%)	97 (60.2%)	0 (0.0%)	199 (36.3%)	
3, 4	18 (10.4%)	18 (12.4%)	0 (0.0%)	38 (6.9%)	
Not available	1 (0.6%)	10 (6.2%)	215 (100.0%)	226 (41.2%)	
RECIST					0.0016*
Resistant	90 (52.3%)	92 (57.1%)	145 (67.4%)	327 (59.7%)	
Sensitive	75 (43.6%)	40 (24.8%)	58 (27.0%)	173 (31.6%)	
Not available	7 (4.1%)	29 (18.0%)	12 (5.6%)	48 (8.8%)	
Site of primary tumor					<0.0001*
Colon	107 (62.6%)	81 (50.3%)	119 (55.3%)	307 (56.0%)	
Rectum	57 (33.3%)	3 (1.9%)	82 (38.1%)	142 (25.9%)	
Both	7 (4.1%)	2 (1.2%)	12 (5.6%)	21 (3.8%)	
Not available	1 (0.6%)	75 (46.6%)	2 (0.9%)	78 (14.2%)	
Median (range) OS from metastasis (months)	30.1 (0-131.5)	32.6 (0-144.8)	41.1 (0-180.3)	34.5 (0-180.3)	<0.0001*
Not available	0 (0.0%)	9 (5.6%)	0 (0.0%)	9 (1.6%)	
Median (range) OS from irinotecan-based treatment (months)	19.0 (1.0-130.8)	18.4 (0.7-126.3)	23.35 (0.7-128.4)	20.3 (0.7-130.8)	<0.0001*
Not available	8 (4.7%)	10 (6.2%)	1 (0.5%)	19 (3.5%)	
Median (range) PFS (months)	9.1 (0.7-127.5)	6.9 (0.7-124.5)	7.1 (0.1-123.8)	8.0 (0.1-127.5)	0.1911
Not available	1 (0.6%)	12 (7.5%)	5 (2.3%)	18 (3.3%)	

Notes: Regimens: 1. Irinotecan-based treatment including other approved drugs but without targeted therapy such as FOLFIRI (folinic acid, 5-FU and irinotecan)/FUIRI (5-FU and irinotecan)/XELIRI (Capecitabine [Xeloda] and irinotecan)/FOLFOXIRI (folinic acid, 5-FU, oxaliplatin, and irinotecan). 2. Monotherapy - irinotecan alone. 3. Irinotecan-based treatment, including targeted therapeutic drugs. Chi-square was conducted to compare gender, grade, TNM, regimen, line, RECIST, and site of primary tumor between the three cohorts. Mann-Whitney U test was used to compare age between the three cohorts. Log-rank test was conducted to compare overall survival (OS) from metastasis, overall survival from irinotecan-based treatment, and progression-free survival (PFS). *: P-value < 0.05.

Since SLC29A1 immunostaining was heterogeneous in different areas of the tumors (**Figure 55**), the intensity of the staining was scored by a trained pathologist using a semiquantitative scale that ranged from 0 (no staining) to 3 (highest intensity). In addition, the percentage of positively stained tumor cells in each tumor was assessed. These two values were combined to obtain the final H-score for each tumor sample in the TMA.

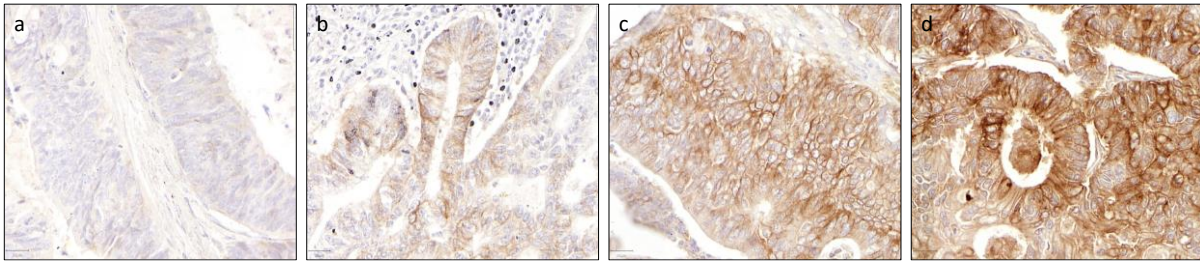


Figure 55: Representative images of the pattern of SLC29A1 immunostaining in tissue microarrays from colorectal cancer patients.

a: negative/non-detectable, score = 0; b: low intensity, score = 1; c: intermediate intensity, score = 2; d: high intensity, score = 3.

Association between SLC29A1 expression and overall or progression-free survival after irinotecan-based treatment. To investigate the possible association between SLC29A1 expression and the overall survival (OS) or progression-free survival (PFS) after irinotecan-based treatment in metastatic colorectal cancer, patients were dichotomized in two groups (high and low) according to tumor SLC29A1 expression. Stratification was conducted using the cutoff expression value that maximized the overall survival differences between the two groups (lowest log-rank test P-value; **Figure 56** and **Figure 57a**). Low levels of expression of SLC29A1 were significantly associated with shorter overall survival after irinotecan-based treatment compared to patients with high expression (P-value = 0.015). However, there was no significant difference in progression-free survival between the low- and high-level groups of patients (P-value = 0.522; **Figure 57c-d**).

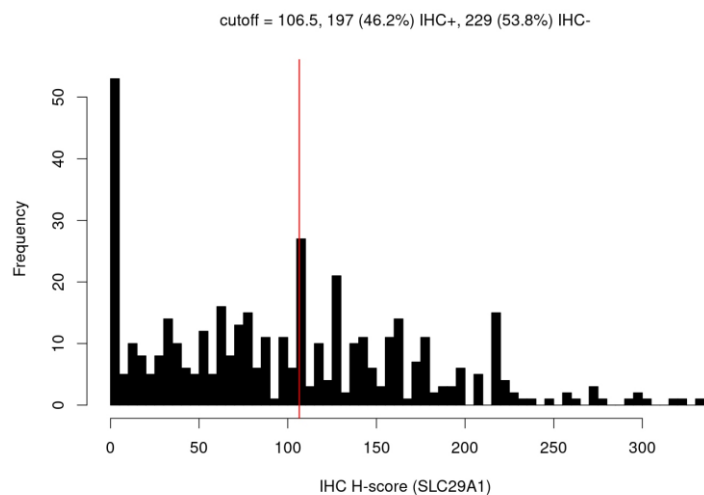


Figure 56: Distribution of SLC29A1 immunohistochemistry staining score in colorectal cancer patients from the three cohorts.

The cutoff (red line) showed maximized overall survival differences between the tumors with high and low expression (*i.e.*, lowest P value in the log-rank test).

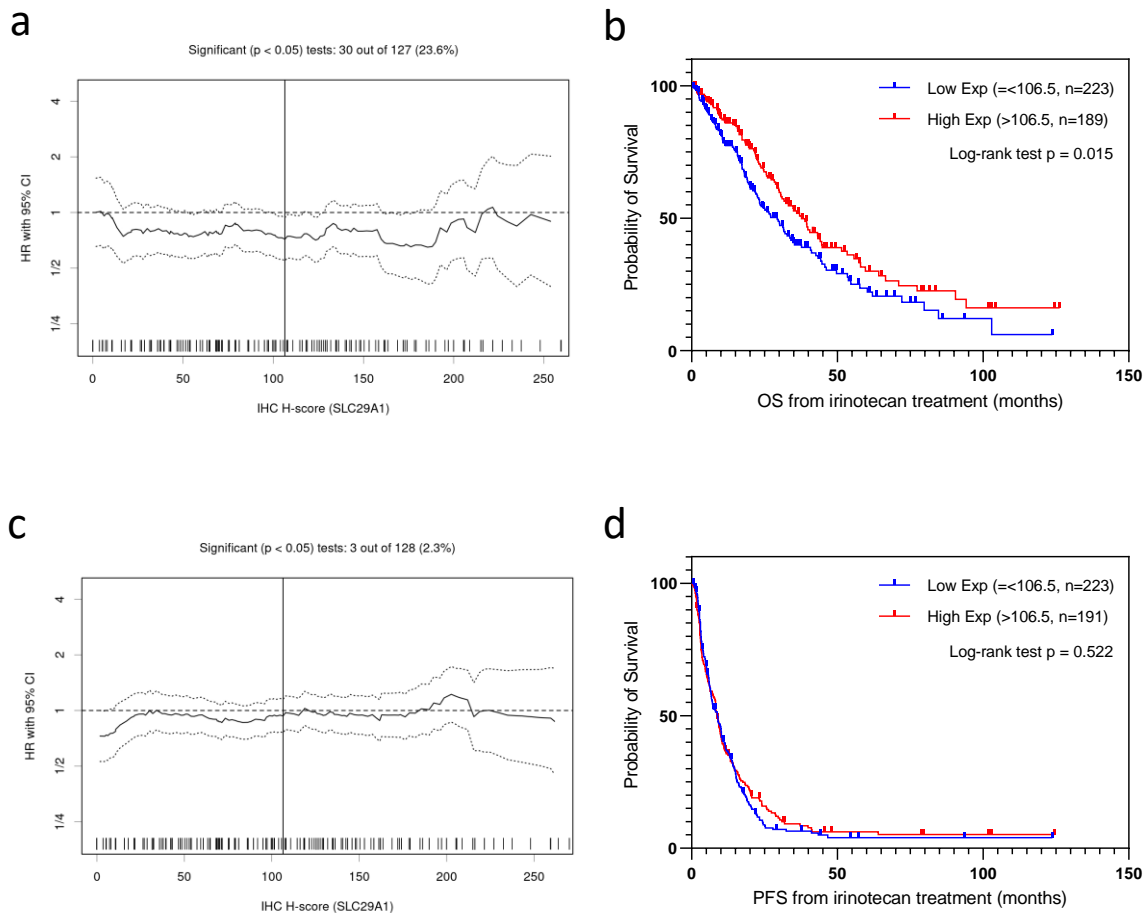


Figure 57: Survival analysis of SLC29A1 based on colorectal cancer patient samples from three cohorts. **a, c** Hazard ratio (HR) distribution on every cutoff of overall survival (a) and progression-free survival (c). **b** Overall survival (OS) by SLC29A1 immunohistochemistry staining score. **d** Progression-free survival (PFS) by SLC29A1 immunohistochemistry staining score. Cutoff = 106.5.

Multivariate analysis. To further investigate whether SLC29A1 expression was associated with the response to irinotecan-based treatment, a multivariate Cox regression (or proportional hazards regression) analysis was used. This analysis prevents the possible bias introduced by the dichotomization process of the patients into high and low-expression groups. Variables associated with the overall survival (patient age, tumor grade, treatment regimen and the site of the primary tumor; **Table 13**) or progression-free survival (patient gender, age, treatment regimen, treatment line and tumor site; **Table 13**) after irinotecan-based treatment (Fisher's exact test P-value < 0.1) when considered as a single variable were included in the respective multivariate analyses.

Table 13: Association between patients and tumors clinicopathological features and patient survival.

Characteristic variables	Survival analysis	
	OS	PFS
Gender	0.7675	0.0931*
Male		
Female		
Median (range) Age (years)	< 0.0001*	0.0050*
Grade	0.0318*	0.7259
Low grade		
High grade		
T of first diagnosis	0.0001*	0.9448
T1+T2		
T3+T4		
N of first diagnosis	0.1064	0.3748
N0+N1		
N2+N3		
M of first diagnosis	0.1032	0.2204
M0		
M1		
Regimen	0.0001*	0.0007*
1		
2		
3		
Line	0.0911*	0.0008*
1		
2		
3, 4		
Site of primary tumor	0.0132*	0.0685*
Colon		
Rectum		
Both		

OS: overall survival after irinotecan-based treatment start; PFS: progression-free survival after irinotecan-based treatment start. *: univariate Cox regression P-value < 0.1. Regimens: 1. Irinotecan-based treatment including other toxic drugs but without targeted therapy such as FOLFIRI (folinic acid, 5-FU and irinotecan)/FUIRI (5-FU and irinotecan)/XELIRI (Capecitabine [Xeloda] and irinotecan)/FOLFOXIRI (folinic acid, 5-FU, oxaliplatin, and irinotecan). 2. Monotherapy - irinotecan alone. 3. Irinotecan-based treatment, including targeted therapeutic drugs. *: P-value < 0.1.

Cox regression analysis showed that SLC29A1 expression was not independently associated with the overall survival or progression-free survival of colorectal cancer patients after irinotecan-based treatment (P-value = 0.591 and 0.145, respectively, **Table 14**). The variable independently associated with the overall patient survival was the age of the patient at the time of the diagnosis and the grade of the tumor. Older patients had shorter overall survival after

irinotecan-based treatment compared with younger patients (P-value = 0.021; **Table 14**). Patients with a higher grade of the tumor had shorter overall survival compared with lower-grade patients (P-value = 0.011; **Table 14**). On the other hand, the only variable independently associated with progression-free survival was the treatment regime (**Table 15**). Patients who received irinotecan alone had significantly shorter progression-free survival compared to irinotecan-based treatment without targeted therapy (P-value = 0.031; **Table 15**) and including targeted therapy (P-value = 0.002; **Table 15**).

Table 14: Multivariate Cox regression of SLC29A1 in overall survival analysis.

		B	SE	Wald	Sig.	OR	95% C.I. for OR	
							Lower	Upper
SLC29A1 IHC		-0.001	0.001	0.289	0.591	0.999	0.997	1.002
Age		0.025	0.011	5.337	0.021	1.026	1.004	1.048
Grade		0.648	0.253	6.540	0.011	1.911	1.163	3.140
T		0.078	0.168	0.217	0.641	1.082	0.778	1.504
Regimen	Targeted therapy included (control)			0.444	0.801			
	Irinotecan alone	-0.075	0.268	0.078	0.780	0.928	0.549	1.569
	Targeted therapy excluded	0.106	0.236	0.200	0.655	1.111	0.699	1.766
Line		0.141	0.154	0.843	0.359	1.151	0.852	1.556
Sites	Both (control)			0.748	0.688			
	Colon	-0.494	0.606	0.664	0.415	0.610	0.186	2.002
	Rectum	-0.549	0.638	0.742	0.389	0.577	0.165	2.015

Table 15: Multivariate Cox regression of SLC29A1 in progression-free survival analysis.

		B	SE	Wald	Sig.	OR	95% C.I. for OR	
							Lower	Upper
SLC29A1_IHC		-0.001	0.001	2.129	0.145	0.999	0.997	1.000
Gender		0.027	0.161	0.028	0.866	1.028	0.750	1.409
Age		0.012	0.007	2.761	0.097	1.012	0.998	1.027
Regimen	Targeted therapy excluded (control)			9.645	0.008			
	Irinotecan alone	0.471	0.218	4.672	0.031	1.601	1.045	2.453
	Targeted therapy included	-0.134	0.183	0.538	0.463	0.875	0.611	1.251
	Targeted therapy included (control)			9.645	0.008			
	Irinotecan alone	0.605	0.196	9.532	0.002	1.830	1.247	2.687
	Targeted therapy excluded	0.134	0.183	0.538	0.463	1.143	0.799	1.636
Line		0.175	0.119	2.146	0.143	1.191	0.943	1.505
Sites	Colon (control)			0.039	0.981			
	Rectum	0.017	0.175	0.010	0.920	1.018	0.723	1.433
	Both (control)	0.077	0.426	0.032	0.857	1.080	0.469	2.487

Association between SLC29A1 expression and the objective response rate after irinotecan-based treatment. The objective response rate (ORR) is used to assess the tumor burden after treatment in patients with solid tumors and is often used as a surrogate of the

percentage of patients who have a partial or complete response to the treatment. No significant difference was found in the average expression of SLC29A1 in patients showing an objective response (complete response or partial response) compared to patients that did not respond to the treatment (stable disease or progressive disease; **Figure 58a**).

Then, we determined the cutoff expression value for SLC29A1 that maximized the overall survival differences between high and low groups in the three patient cohorts. Using this cutoff value (H-score = 106.5), the protein level of SLC29A1 could not predict the response to irinotecan (P-value = 0.175; **Figure 58b**).

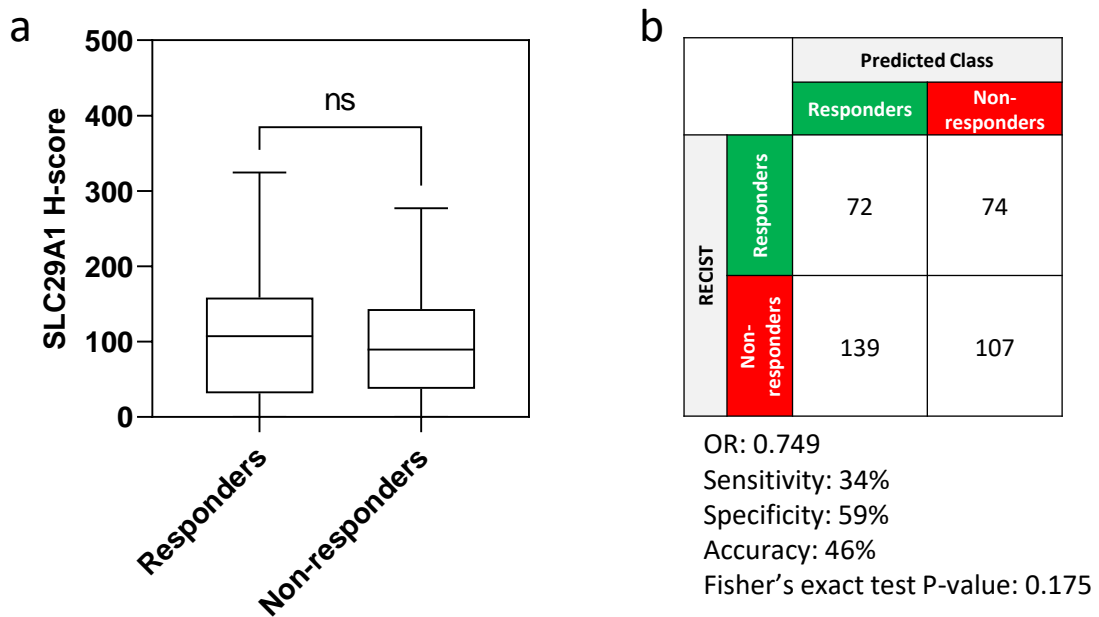


Figure 58: Association between objective response to irinotecan-based treatment and SLC29A1 protein levels.

a) The difference in tumor SLC29A1 protein levels between responder and non-responder patients. ns, $P > 0.05$. **b)** The performance of the prediction. OR: Odds ratio.

Multivariate logistic regression was used to assess whether SLC29A1 expression was independently associated with the objective response rates (ORR) after irinotecan-based treatment. Variables included in this multivariate analysis were SLC29A1 expression, patient age and treatment regimen (**Table 16**). The multivariate logistic regression results showed that SLC29A1 expression was not independently associated with the ORR after irinotecan-based treatment (P value = 0.374, **Table 17**). Older age was independently associated with higher resistance to irinotecan (P-value = 0.007, **Table 17**). In addition, irinotecan monotherapy was independently associated with an increased risk of resistance (P-value = 0.008). Irinotecan

monotherapy exhibited a significantly higher risk of irinotecan resistance compared to irinotecan-based regimens, including targeted therapy (P-value = 0.007; **Table 17**) and irinotecan-based regimens without targeted therapy (P-value = 0.008; **Table 17**).

Table 16: Association between patient and tumors clinicopathological features and irinotecan-based treatment responders and non-responders.

Characteristic variables	Non-responders (n=327)	RECIST		P value
		Mean	Responders (n=173)	
Gender				0.7686
Male	207 (63.7%)		113 (65.3%)	
Female	118 (36.3%)		60 (34.7%)	
Median (range) Age (years)		67.7 (34.2-84.2)		0.0224*
Grade				
Low grade	246 (78.8%)		128 (82.1%)	0.4637
High grade	66 (21.2%)		28 (17.9%)	
T of first diagnosis				0.1913
T1+T2	26 (8.4%)		8 (4.8%)	
T3+T4	285 (91.6%)		159 (95.2%)	
N of first diagnosis				0.2906
N0+N1	165 (53.9%)		81 (48.8%)	
N2+N3	141 (46.1%)		85 (51.2%)	
M of first diagnosis				>0.9999
M0	43 (16.2%)		24 (16.6%)	
M1	223 (83.8%)		121 (83.4%)	
Regimen				0.0565*
1	89 (27.3%)		56 (32.6%)	
2	60 (18.4%)		18 (10.5%)	
3	177 (54.3%)		98 (57.0%)	
Line				0.5883
1	45 (24.7%)		28 (30.1%)	
2	114 (62.6%)		54 (58.1%)	
3, 4	23 (12.6%)		10 (10.8%)	
Site of primary tumor				0.4286
Colon	185 (65.6%)		92 (60.1%)	
Rectum	86 (30.5%)		52 (34.0%)	
Both	11 (3.9%)		9 (5.9%)	

*: Fisher's exact test P-value < 0.1. Regimens: 1. Irinotecan-based treatment including other approved drugs but without targeted therapy such as FOLFIRI (folinic acid, 5-FU and irinotecan)/FUIRI (5-FU and irinotecan)/XELIRI (Capecitabine [Xeloda] and irinotecan)/FOLFOXIRI (folinic acid, 5-FU, oxaliplatin, and irinotecan). 2. Monotherapy - irinotecan alone. 3. Irinotecan-based treatment, including targeted therapeutic drugs. Fisher's exact test was performed to compare gender, grade, and TNM between the three cohorts. Chi-square was conducted to compare the regimen, line, and site of primary tumor between the three cohorts. Mann-Whitney U test was used to compare age between the three cohorts. *: P-value < 0.1.

Table 17: Multivariate logistic regression of SLC29A1

		B	S.E.	Wald	Sig.	OR	95% C.I. for OR	
							Lower	Upper
SLC29A1 IHC		0.001	0.001	0.791	0.374	1.001	0.998	1.004
Age		0.010	0.004	7.350	0.007	1.010	1.003	1.017
Irinotecan-based regimen	Targeted therapy excluded (control)			8.220	0.016			
	Irinotecan alone	-0.969	0.362	7.143	0.008	0.380	0.187	0.772
	Targeted therapy included	-0.028	0.242	0.014	0.907	0.972	0.605	1.563
Irinotecan-based regimen	Targeted therapy included (control)			8.220	0.016			
	Irinotecan alone	-0.940	0.346	7.376	0.007	0.390	0.198	0.770
	Targeted therapy included	0.028	0.242	0.014	0.907	1.029	0.640	1.654

Analysis of SLC29A1 expression and response to irinotecan-based treatment in the Barcelona, Madrid and Germany cohorts separately. We then investigated whether the association of SLC29A1 protein levels and the response to irinotecan-based treatment were consistent across the three cohorts.

Association between SLC29A1 expression and overall or progression-free survival after irinotecan-based treatment. As done for the analysis carried out with the three cohorts of colorectal cancer patients combined, the cutoff expression value for SLC29A1 that maximized the overall survival differences (*i.e.*, lowest log-rank P value) was calculated for each of the patient cohorts.

In the *Madrid patient cohort*, low expression of SLC29A1 was significantly associated with poorer overall survival compared to the high expression group (P-value = 0.001, **Figure 59b**). However, there was no significant difference in progression-free survival between the low- and high-SLC29A1 group (P-value = 0.102; **Figure 59d**).

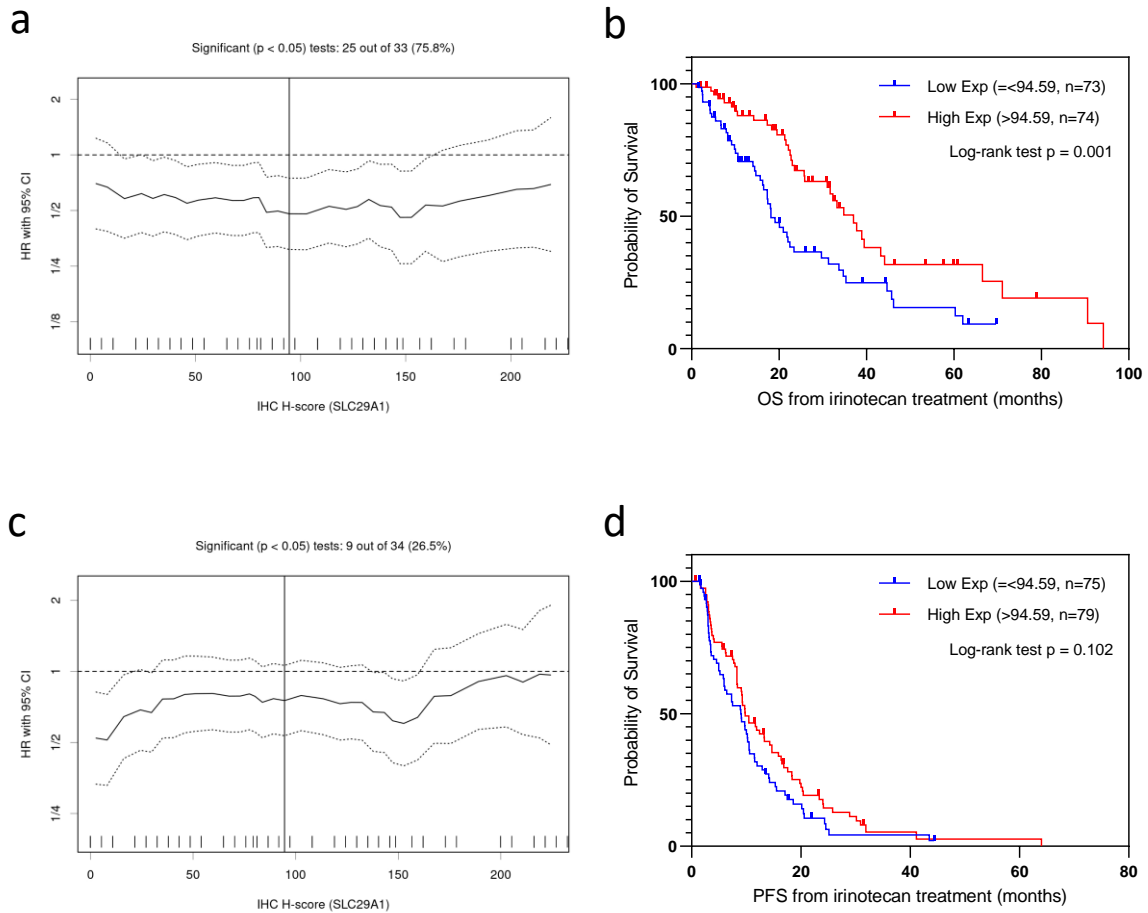


Figure 59: Survival analysis of SLC29A1 based on colorectal cancer patients samples from the Madrid patients cohort.

a, c Hazard ratio (HR) distribution on every cutoff of overall survival (a) and progression-free survival (c). **b** Overall survival (OS) by SLC29A1 immunohistochemistry staining score for the Madrid cohort. **d** Progression-free survival (PFS) by SLC29A1 immunohistochemistry staining score for Madrid patients cohort. Cutoff = 94.59.

In the *Barcelona patient cohort*, no significant associations were found between the overall survival or the progression-free survival and the expression of SLC29A1 (P-value = 0.067 and 0.112, respectively; **Figure 60b, d**).

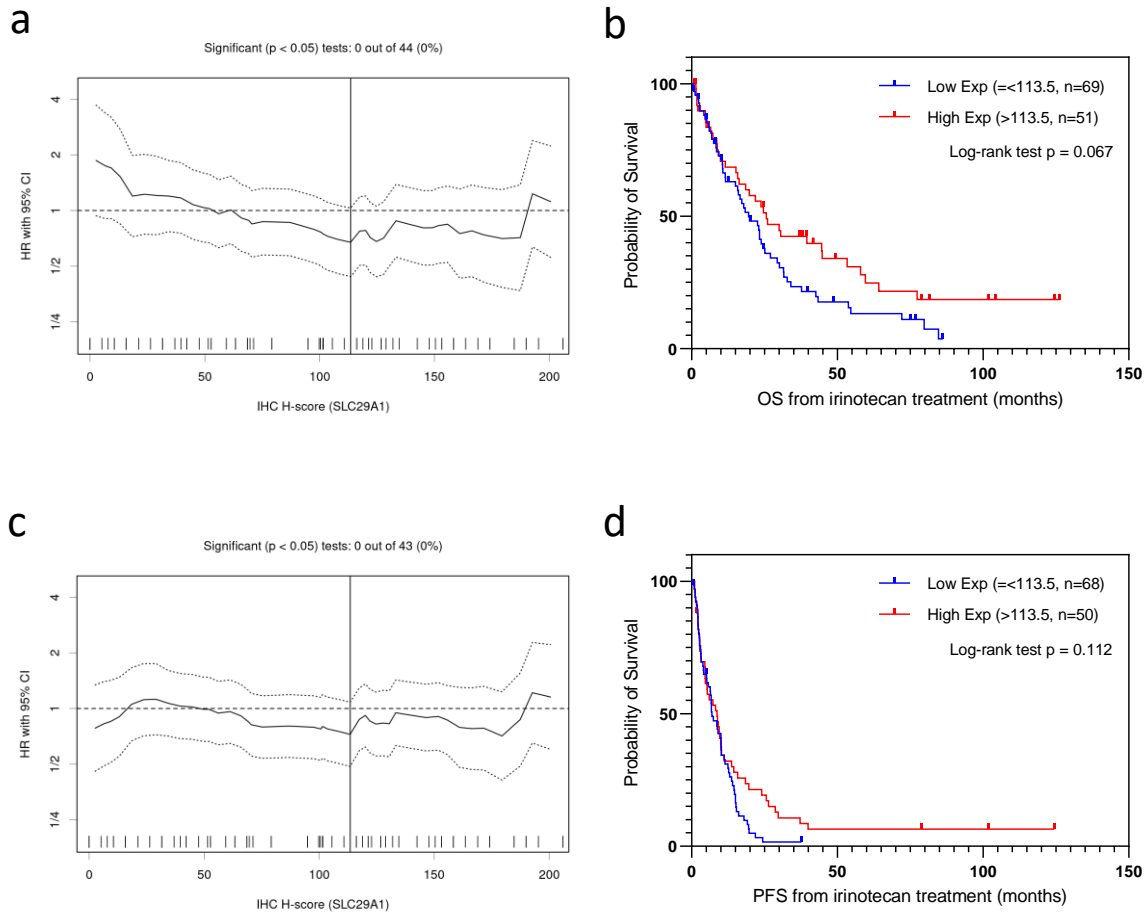


Figure 60: Survival analysis of SLC29A1 based on colorectal cancer patients samples from the Barcelona patients cohort.

a, c Hazard ratio (HR) distribution on every cutoff of overall survival (a) and progression-free survival (c). **b** Overall survival (OS) by SLC29A1 immunohistochemistry staining score for the Barcelona cohort. **d** Progression-free survival (PFS) by SLC29A1 immunohistochemistry staining score for Barcelona patients cohort. Cutoff = 113.5

In the *Germany patient cohort*, low expression of SLC29A1 was associated with significantly better overall survival than the high expression group (P-value = 0.046, **Figure 61b**), while there was no significant difference in progression-free survival between the low- and high-SLC29A1 groups (P-value = 0.367; **Figure 61d**).

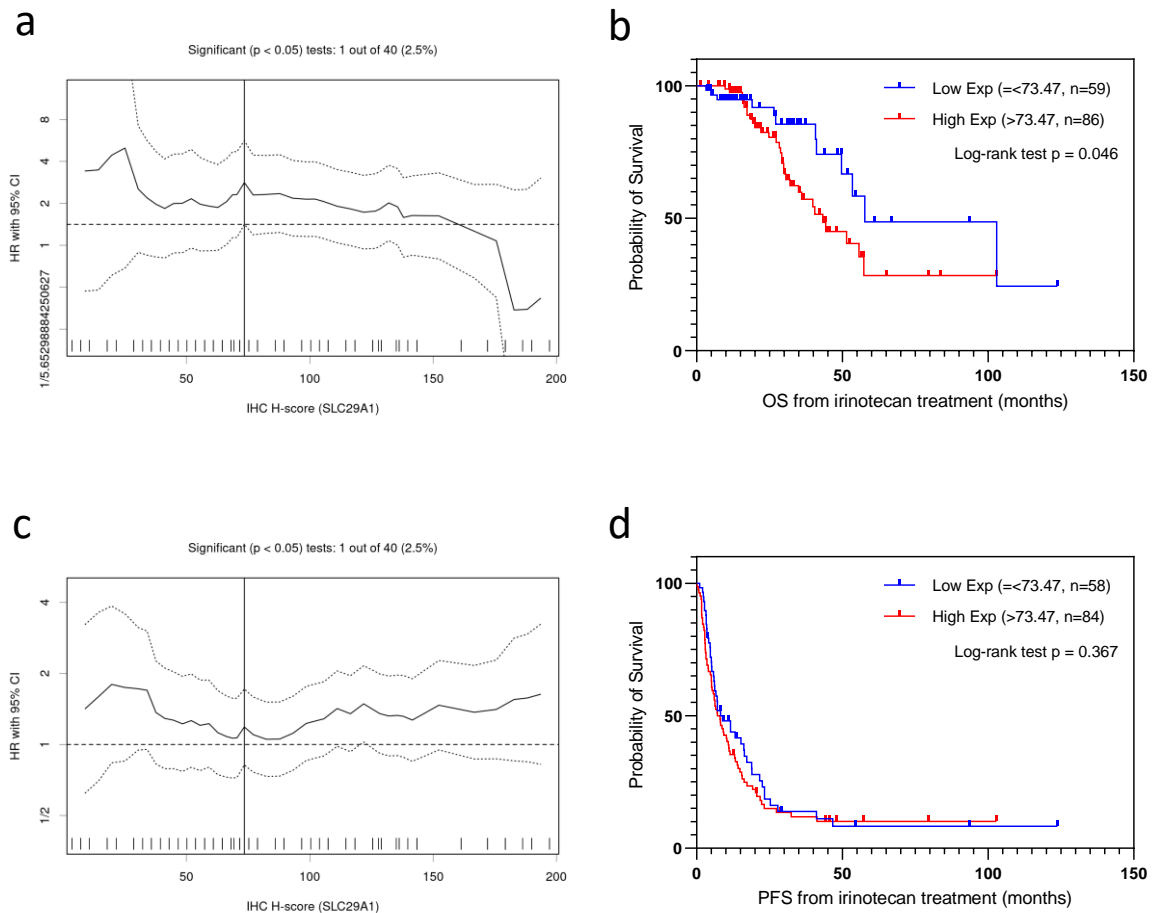


Figure 61: Survival analysis of SLC29A1 based on colorectal cancer patients samples from the Germany cohort.

a, c) Hazard ratio (HR) distribution on every cutoff of overall survival (a) and progression-free survival (c). **b)** Overall survival (OS) by SLC29A1 immunohistochemistry staining score for the Germany cohort. **d)** Progression-free survival (PFS) by SLC29A1 immunohistochemistry staining score for the Germany cohort. Cutoff = 73.47

Association between SLC29A1 expression and the objective response rate after irinotecan-based treatment. The average expression of SLC29A1 in the responder and non-responder patients in the three patient cohorts was then analyzed separately. In the **Madrid** patients cohort, SLC29A1 expression was significantly higher in patients that showed an objective response compared to non-responding patients (P-value = 0.013; **Figure 62a**). However, this difference is opposite to the expected result since the cell lines used to identify candidate biomarkers lines that showed a positive objective response according to the ccIRECIST criteria exhibited lower SLC29A1 expression compared to non-responding lines. In the **Barcelona** patients cohort, no significant difference in SLC29A1 expression was observed between responder and non-responder patients (P-value = 0.622; **Figure 62b**). Finally, in the Germany patients cohort, the

SLC29A1 protein level was significantly lower in responders compared to non-responder patients (P-value = 0.020; **Figure 62c**).

For predictive performance, the cutoff value of SLC29A1 protein expression that maximized overall survival differences in each of the 3 cohorts individually was used. In the Madrid patients cohort, the SLC29A1 protein levels could not predict the response to irinotecan (P-value = 0.033; **Figure 62d**). For the Barcelona patients cohort, the SLC29A1 protein levels in this cohort could not predict either the response to irinotecan (P-value = 0.685; **Figure 62e**). However, in the Germany patients cohort, the SLC29A1 protein level had a high performance in predicting the irinotecan response (P-value = 0.038; **Figure 62f**). Collectively, these results indicate that SLC29A1 expression was unable to consistently predict the response to irinotecan-based treatment in metastatic colorectal cancer patients.

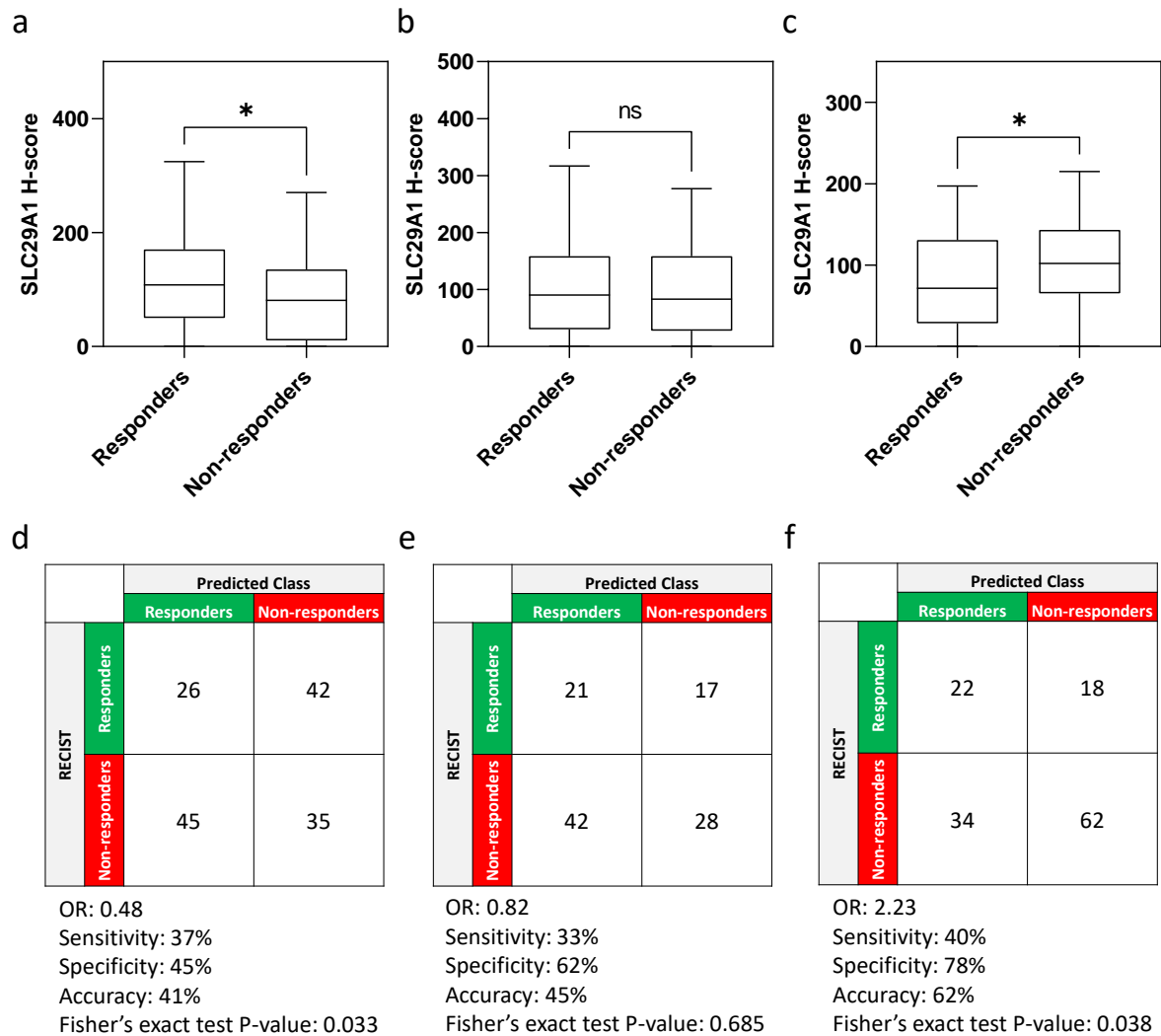


Figure 62: Association between objective response to irinotecan-based treatment and SLC29A1 expression.

a-c) shows the difference in tumor SLC29A1 protein expression between responder and non-responder patients in Madrid (a), Barcelona (b), and Germany (c) cohorts. ns, $P > 0.05$; * $P \leq 0.05$. **d-f)** shows the performance of the response prediction based on dichotomized SLC29A1 expression values of the Madrid (d), Barcelona (e), and Germany (f) cohorts. OR: Odds ratio.

Altogether, these results indicated that SLC29A1 expression could not consistently predict the response rates or the survival (OS and PFS) after irinotecan-based treatment in the three cohorts of metastatic colorectal cancer patients used in this study.

4.9 Validation of the potential of GUCY2C expression to predict response to irinotecan-based treatment for colorectal cancer patients

Guanylate cyclase 2C (GUCY2C) is a transmembrane protein located on the luminal surface of the intestinal mucosa that functions as a receptor for the endogenous peptides guanylin and uroguanylin (Potter, 2011). Activated GUCY2C converts guanosine-5-triphosphate (GTP) to cGMP and increases cGMP levels, causing the activation of downstream effectors in the cell, involving DNA damage repair and cellular metabolism (Lin et al., 2009). Studies have shown GUCY2C as a potential tumor suppressor in intestinal tumorigenesis (Gibbons et al., 2013; P. Li et al., 2007).

4.9.1 Validation of GUCY2C mRNA levels as a biomarker of response to irinotecan in an external cohort of 221 metastatic colorectal cancer patients

We first confirmed that the expression of GUCY2C shows a significant correlation at the mRNA and protein levels (**Figure 63a**) to interrogate the association of GUCY2C mRNA expression in an external validation cohort of 221 metastatic colorectal cancer patients that received irinotecan-based treatment and for which response data were available (ROC Plotter; <https://rocplot.org/>). Consistent with the findings in the panel of 90 cell lines, high levels of GUCY2C mRNA expression were significantly associated with lower response rates to irinotecan (P-value < 0.001, AUC = 0.59; **Figure 63b-c**).

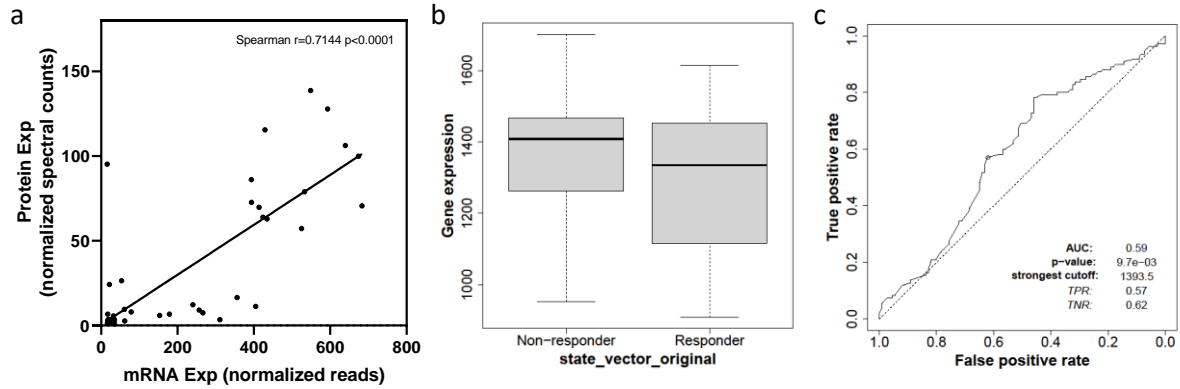


Figure 63: Validation of GUCY2C predictive power in an external cohort of 221 metastatic colorectal cancer patients treated with irinotecan.

a) Correlation between GUCY2C mRNA and protein expression in a panel of colorectal cancer cell lines (Spearman correlation). **b)** The GUCY2C mRNA expression in irinotecan responder and non-responder patients in an external cohort of 221 metastatic colorectal patients (Mann-Whitney test P-value = 0.020). **c)** ROC curve (receiver operating characteristic curve) reveals the ability of this model to classify the patients in this cohort as responders and non-responders at all expression thresholds that can be used to dichotomize patients as responders and non-responders.

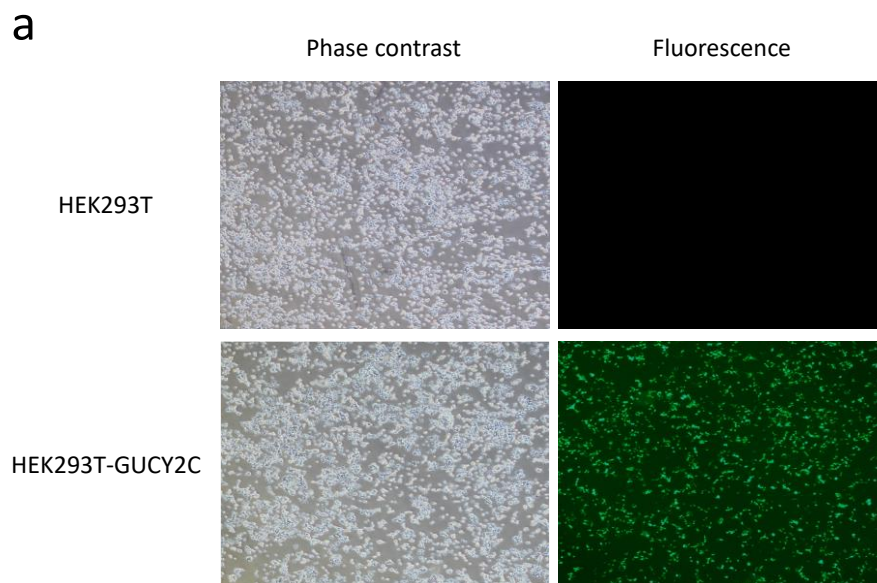
4.9.2 Validation of GUCY2C protein levels as a biomarker of response to irinotecan in a large cohort of metastatic colorectal cancer patients.

4.9.2.1 Confirmation of the specificity of the anti-GUCY2C antibody

Before measuring the levels of protein in formalin-fixed, paraffin-embedded (FFPE) tumors in a large cohort of metastatic colorectal cancer patients by immunohistochemistry, we rigorously confirmed the specificity and sensitivity of the commercial antibody selected against GUCY2C (**Figure 53**).

To this aim, first, a cell line system with forced transient overexpression GUCY2C was engineered. HEK293T cells were transiently transfected with pHAGE-GUCY2C together with pINDUCER20-EGFP (**Figure 64a**). Then, cell pellets were processed for both, protein extraction and Western blotting, and were formalin-fixed and paraffin-embedded as routinely conducted for primary tumor samples in the pathology department of clinical institutions worldwide. Then the overexpression of GUCY2C and the specificity of the antibody were tested by Western blotting in control HEK293T cells, and HEK293T cells transfected with GUCY2C. Robust overexpression of the protein of interest was confirmed by the presence of

a band of 130 KDa in the GUCY2C overexpressing cells (**Figure 64b**). Next, histological sections of the paraffin blocks containing cell pellets were used to assess the capacity of the antibody to detect GUCY2C in FFPE samples. Significantly stronger immunostaining was observed in HEK293T cells overexpressing GUCY2C compared to control HEK293 cells (**Figure 64c**). Finally, to further confirm the specificity of the antibody to detect GUCY2C on FFPE samples, a tissue microarray (TMA) containing 54 colorectal cancer cell lines for which shotgun LC-MS proteomic data was available was used. TMA sections were immunostained with the anti-GUCY2C antibody, and the relative expression levels were quantified using QuPath software (Bankhead et al., 2017). Immunohistochemistry staining was significantly correlated with LC-MS proteomic GUCY2C expression data across the cell line panel (Spearman's correlation P-value = 0.034; **Figure 64d**). These results confirmed that the anti-GUCY2C antibody has high sensitivity and specificity to determine GUCY2C protein levels in FFPE tumor samples.



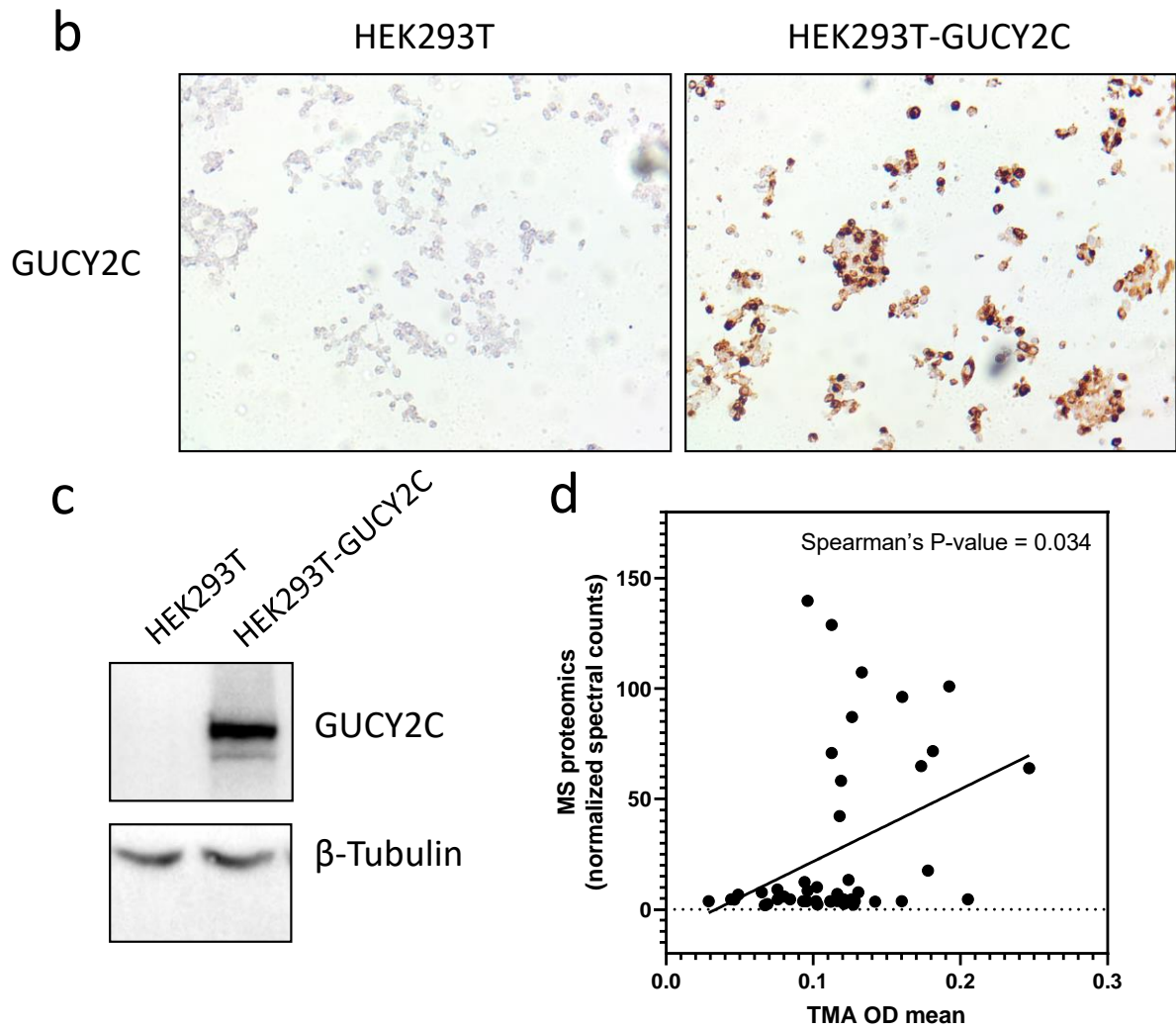


Figure 64: Validation of the anti-GUCY2C antibody in formalin-fixed, paraffin-embedded samples.

a) Transient transfection efficiency in the GUCY2C overexpression model. Fluorescence microscopy images show the expression of EGFP in HEK293T-GUCY2C cells 48 h after the transfection with overexpression plasmid (pHAGE-GUCY2C) and pINDUCER20-EGFP. **b, c)** Immunohistochemistry (**b**) and Western blot (**c**) of parental and GUCY2C-overexpressing HEK293T cells. **d)** Spearman's correlation and the corresponding P-value of GUCY2C expression between mass spectrometry-based proteomics and immunohistochemistry. MS, mass spectrometry-based proteomics; TMA, tissue microarray; OD, Optical Density.

4.9.2.2 Validation of GUCY2C protein expression as a biomarker of response to irinotecan-based treatment using a tissue microarray of 548 metastatic colorectal cancer patients

A TMA containing duplicate/triplicate tumor samples from a total of 548 patients who received irinotecan-based treatment was utilized to further evaluate the potential of GUCY2C as a biomarker of response to this drug in patients with metastatic colorectal cancer. The TMA included tumor specimens gathered from three different medical facilities in Madrid (n = 172), Barcelona (n = 161), and Germany (n = 215). **Table 12** lists the clinicopathological characteristics of the patients.

Contrary to SLC29A1, GUCY2C immunostaining was homogeneous in different areas of the tumors (**Figure 65**). The intensity of the staining was scored by a trained pathologist using a semiquantitative scale that ranged from 0 (no staining) to 3 (highest intensity).

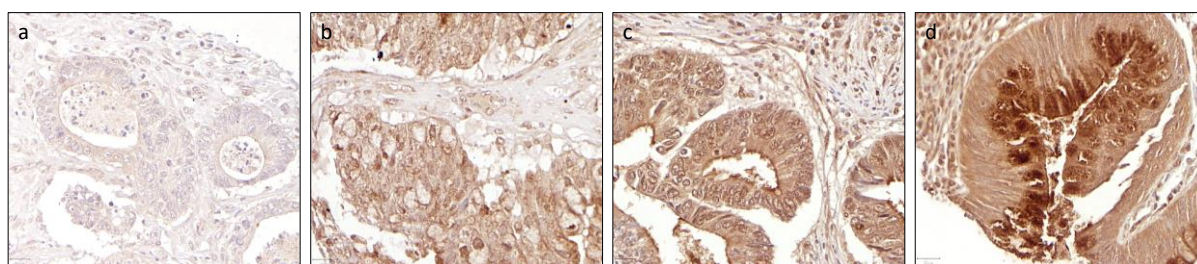


Figure 65: Representative images of the pattern of GUCY2C immunostaining in tissue microarrays from colorectal cancer patients.

a: negative/non-detectable, score = 0; b: low intensity, score = 1; c, intermediate intensity, score = 2; high intensity, score = 3.

Association between GUCY2C expression and overall or progression-free survival after irinotecan-based treatment. To investigate the possible association between GUCY2C expression and the overall survival (OS) or progression-free survival (PFS) after irinotecan-based treatment in metastatic colorectal cancer, patients were dichotomized into high and low groups according to tumor GUCY2C. Stratification was conducted using the cutoff expression value that maximized the overall survival differences between the two groups (*i.e.*, lowest log-rank test P-value; **Figure 66** and **Figure 67a**). Low levels of expression of GUCY2C were significantly associated with longer overall survival after irinotecan-based treatment compared to patients with high expression (P-value = 0.001, **Figure 67b**). However, there was no significant difference in progression-free survival between the low- and high-expression groups of patients (P-value = 0.790; **Figure 67d**).

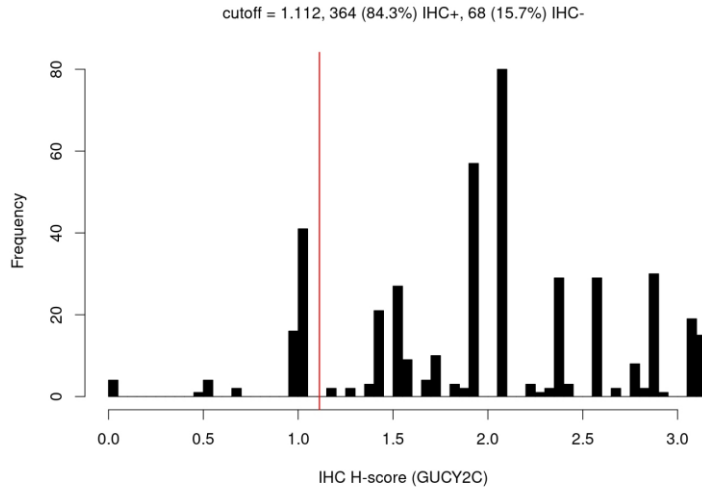


Figure 66: Distribution of GUCY2C immunohistochemistry staining score in colorectal cancer patients from the three cohorts.

The cutoff selected (red line) maximized the overall survival differences between the tumors with high and low GUC2C expression (i.e., lowest P value in the log-rank test).

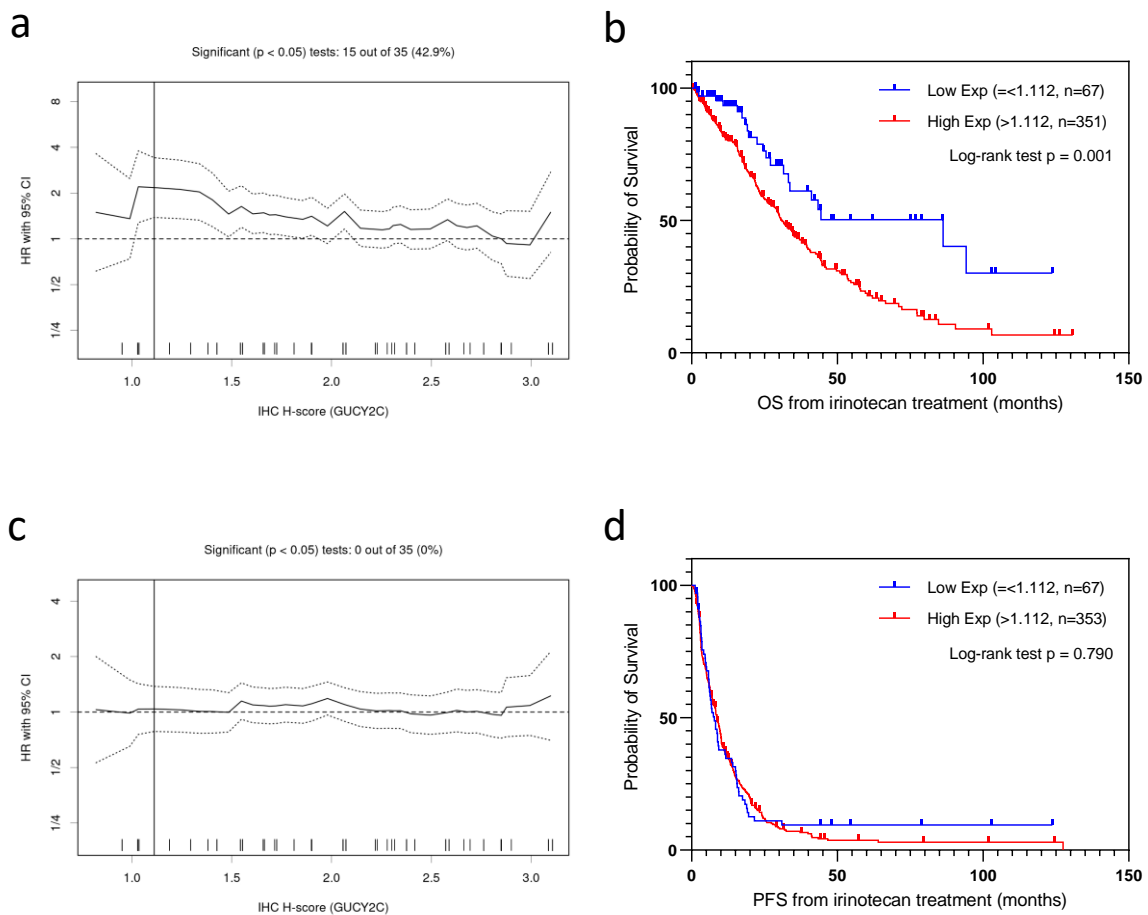


Figure 67: Survival analysis of colorectal cancer patients based on GUCY2C levels on tumor patient samples from three cohorts.

a, c) Hazard ratio (HR) distribution on every cutoff of overall survival (a) and progression-free survival (c). **b)** Overall survival (OS) and **d)** Progression-free survival (PFS) as a function of GUCY2C immunohistochemistry staining score for three patient cohorts. Cutoff = 1.112

To further investigate whether GUCY2C expression was associated with the response to irinotecan-based treatment, a multivariate Cox regression (or proportional hazards regression) analysis was used. Variables associated with the overall survival (patient age, tumor grade, treatment regimen and the site of the primary tumor; **Table 15**) or progression-free survival (patient gender, age, treatment regimen, treatment line and tumor site; **Table 15**) after irinotecan-based treatment (Fisher's exact test P-value < 0.1) when considered as a single variable were included in the respective multivariate analyses.

Cox regression analysis showed that GUCY2C expression was independently associated with the patient's overall survival after irinotecan-based treatment (P-value = 0.027, **Table 18**). No association was observed, however, between GUCY2C expression and progression-free survival (P-value = 0.615, **Table 18**). In addition, both patients at the time of the diagnosis and the grade of the tumor were independently associated with overall patient survival. Older patients had shorter overall survival after irinotecan-based treatment compared with younger patients (P-value = 0.025; **Table 18**). Patients diagnosed with a high-grade tumor had shorter overall survival compared to patients with lower-grade tumors (P-value = 0.007; **Table 18**). On the other hand, the only variable independently associated with progression-free survival was the treatment regimen (**Table 19**). Patients who received irinotecan in monotherapy had significantly shorter progression-free survival compared to patients receiving irinotecan-based regimes that incorporated targeted therapy (P-value = 0.005, **Table 19**). There was no difference in the progression-free survival of patients receiving irinotecan alone and combination irinotecan-based regimes that did not include targeted therapy (P-value = 0.074; **Table 19**) or irinotecan-based treatment with and without targeted therapy (P-value = 0.376, **Table 19**).

Table 18: Multivariate Cox regression of GUCY2C in overall survival analysis.

		B	SE	Wald	Sig.	OR	95% C.I. for OR	
						Lower		Upper
GUCY2C IHC		0.340	0.154	4.879	0.027	1.405	1.039	1.900
Age		0.025	0.011	5.019	0.025	1.025	1.003	1.047
Grade		0.679	0.253	7.219	0.007	1.972	1.202	3.235
T		0.131	0.166	0.616	0.433	1.140	0.822	1.579
Regimen	Targeted therapy excluded (control)			0.159	0.924			
	Irinotecan alone	-0.039	0.277	0.019	0.890	0.962	0.559	1.657
	Targeted therapy included	-0.093	0.236	0.156	0.693	0.911	0.574	1.447
Line		0.164	0.152	1.173	0.279	1.179	0.875	1.587
Sites	Colon (control)			3.072	0.215			
	Rectum	-0.045	0.258	0.030	0.862	0.956	0.576	1.587
	Both	1.083	0.631	2.944	0.086	2.954	0.857	10.179

Table 19: Multivariate Cox regression of GUCY2C in progression-free survival analysis.

		B	SE	Wald	Sig.	OR	95% C.I. for OR	
						Lower		Upper
GUCY2C IHC		-0.056	0.112	0.253	0.615	0.945	0.759	1.177
Gender		-0.006	0.159	0.002	0.967	0.994	0.727	1.358
Age		0.011	0.008	2.052	0.152	1.011	0.996	1.027
Regimen	Targeted therapy excluded (control)			8.044	0.018			
	Irinotecan alone	0.384	0.215	3.188	0.074	1.468	0.963	2.236
	Targeted therapy included	-0.159	0.180	0.783	0.376	0.853	0.599	1.213
	Targeted therapy included (control)			8.044	0.018			
	Irinotecan alone	0.543	0.191	8.040	0.005	1.721	1.182	2.504
Line		0.159	0.180	0.783	0.376	1.173	0.824	1.668
Sites	Targeted therapy excluded	0.184	0.122	2.258	0.133	1.202	0.946	1.527
	Colon (control)			0.143	0.931			
	Rectum	0.050	0.169	0.089	0.766	1.052	0.756	1.463
Both		0.137	0.524	0.069	0.793	1.147	0.411	3.206

Association between GUCY2C expression and the objective response rate after irinotecan-based treatment. No significant difference was found in the average expression of tumor GUCY2C in patients showing an objective response (complete response or partial response) compared to patients that did not respond to the treatment (stable disease or progressive disease; **Figure 68a**). Moreover, the protein expression of GUCY2C could not predict the objective response to irinotecan (P-value = 0.568; **Figure 68b**).

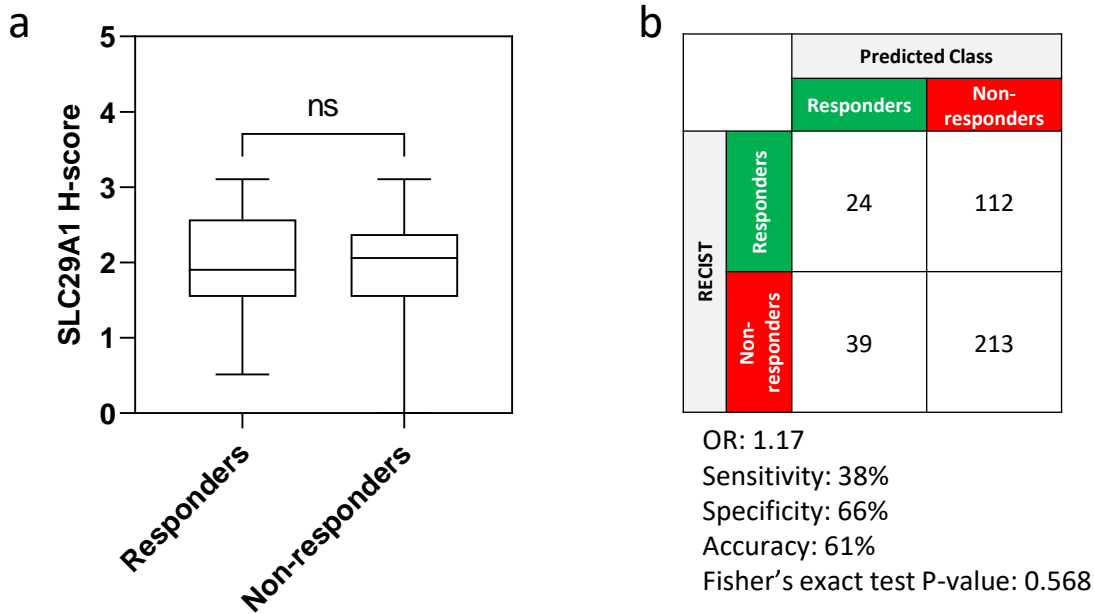


Figure 68: Association between objective response to irinotecan-based treatment and GUCY2C protein levels.

a) The difference in tumor GUCY2C protein levels between responder and non-responder patients. ns, $P > 0.05$. **b)** The performance of the prediction. OR: Odds ratio.

Multivariate logistic regression was used to assess whether GUCY2C expression was independently associated with the objective response rates (ORR) after irinotecan-based treatment. Variables included in this multivariate analysis were GUCY2C expression, patient age and treatment regimen (**Table 13**). The multivariate logistic regression results showed that GUCY2C expression was not independently associated with the ORR after irinotecan-based treatment (P -value = 0.378, **Table 20**). Irinotecan monotherapy was independently associated with an increased risk of resistance (P -value = 0.022, **Table 20**). Irinotecan monotherapy had a significantly lower response rate to the treatment compared to irinotecan-based regimens, including targeted therapy (P -value = 0.007; **Table 20**) and irinotecan-based regimen without targeted therapy (P -value = 0.024; **Table 20**).

Table 20.

Table 20: Multivariate logistic regression of GUCY2C.

	B	S.E.	Wald	Sig.	OR	95% C.I. for OR	
						Lower	Upper
GUCY2C IHC	0.180	0.204	0.776	0.378	1.197	0.802	1.787
Age	-0.012	0.013	0.819	0.365	0.988	0.964	1.014
Irinotecan-based regimen			7.677	0.022			
Targeted therapy excluded (control)							
Irinotecan alone	-0.870	0.385	5.107	0.024	0.419	0.197	0.891
Targeted therapy included	0.123	0.297	0.171	0.679	1.131	0.632	2.024
Irinotecan-based regimen			7.677	0.022			
Targeted therapy included (control)							
Irinotecan alone	-0.993	0.365	7.387	0.007	0.370	0.181	0.758
Targeted therapy included	-0.123	0.297	0.171	0.679	0.884	0.494	1.583

Analysis of GUCY2C expression and response to irinotecan-based treatment in the Barcelona, Madrid and Germany cohorts separately. We then investigated whether the association of GUCY2C protein levels and the response to irinotecan-based treatment were consistent across the three cohorts.

Association between GUCY2C expression and overall or progression-free survival after irinotecan-based treatment. As done for the analysis carried out with the three cohorts of colorectal cancer patients combined, the cutoff expression value for GUCY2C that maximized the overall survival differences (*i.e.*, lowest log-rank P value) was calculated for each of the patient cohorts.

In the *Madrid patient cohort*, low expression of GUCY2C tended to have a longer overall survival compared to the high expression group, although the difference was not statistically significant (P-value = 0.144, **Figure 69b**). There was no significant difference in progression-free survival between the low- and high- GUCY2C groups (P-value = 0.981; **Figure 69d**).

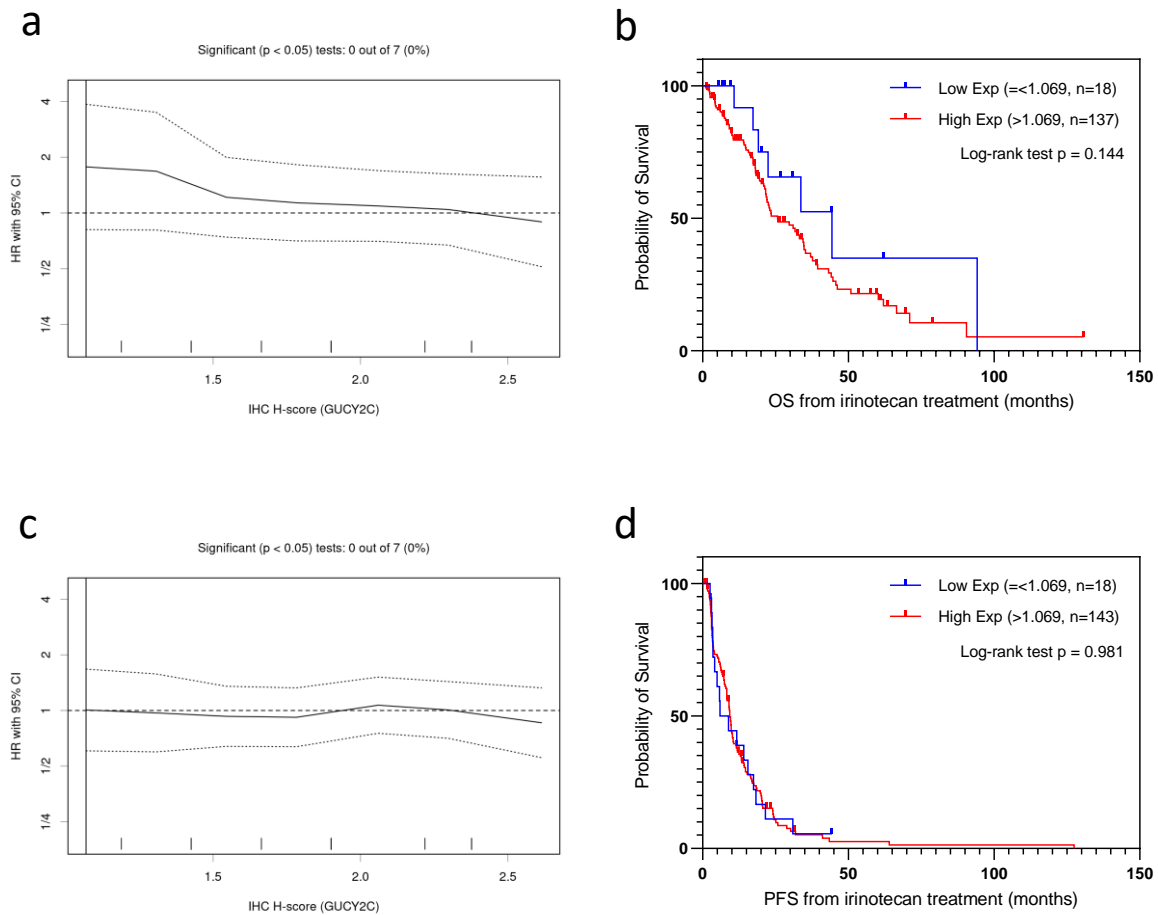


Figure 69: Survival analysis of GUCY2C based on colorectal cancer patients samples from the Madrid patient cohort.

69.

a, c) Hazard ratio (HR) distribution on every cutoff of overall survival (a) and progression-free survival (c). **b)** Overall survival (OS) by GUCY2C immunohistochemistry staining score for the Madrid cohort. **d)** Progression-free survival (PFS) by GUCY2C immunohistochemistry staining score for the Madrid patient cohort. Cutoff = 1.069.

In the *Barcelona patient cohort*, patients with low expression of GUCY2C had significantly longer overall survival compared to the high expression group (P-value = 0.007; **Figure 70b**). There was no significant difference in progression-free survival between the low- and high-GUCY2C groups (P-value = 0.991; **Figure 70d**).

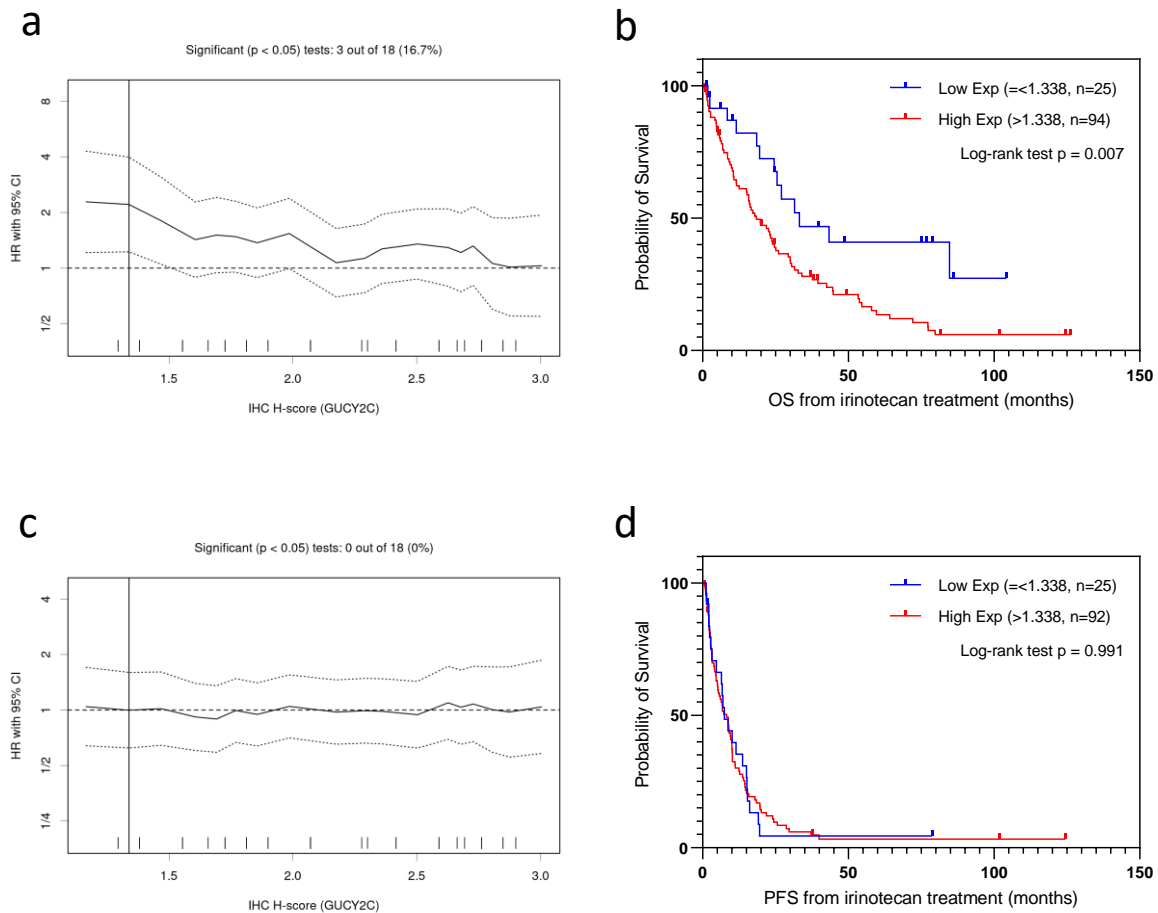


Figure 70: Survival analysis of GUCY2C based on colorectal cancer patients samples from the Barcelona patient cohort.

a, c Hazard ratio (HR) distribution on every cutoff of overall survival (a) and progression-free survival (c). **b** Overall survival (OS) by GUCY2C immunohistochemistry staining score for the Barcelona cohort. **d** Progression-free survival (PFS) by GUCY2C immunohistochemistry staining score for Barcelona patient cohort. Cutoff = 1.338.

In the *Germany patient cohort*, low expression of GUCY2C was associated with significantly longer overall survival than the high expression group (P-value = 0.043; **Figure 71b**), while there was no significant difference in progression-free survival between the low- and high-GUCY2C groups (P-value = 0.634; **Figure 71d**).

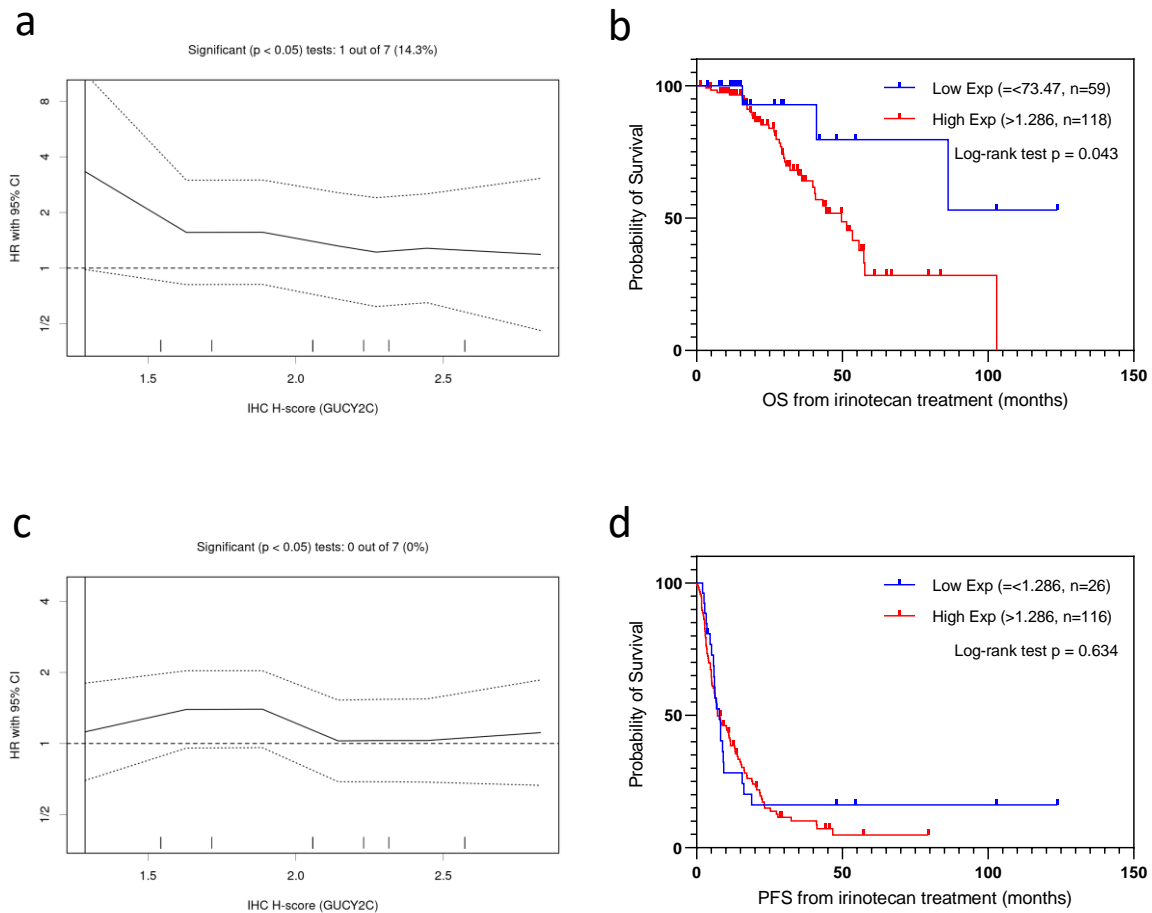


Figure 71: Survival analysis of GUCY2C based on colorectal cancer patients samples from the Germany patient cohort.

71. a, c Hazard ratio (HR) distribution on every cutoff of overall survival (a) and progression-free survival (c). **b** Overall survival (OS) by GUCY2C immunohistochemistry staining score for the Germany cohort. **d** Progression-free survival (PFS) by SLC29A1 immunohistochemistry staining score for the Germany patient cohort. Cutoff = 0.634.

Association between GUCY2C expression and the objective response rate after irinotecan-based treatment. The average expression of GUCY2C in the responder and non-responder patients in the three patient cohorts was then analyzed separately. In the **Madrid** patient cohort, no significant difference in GUCY2C expression was observed between responder and non-responder patients (P -value > 0.999 ; **Figure 72a**). In the **Barcelona** patient cohort, GUCY2C expression was significantly higher in patients that showed an objective response compared to non-responding patients (P -value = 0.025; **Figure 72b**). However, this difference was opposite to the expected result, since in the *in vitro* screening to identify candidate biomarkers, those cell lines that showed an objective response according to the ccIRECIST criteria exhibited lower GUCY2C expression compared to non-responding lines. Finally, in the **Germany**

patient cohort, GUCY2C protein expression was not significantly different in responder and non-responder patients (P-value = 0.213; **Figure 72c**).

For predictive performance, the cutoff value for GUCY2C protein expression that maximized overall survival differences in the 3 cohorts individually was used. In the Madrid patient cohort, the levels of the protein of interest could not predict the response to irinotecan (P-value = 0.608; **Figure 72a, d**). For the Barcelona patient cohort, the protein levels of GUCY2C again could not predict the response to irinotecan (P-value = 0.219; **Figure 72b, e**). In the Germany patient cohort, the GUCY2C protein expression could not predict the response to irinotecan (P-value = 0.097; **Figure 72c, f**). When considered together, these results indicate that GUCY2C expression was unable to consistently predict the response to irinotecan-based treatment in metastatic colorectal cancer patients.

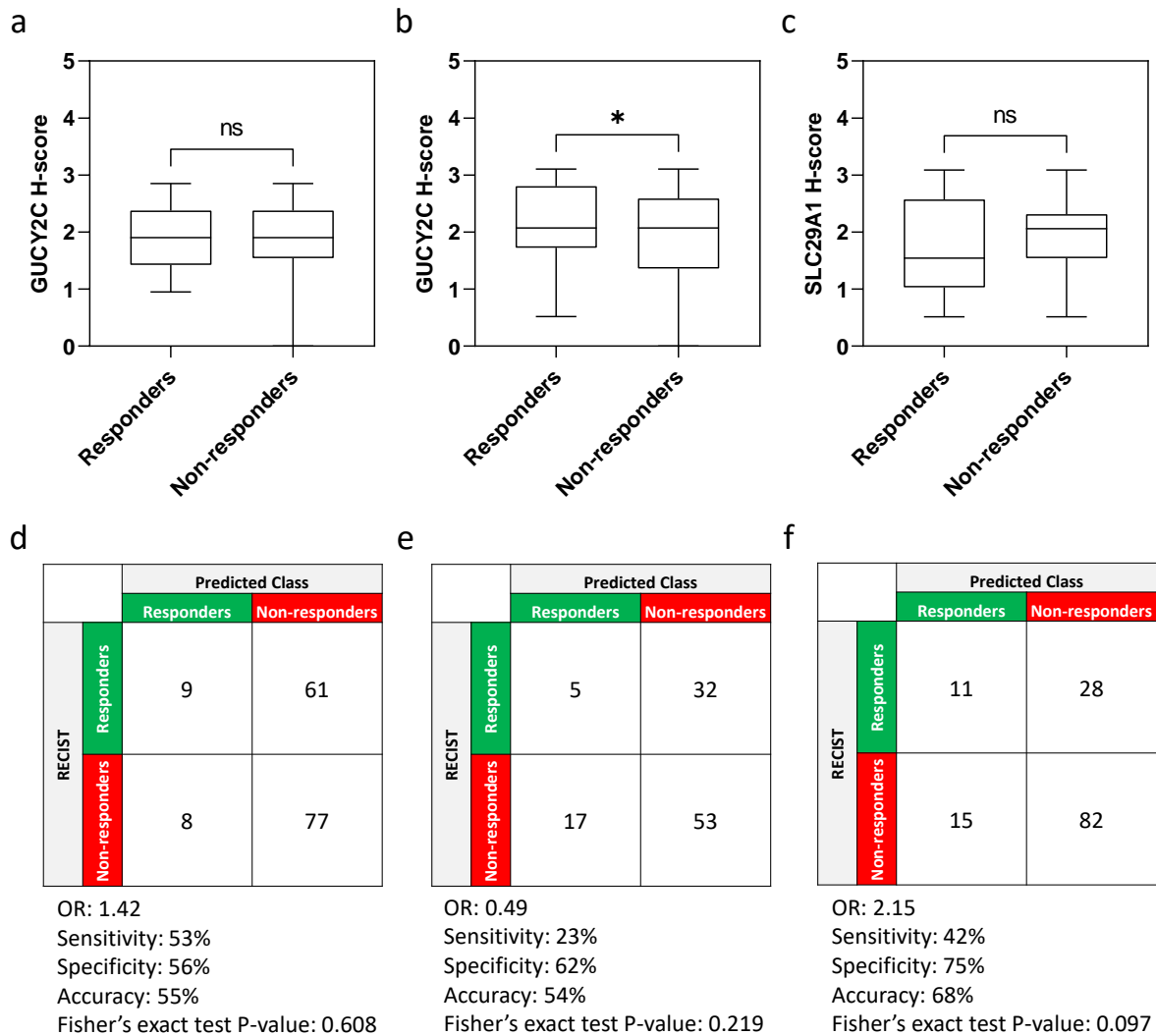


Figure 72: Association between objective response to irinotecan-based treatment and SLC29A1 expression.

72a-c) shows the difference in GUCY2C protein expression between responder and non-responder patients in Madrid (a), Barcelona (b), and Germany (c) cohorts. ns, $P > 0.05$; * $P \leq 0.05$. **d-f)** shows the performance of the response prediction based on dichotomized GUCY2C expression values of the Madrid (d), Barcelona (e), and Germany (f) cohorts. OR: Odds ratio.

Altogether, these results indicated that GUCY2C expression could not predict the response to irinotecan-based treatment in the three cohorts. However, GUCY2C expression could predict the three cohorts' overall survival collectively when using the cutoff at 1.112. When analyzing the three cohorts individually, high GUCY2C expression was statistically associated with shorter survival in Barcelona (cutoff = 1.338) and Germany cohorts (cutoff = 1.286), respectively. Although the difference between high and low GUCY2C expression was not significantly different in the Madrid cohort (cutoff = 1.069), it has the same trend as the other two cohorts.

5 DISCUSSION

Colorectal cancer is currently the third most common cancer and the second leading cause of cancer death worldwide. 5-FU, oxaliplatin, and irinotecan are the most common cytotoxic agents in colorectal cancer chemotherapy after the surgery of patients to resect the primary tumor masses. The sensitivity of a single chemotherapeutic drug is only about 10-30%. Oncologists need to use regimens including more than one agent to increase the treatment efficacy and try different regimens according to the response rate of the treatment. The discovery of biomarkers to predict drug response contributes to personalized medicine, which helps select appropriate and optimal therapies based on the molecular features of the tumors.

With the rapid development of omics in recent years, high-throughput technologies such as RNA-sequencing and mass spectrometry-based proteomics provide opportunities to identify novel predictive biomarkers. However, although many scientific publications reported the discovery of potential biomarkers, only very few biomarkers have reached clinical use or are close to entering clinical use. The most common reasons for this difficulty include the different technologies used between laboratories, insufficient validation of potential biomarkers, inadequate patient cohort size, and a low number of well-designed clinical trials (Koncina et al., 2020a).

In this study, we characterized the sensitivity to the FDA/EMA-approved chemotherapeutic agents 5-FU, oxaliplatin, and irinotecan of a large panel of colorectal cancer cell lines, identified potential predictive biomarkers for the three drugs using omics data and validated the predictive power of a subset of them in a large cohort of patient samples.

5.1 Characterization of drug sensitivity of colorectal cancer cell lines

5.1.1 Cell line as a tool to identify predictive biomarkers

We used colorectal cancer cell lines as the model to investigate potential biomarkers in this study. Cell lines have been widely used as preclinical model systems and are a fundamental tool for the discovery of novel biomarkers (Goodspeed et al., 2016; Mirabelli et al., 2019).

Compared with clinical tumor tissue, cell lines exhibit several advantages: **1)** *In vitro* cell line models allow it possible to investigate the response to a single drug. The current clinical treatment of colorectal cancer patients mainly uses the combination of several anticancer drugs. The results are affected by multiple drug interactions if clinical samples are used for studies. Indeed, in responder patients, it is not possible to know to what agent(s) they are responding. Accordingly, we treated colorectal cancer cells with a single drug and eliminated interference from other drugs. **2)** *In vitro* cell lines are not “contaminated” with non-tumor cells, such as blood/lymphatic vessels, infiltrating immune cells, and stromal cells. These non-tumor cells are usually present in different proportions in primary tumor samples, and may account for more than 50% of the cells, thus significantly complicating the interpretation of the results. **3)** Cell lines are easy to culture and have relatively stable biological features (*e.g.*, response to chemotherapeutic agents). **4)** Cancer cell lines retain most of the genetic properties of the cancer of origin (Mouradov et al., 2014). What is discovered in the cell lines reflects similar molecular properties in the clinical samples. **5)** A great number of cell lines have been derived from colorectal cancer patients, while other tumor types have substantial limitations in the number of pre-clinical models available. In addition, cancer cell lines are easy to handle, low cost, and can be easily genetically modified to confirm the results obtained.

Also, there are some disadvantages to using cell lines as a model system for biomarker discovery. **1)** The effect of the tumor microenvironment on tumor cell growth is neglected. Oxygen deficiency, nutrient limitations, accumulation of waste products, and acidity may affect the drug sensitivity of tumor cells *in vivo* (Baghban et al., 2020). **2)** The immune system impacts the drug response of the tumor, which cannot be investigated in cell line models. A study has shown that tumor-associated macrophage infiltration correlates with postsurgical adjuvant chemotherapy in patients with pancreatic ductal adenocarcinoma (A. Mantovani et al., 2017). In addition, some tumor cells escape the elimination of immune cells through the expression of PD-L1 or the secretion of suppressive cytokines, which cannot be reflected in the cell line model.

Other commonly used cancer models to identify predictive biomarkers include patient-derived xenografts (PDXs) and organoids. **Patient-derived xenografts** are established by engrafting and growing human tumor tissue/cells in immunodeficient animal hosts, *e.g.*, mice (Hidalgo et al., 2014) and recapitulate the complexity of tumor heterogeneity in the clinic. However,

compared with the cell line model, patient-derived xenografts are more time- and cost-consuming, limiting the study's sample size. Also, the original human stromal and immune cells are replaced by mouse stromal cells, losing the real clinical tumor microenvironment, *i.e.*, the adaptive immune system (Cassidy et al., 2015). **Organoids** derived from one or a few embryonic stem cells or induced pluripotent stem cells from a tissue, which have self-renewal and differentiation capacities and are supported by a 3D matrix, such as Matrigel and growth factors and inhibitors (Fujii & Sato, 2021). Tumor organoids mimic a higher degree of similarity to the clinical tumor samples than cell lines as they better model patient tissue heterogeneity (H. Yang et al., 2018). Nevertheless, some disadvantages, such as the lack of robust culturing strategies, which may significantly affect drug response (Huch et al., 2017), and varied growth dynamics hamper the use of these models.

Each cancer model has its advantages and limitations. The widely used omics technologies and more cell lines omics data available online allowed us to identify new predictive biomarkers by exploring the large colorectal cancer cell line collection available from our laboratory (n = 91).

5.1.2 Characterization of cell lines' sensitivity to 5-FU, oxaliplatin, and irinotecan

5.1.2.1 Inconsistency in large pharmacogenomic studies

In 1988, the first large-scale pharmaceutical screening in cancer therapy called NCI-60 was generated, including 60 human cancer cell lines (Alley et al., 1988). Cell line genomic data was also developed and resulted in a better understanding of the relationship between the molecular features of cancers and treatment outcomes. Subsequently, the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012a), including the sensitivity of 1036 cell lines against 24 drugs, the Genomics of Drug Sensitivity in Cancer (GDSC) (W. Yang et al., 2012) integrating the sensitivity of 1,001 cancer cell lines to 265 anti-cancer drugs were published in 2012. The Cancer Therapeutics Response Portal (CTRP) dataset measured the sensitivity to 481 small molecules in 224 cancer cell lines and was released in 2013 (Basu et al., 2013). Many

laboratories use these databases to guide research on cancer's molecular mechanisms and biomarker identification.

However, the reported drug sensitivity measures between the most commonly used cell line datasets, CCLE and GDSC, were discrepant, and the difference propagates into the gene-drug correlations found by different datasets (Haibe-Kains et al., 2013). In addition, we showed the inconsistency of the sensitivity data of 5-FU, oxaliplatin and irinotecan between three datasets: CCLE, GDSC, and CTRP. The reasons for this difference likely include cell culture seeding density, plating efficiency, growth rate, and time of exposure to the drug, which might affect cell metabolism and drug responsiveness. In addition, three datasets used different cell viability assays. CCLE and CTRP used Cell Titer Glo, an ATP detection cell viability assay for drug screening, which is easy to manipulate but could be interfered with by cell metabolism. GDSC used Syto60 cell viability assays containing a fluorescent nucleic acid stain, which could stain the DNA of intact but not viable cells (Hatzis et al., 2014). As a comparison, we used Sulforhodamine B (SRB) protein stain assay, which is independent of cellular metabolism and displays a high reproducibility despite the need for more manipulation steps, such as fixation and washing steps.

5.1.2.1 Sulforhodamine B (SRB) protein stain assay

Cellular response profiling of tumor models is important for predictive biomarker discovery. Dose-response analyses have been used for decades to screen cell line sensitivity to drugs. As direct cell line counting is time-consuming, cell viability or density is thought to be the primary surrogate of cell line counting. Since 1983, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) assay was a standard method for the determination of cell viability (Mosmann, 1983). This assay measures cell viability by detecting reductive activity in the mitochondria of living cells, which have dehydrogenases to convert the tetrazolium compound to water-insoluble formazan crystals. The crystal is purple and can be detected by spectrophotometer after solubilization.

Sulforhodamine B (SRB) protein stain assay, which was developed in 1990 (Skehan et al., 1990), is another commonly used cell dye-based cytotoxicity assay for screening cell line drug

sensitivity. The assay relies on the ability of SRB to bind to amino-acid residues of cells that have been fixed to culture plates. Compared to the MTT assay, SRB is not dependent on cell viability; instead, it measures total cellular protein content, which is proportional to the cell number.

Each cell viability cytotoxicity assay has its advantages and disadvantages. Compared to other methods, the SRB assay shows good linearity with cell number, has high reproducibility and sensitivity threshold (able to be sensitive to detect very few amounts of cells) and is independent of cellular metabolism (Aslantürk, 2018). These merits make the SRB assay ideal for our research. However, as a colorimetric assay, evaluating the number of cells to be seeded is important, as the linearity between cell number and absorbance can be lost at high cell densities (Skehan et al., 1990). To solve this, we did a growth test for every cell line to ensure the cell density was below 80% before fixation in the untreated control cells.

5.1.2.2 Cell line adapted response evaluation criteria in solid tumors (ccIRECIST)

In addition to the cell viability method, the choice of dose-response curves and related measures of drug response is also vital to cell line characterization. The most widely used and routinely accepted methods for measuring drug sensitivity are the IC_{50} (Inhibitory concentration 50) value, and the area under the dose-response curve (AUC), used in both CCLE and GDSC repositories. IC_{50} indicates the concentration of the drug that leads to inhibition of 50% growth over a period of time after exposure to the drug. AUC indicates the area under the dose-response curve. GI_{50} (growth inhibition 50) is similar to IC_{50} , precisely the drug concentration resulting in a 50% reduction in the net increase in control cells during the drug incubation. Other measurements include EC_{50} (half maximal effective concentration), Hill slope (steepness of the slope observed in the exponential part of the curve) and E_{max} (maximum effect of the drug is achieved).

However, no standard measures have been shown to reflect the response as it is routinely done in the clinical setting, and each parameter may convey different information (Fallahi-Sichani et al., 2013). Most measurements are dependent on and sensitive to the drug dose range, which

was not the same between different cell line repositories. In addition, due to the large number of compounds to be tested in the datasets, limited doses range were used in these studies. NCI-60, CCLE, and GDSC used 5, 8, and 9 different drug concentrations, respectively.

In this study, we proposed a new alternative classification system adapted from Response Evaluation Criteria In Solid Tumors (RECIST), which is used in the clinic to evaluate the patient's response to the treatment. We called the classification criteria cclRECIST. The concordance of biomarkers identified by CL-RECIST classification with the clinical study provides a practical new approach to maximize the chances of validating biomarkers in the clinical samples.

Overall, the current study has the following advantages compared with previous large high throughput studies in terms of characterization of cell lines sensitivity to drugs: 1) used an optimized number of cells inoculated in the plates for each cell line depending on the doubling time of individual cell lines, to control confounding factors such as different growth rates; 2) selected different drug concentrations for screening by detailed examination and optimization of response curves for each drug; 3) treated cells with 12 different concentrations of drug to more accurately get the dose-response curve; 4) kept high reproducibility between experimental replicates by calculating the coefficient of variation; 5) set a control group to get the net increase/decrease of cell growth, reducing the effect of the different growth rate of the individual cell line; 6) a large number of colorectal cancer cell lines ($n = 92$) were included, compared to NCI-60 ($n = 9$), CCLE ($n = 13$), GDSC ($n = 46$), and CTRP ($n = 47$); 7) use sulforhodamine B (SRB) protein stain assay as the viability assay, which has high reproducibility and is independent of cellular metabolism.

5.2 Consensus molecular subtype 1 and microsatellite instability are associated with a higher response rate to irinotecan

Our results showed that cell lines with microsatellite instability (MSI) and consensus molecular subtype 1 (MSI Immune) have a significantly higher response rate to irinotecan than microsatellite-stable cell lines. DNA mismatch repair errors often occur at areas of DNA repeats (microsatellites), and these mutations accumulate in colorectal tumors with defects in

mismatch repair mechanisms, resulting in a phenotype known as microsatellite instability (MSI). The increased sensitivity to irinotecan in MSI tumors may be caused by irinotecan and active metabolites SN-38 breaking the DNA double-strand, which cannot be repaired because of the deficiency in the mismatch repair system (Bhonde et al., 2010). A study involving five colorectal cancer cell lines reported consistent results with ours (Vilar et al., 2008). In addition, Fallik et al. (Fallik et al., 2003) showed that microsatellite instability predicts improved response in metastatic colorectal cancer patients involving 72 patients treated with irinotecan alone or combined with 5-FU. In another clinical study including 723 colorectal cancer patients, Bertagnolli et al. (Bertagnolli et al., 2009) found that microsatellite instability predicts improved response in stage III patients treated with irinotecan-based chemotherapy. However, this discovery was not validated in a study with 297 metastatic colorectal cancer patients (J. E. Kim et al., 2011) and neither in a trial containing 1,254 stage II and III colorectal cancer patients (Klingbiel et al., 2015).

Regarding the association with the consensus molecular subtypes (CMS) in colorectal cancer, a clinical trial showed that the prognosis of oxaliplatin-based treatment in metastatic colorectal cancer in CMS1 was poor compared to other CMS subtypes. In contrast, irinotecan-based therapy was effective in all subtypes of CMS (Yuki et al., 2020). Although we also found CMS1 cell lines have a poor response rate to oxaliplatin, this difference did not reach statistical significance. Considering that CMS1 subtypes include most cases of microsatellite instability (MSI), we think these findings provided evidence that MSI/CMS1 predicted a better response to irinotecan than MSS and other CMS subtypes.

Currently, it is recommended that all metastatic colorectal cancer patients undergo testing for microsatellite status to identify patients who could be treated with anti-PD-1 immunotherapy (Marcus et al., 2019), but it cannot guide decisions on the choice of chemotherapy (Overman, Ernstoff, et al., 2018). Data from these *in vitro* and clinical studies regarding the predictive value of microsatellite status on irinotecan response are inconsistent, indicating investigations of the effects of these agents are challenging, partly due to differences in the immune system and tumor microenvironment between patients, the different stages of patients in the studies, and differences *in vitro* and patients' drug metabolism. The association between microsatellite instability and the benefit of irinotecan-based adjuvant chemotherapy in patients with primary colorectal cancer is not confirmed and needs further investigation. In conclusion, our result

suggested an increased sensitivity to irinotecan in microsatellite instable colorectal cancer cell lines.

5.3 *KRAS* wild type is associated with a higher response rate to irinotecan

We observed that *KRAS* wild type was associated with a higher response rate to irinotecan in the cell line model. *KRAS* is a downstream effector of epidermal growth factor receptor (EGFR) in the RTK-RAS pathway. *KRAS* is mutated in approximately 36% of colorectal cancer (Chang et al., 2013) and produces a constitutively active Ras protein, leading to constitutive activation of the RTK-RAS pathway. *KRAS* mutational status was widely investigated and approved as a predictive biomarker to identify patients most likely to benefit from anti-EGFR targeted therapy (Amado et al., 2008; Bokemeyer et al., 2011; van Cutsem et al., 2011). However, the role of *KRAS* status in response to irinotecan has rarely been investigated. Some studies showed metastatic colorectal cancer patients with wild-type *KRAS* achieve considerable benefit from irinotecan-based treatment (André et al., 2013; Loupakis et al., 2009; Tejpar et al., 2008). Nevertheless, some independent studies failed to demonstrate a relevant predictive impact (Ergun et al., 2019; Richman et al., 2009). Notably, all these irinotecan-based treatments included anti-EGFR targeted therapy, which inevitably affected irinotecan's contribution. Spindler et al. found that in metastatic colorectal cancer patients treated with irinotecan alone, *KRAS* mutations detected in plasma cell-free DNA were a predictive and prognostic biomarker (Spindler et al., 2013), which was in concordance with our finding. Subsequently, in 2018 and 2020, Yao et al. (Yao et al., 2018) and Garcia-Carbonero et al. (Garcia-Carbonero et al., 2020) reported mutant *KRAS* plasma DNA as a predictor of response and prognosis to irinotecan in metastatic colorectal cancer patients.

Since irinotecan is often used in combination with anti-EGFR in colorectal cancer patients, the possible association of irinotecan and its active metabolites SN-38 with RTK-RAS pathways has been investigated. In 2004, studies reported that anti-EGFR targeted therapy overcome acquired resistance to irinotecan in metastatic colorectal cancer (Cunningham et al., 2004). Yashiro et al. showed that EGFR inhibition enhances the efficacy of SN-38 by increasing the

levels of apoptosis-related molecules, caspase-6, p53, and DAPK-2 (Yashiro et al., 2011). In 2013, Petitprez and K Larsen indicated irinotecan resistance was accompanied by the upregulation of EGFR (Petitprez & K Larsen, 2013). In 2016, Shao et al. found that EGFR inhibitor synergizes with SN-38 to inhibit the proliferation of hepatocellular carcinoma cells *in vitro* (Shao et al., 2016). Furthermore, the RTK-RAS pathway was reported to be activated by irinotecan monotherapy (Mahli et al., 2018; Mazard et al., 2013). In conclusion, although the underlying mechanism was not clearly identified, these studies suggested that alterations in *KRAS* status and RTK-RAS pathways could affect the cell's response to irinotecan.

Our study reinforced the current knowledge of *KRAS* status as a predictive biomarker in colorectal cancer patients. We showed that *KRAS* mutation status was associated with a better response of cell lines to irinotecan. It is important to note that *KRAS* status was evaluated in the clinical setting to guide anti-EGFR therapy but not for irinotecan. Our finding provides evidence that evaluation of *KRAS* status could also be valuable in selecting irinotecan-based or monotherapy treatment in colorectal cancer.

5.4 Omics data merging

Comparison of gene expression profiles between diseased and normal tissue or treated and controls provides a deep understanding of disease mechanisms and response to therapeutic interventions. However, the number of samples is limited in most studies due to the elevated cost of expression profiling techniques. The systematic differences between the measurements done at different times or by different labs make the comparison of samples in different batches challenging due to what is known as the “batch effect”. Different approaches to “normalize” genome-wide expression profiling exists. In 2007, Johnson et al. introduced an empirical Bayes method called Combatting batch effects when combining batches of microarray data (ComBat) (W. E. Johnson et al., 2007). Subsequently, this method was applied in RNA-sequencing (Graf et al., 2017) and mass spectrometry-based proteomics (A. H. Lee et al., 2019b; Seyfried et al., 2017).

Here, Principal Component Analysis (PCA) and hierarchical clustering analysis methods were used to evaluate the effect of batch among the different datasets available for the study and also

for the elimination upon data curation. PCA is a method commonly used for reducing the dimensionality of large data sets by transforming multivariable into few principal components to obtain lower-dimensional data while still containing most of the information in the large data sets (Jolliffe & Cadima, 2016). Principal component analysis restricted to two dimensions is the most commonly used. The first principal component (PC1) score summarizes the greatest variability that presents the most variability among the samples included in the test. PC2 score indicates the second-greatest variability among samples. When plotted, PC1 and PC2 are represented in X and Y axis, respectively. Hierarchical clustering is an unsupervised machine-learning algorithm that groups similar unlabeled objects into groups called clusters, usually presented in a dendrogram (Nielsen, 2016). In the present study, low PC1 and PC2 scores and undetectable batch clustering in the hierarchical clustering analysis dendrogram indicated a successful removal of the between-batch bias across datasets.

5.5 Association of OMICS data with the response.

We identified 16 putative predictive biomarkers for irinotecan response by mass spectrometry-based proteomics. Among these potential biomarkers identified as a result of the cell line model, we first selected SLC29A1 and GUCY2C for validation in colorectal cancer patients' tumor samples.

Solute Carrier Family 29 Member 1 (SLC29A1) is an equilibrative nucleoside transporter localized to the plasma and mitochondrial membranes and mediates the cellular uptake of nucleosides. SLC29A1 is required to uptake nucleotide synthesis in the cells that depend on the salvage pathway nucleotide synthesis. It also mediates the transport of nucleotide analog chemotherapeutic agents such as 5-FU (Baldwin et al., 2004). Its associations with the response to the nucleotide analog have been well studied. Its role in drug resistance is varied in different tumor types. Low SLC29A1 levels were associated with drug resistance in leukemia and lung cancer as it plays a vital role in cellular uptake (Y. Huang, Anderle et al., 2004; L. Li et al., 2012). On the other hand, results from a functional study indicated that the levels of SLC29A1 correlated inversely with the response to 5-FU in colorectal cancer cells, and SLC29A inhibitor (NBMPR) increased the cytotoxicity effects (Phua et al., 2013). Overexpression of SLC29A1 was reported in colorectal cancer tumors compared to adjacent normal tissues (Kunicka et al.,

2016). In addition, SLC29A upregulation also contributed to cisplatin resistance in colorectal, astroglial, and breast cancer cells (Gotovdorj et al., 2014). There is currently no study reporting the association between SLC29A1 and irinotecan response. Since previous studies showed that SLC29A1 plays an important role in transporting nucleotides, and impacts cell viability, it could influence other anti-cancer drugs, not only antimetabolite nucleotide drugs. Its role in the resistance to platinum compounds also suggests it could contribute to the irinotecan resistance in colorectal cancer.

Guanylate cyclase 2C (GUCY2C) is a transmembrane protein located on the luminal surfaces of the intestinal mucosa that functions as a receptor for endogenous peptides guanylin and uroguanylin (Potter, 2011). It can also be activated by the heat-stable *Escherichia coli* enterotoxin (Arshad & Visweswariah, 2012). Activated GUCY2C converts guanosine-5'-triphosphate (GTP) to cGMP and thus increases cGMP levels which cause the activation of downstream effectors in the cell involving DNA damage repair and cellular metabolism (Lin et al., 2009). Bashir et al. found that loss of GUCY2C ligands-induced silencing of GUCY2C signaling contributed to the tumorigenesis of colorectal cancer (Bashir et al., 2019). Conversely, GUCY2C was also reported to be overexpressed in colorectal cancer (Magee et al., 2016). In addition, the detection of GUCY2C expression in lymph nodes is indicative of the presence of micrometastasis and was an independent marker of prognosis. Colorectal cancer patients with higher GUCY2C expression level in lymph nodes was associated with earlier time to recurrence and reduced disease-free survival (Waldman et al., 2009). In recent years, GUCY2C-based precision therapies have been explored as a reliable targeted approach to cancer treatment (Bashir et al., 2019; Magee et al., 2018). However, no evidence has been reported of an association between GUCY2C expression levels and the response to irinotecan in colorectal cancer. In conclusion, since GUCY2C plays an essential role in cellular homeostasis and has been found as a prognostic biomarker in colorectal cancer, it could impact the response to chemotherapy, like irinotecan.

5.6 Validation of potential biomarkers

Although numerous candidate biomarkers have been investigated and reported, there are currently few approved predictive biomarkers of treatment response in colorectal cancer. A

qualified biomarker should pass through three phases of the development process: preanalytical and analytical validation, clinical validation, and clinical utility (Masucci et al., 2016). The sensitivity, specificity, and usability of each will determine whether the biomarker will be translated into an effective biomarker for patients (J. W. Lee et al., 2006).

To proceed with the validation of the most promising biomarkers results from an external online dataset ROC Plotter (<https://rocplot.org/>), including 221 colorectal cancer patients treated with irinotecan, were used. Interestingly, high mRNA expression levels of SLC29A1 and GUCY2C were associated with a lower response rate to irinotecan, which was in accordance with the findings in our cell line panel.

Therefore, we performed analytical validation for SLC29A1 and GUY2C. A cell line overexpression model was initially used to validate the sensitivity and specificity of the related antibodies. Western blot and immunohistochemistry results showed that the antibodies recognizing SLC29A1 and GUCY2C were highly sensitive and specific to the target antigen. Subsequently, we validated these two proteins in formalin-fixed, paraffin-embedded (FFPE) tissue microarrays from 91 colorectal cancer cell lines using immunohistochemistry. GUCY2C and SLC29A1 antibodies showed a significant correlation between immunohistochemistry and mass spectrometry-based proteomics, indicating that the association of protein and drug response identified by mass spectrometry-based proteomics can also be translated to immunohistochemistry. These results supported that the analytical validation in colorectal cancer cell lines of GUCY2C and SLC29A1 was accurate, precise, and repeatable.

5.7 Validation of potential biomarkers in colorectal cancer patient samples

Our study incorporated a large number of colorectal cancer cell lines to rigorously screen the sensitivity data of 57 cell lines to 5-FU and oxaliplatin and 90 lines to irinotecan, merged and analyzed omics data from the cell lines, including DNA, mRNA and protein. According to our findings, microsatellite instability (MSI), consensus molecular subtype 1 (MSI Immune) and *KRAS* status are able to predict irinotecan response in colorectal cancer cell lines, indicating that these molecular features could be valuable in selecting for irinotecan-based or

monotherapy treatment in colorectal cancer. Unfortunately, information on the MSI status, *KRAS* mutations or the consensus molecular subtype of the tumors in our cohort of 548 mCRC cases was not available, and these results await future validation in the clinical setting.

We evaluated whether *GUCY2C* and *SLC29A1* could be predictive biomarkers for irinotecan response using 548 patients' samples from three independent cohorts treated with irinotecan-based regimens. Our results showed no significant association between the levels of expression of these two proteins and the response rates to irinotecan. Of note, the findings were inconsistent across the individual cohorts. These inconsistent results in the three cohorts could be related to several differences in characteristic variables between different cohorts, such as gender, grade of the tumor, TNM stages, and site of the primary tumor. Thus, we evaluated these factors by conducting multivariate logistic regression. The results indicated *SLC29A1* and *GUCY2C* were not independent factors for irinotecan response. In addition, other anti-cancer drugs included in the regimens, such as 5-FU, oxaliplatin, and targeted therapy, may disturb the impact of biomarker levels on the irinotecan response rate.

Furthermore, we explored whether *SLC29A1* and *GUCY2C* expression levels were associated with overall survival and progression-free survival after irinotecan-based treatment in colorectal cancer patients. For *SLC29A1*, the high protein expression was statistically associated with longer overall survival, although there was no association between *SLC29A1* expression and progression-free survival. However, when we analyzed the three cohorts separately, one cohort showed that patients with high *SLC29A1* expression had longer overall survival. In contrast, one cohort showed opposite results, indicating the challenge of an extensive evaluation of potential biomarkers across different cohorts. Multivariate Cox regression indicated *SLC29A1* was not an independent factor of overall and progression-free survival. For *GUCY2C*, although no association between *GUCY2C* protein expression and progression-free survival was found, patients with high tumor protein expression showed significantly shorter overall survival in the three cohorts, both combined and when analyzed separately. Moreover, multivariate Cox regression results also showed that *GUCY2C* was an independent factor associated with overall survival but not progression-free survival.

As described before, *GUCY2C* is a transmembrane protein located on the luminal surface of the intestinal mucosa that functions as a receptor for the endogenous peptides guanylin and uroguanylin (Potter, 2011). Loss of its ligands (guanylin and uroguanylin) may contribute to

the inactivation of GUCY2C, causing colorectal cancer tumorigenesis. GUCY2C is activated after binding its ligands and converts guanosine-5-triphosphate (GTP) to cGMP and increases cGMP levels, causing the activation of downstream effectors in the cell, involving DNA damage repair and cellular metabolism (Lin et al., 2009). In addition, activated GUCY2C by guanylin and uroguanylin could increase DNA damage repair, which is also one of the most important mechanisms of drug resistance, especially for chemotherapeutic drugs inhibiting DNA replication and transcription, *e.g.*, irinotecan (Beidler et al., 1996; Xu & Villalona-Calero, 2002). Moreover, alteration of the cellular metabolism caused by activated GUCY2C could result in increased conversion of irinotecan to its inactive metabolite. Studies have shown higher GUCY2C mRNA levels in rectal cancer tumor tissues than in normal tissues of the rectum (Y. Liu et al., 2017) and the association between high mRNA levels of GUCY2C with reduced overall survival and progression-free survival in colorectal cancer patients (Shi et al., 2017). Altogether, our study provided evidence that high GUCY2C protein levels are associated with shorter patient overall survival in patients under irinotecan-based therapy.

An interesting predictive protein biomarker to irinotecan response was the ATP Binding Cassette Subfamily B Member 1 (ABCB1), coding for Multidrug Resistance Protein 1 (also known as P-Glycoprotein 1), which transports various molecules over the cell membrane, including chemotherapeutic drugs. Previous studies have shown that ABCB1 contributes to the elimination of irinotecan and its metabolites from cancer cells. Furthermore, polymorphisms of ABCB1 have been found to be associated with the response to irinotecan-based treatment (Glimelius et al., 2011). However, a study showed the protein level of ABCB1 did not correlate with response to irinotecan in metastatic colorectal cancer patients. Regarding other identified predictive biomarkers, their role in drug response is still poorly understood. Further studies are necessary to determine the predictive power of these candidate biomarkers.

There existed several limitations to the current study. The two potential protein-based biomarkers we selected identified by mass spectrometry-based proteomics were not reproducibly validated in our three independent cohorts of patient samples. Only an external cohort validation showed the expected results for SLC29A1 and GUCY2C at the mRNA level. The limited number of patients treated with irinotecan alone as the first-line therapy, incomplete clinical features of the patients, tumors and follow-up, and differences in baseline characteristics may explain the inconsistency of clinical validation results across cohorts. In

this study, we incorporated genomics, transcriptomics and proteomics data to investigate the predictive biomarkers. However, other omics such as miRNA, lncRNA and/or metabolomics also play a key role in the drug response in cancer cell lines. The integration of several omics data (multi-omics analysis) will also contribute positively to the discovery of biomarkers.

Validation in large-scale, multicenter cohorts is necessary to fully investigate the potential predictive role of these proteins. Besides, we plan to confirm more potential biomarkers identified by mass spectrometry-based proteomics in our cohorts. Rigorous basic experiments are also needed to investigate molecular mechanisms used by these proteins to build drug sensitivity/resistance into cells.

6 CONCLUSIONS

5-fluorouracil (5-FU), oxaliplatin and irinotecan are chemotherapeutic agents for metastatic colorectal cancer. Due to the low response rate to a single agent of colorectal cancer patients, significant side effects, and high medical costs, it is essential to find strategies to guide the use of the different therapeutic agents available and, therefore, improve patients' survival rates. Predictive biomarkers contribute to selecting patients with a high probability of responding to a given chemotherapy. However, there are currently no biomarkers capable of predicting the response to 5-FU, oxaliplatin and irinotecan. In this study, we thoroughly characterized the drug sensitivity and obtained high throughput expression and mutation data of a large panel of colorectal cancer cell lines and identified novel candidate predictive biomarkers to chemotherapeutic agents' response in colorectal cancer patients.

The main conclusions of this study are:

- The sensitivity of 57 colorectal cancer cell lines to 5-FU and oxaliplatin and 90 colorectal cancer cell lines to irinotecan was characterized using a novel method (ccIRECIST) based on the inhibition of cell growth after treatment.
- Microsatellite instability (MSI) is associated with a better response to irinotecan in colorectal cancer cell lines.
- Consensus Molecular subtype (CMS) 2 (Canonical) is associated with a worse response to irinotecan in colorectal cancer cell lines, compared to CMS1 (MSI Immune) and CMS4 (Mesenchymal).
- KRAS mutations are associated with resistance to irinotecan in colorectal cancer cell lines.
- The expression level of sixteen proteins is significantly associated with the response to irinotecan in colorectal cancer cell lines, and constitutes novel candidate biomarkers of irinotecan response in colorectal cancer patients.
- GUCY2C can predict the overall survival of colorectal cancer patients after irinotecan-based treatment.

These results add to the literature novel predictive biomarkers of irinotecan response; further validation of the biomarkers' predictive power is warranted.

7 BIBLIOGRAPHY

- Abdallah, E. A., Fanelli, M. F., Buim, M. E. C., Machado Netto, M. C., Gasparini Junior, J. L., Souza E Silva, V., Dettino, A. L. A., Mingues, N. B., Romero, J. V., Ocea, L. M. M., Rocha, B. M. M., Alves, V. S., Araújo, D. V., & Chinen, L. T. D. (2015). Thymidylate synthase expression in circulating tumor cells: a new tool to predict 5-fluorouracil resistance in metastatic colorectal cancer patients. *International Journal of Cancer*, *137*(6), 1397–1405. <https://doi.org/10.1002/ijc.29495>
- Akhurst, R. J., & Hata, A. (2012). Targeting the TGF β signalling pathway in disease. *Nature Reviews Drug Discovery*, *11*(10), 790–811. <https://doi.org/10.1038/nrd3810>
- Alazzouzi, H., Alhopuro, P., Salovaara, R., Sammalkorpi, H., Järvinen, H., Mecklin, J.-P., Hemminki, A., Schwartz Jr, S., Aaltonen, L. A., & Arango, D. (2005). SMAD4 as a prognostic marker in colorectal cancer. *Clinical Cancer Research*, *11*(7), 2606–2611.
- Alhopuro, P., Alazzouzi, H., Sammalkorpi, H., Dávalos, V., Salovaara, R., Hemminki, A., Järvinen, H., Mecklin, J.-P., Schwartz Jr, S., & Aaltonen, L. A. (2005). SMAD4 levels and response to 5-fluorouracil in colorectal cancer. *Clinical Cancer Research*, *11*(17), 6311–6316.
- Alley, M. C., Scudiero, D. A., Monks, A., Hursey, M. L., Czerwinski, M. J., Fine, D. L., Abbott, B. J., Mayo, J. G., Shoemaker, R. H., & Boyd, M. R. (1988). Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. *Cancer Research*, *48*(3), 589–601.
- Amado, R. G., Wolf, M., Peeters, M., van Cutsem, E., Siena, S., Freeman, D. J., Juan, T., Sikorski, R., Suggs, S., & Radinsky, R. (2008). *Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer.*
- Anand, S., Penrhyn-Lowe, S., & Venkitaraman, A. R. (2003). AURORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol. *Cancer Cell*, *3*(1), 51–62.

- Andre, F., Ismaila, N., Allison, K. H., Barlow, W. E., Collyar, D. E., Damodaran, S., Henry, N. L., Jhaveri, K., Kalinsky, K., Kuderer, N. M., Litvak, A., Mayer, E. L., Puztai, L., Raab, R., Wolff, A. C., & Stearns, V. (2022). Biomarkers for Adjuvant Endocrine and Chemotherapy in Early-Stage Breast Cancer: ASCO Guideline Update. *Journal of Clinical Oncology*, *40*(16), 1816–1837. <https://doi.org/10.1200/JCO.22.00069>
- André, T., Blons, H., Mabro, M., Chibaudel, B., Bachet, J.-B., Tournigand, C., Bennamoun, M., Artru, P., Nguyen, S., Ebenezer, C., Aissat, N., Cayre, A., Penault-Llorca, F., Laurent-Puig, P., & de Gramont, A. (2013). Panitumumab combined with irinotecan for patients with KRAS wild-type metastatic colorectal cancer refractory to standard chemotherapy: a GERCOR efficacy, tolerance, and translational molecular study. *Annals of Oncology*, *24*(2), 412–419. <https://doi.org/https://doi.org/10.1093/annonc/mds465>
- André, T., Boni, C., Mounedji-Boudiaf, L., Navarro, M., Tabernero, J., Hickish, T., Topham, C., Zaninelli, M., Clingan, P., & Bridgewater, J. (2004). Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *New England Journal of Medicine*, *350*(23), 2343–2351.
- Ang, P. W., Loh, M., Liem, N., Lim, P. L., Grieu, F., Vaithilingam, A., Platell, C., Yong, W. P., Iacopetta, B., & Soong, R. (2010). Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC Cancer*, *10*(1), 227. <https://doi.org/10.1186/1471-2407-10-227>
- Arango, D., Laiho, P., Kokko, A., Alhopuro, P., Sammalkorpi, H., Salovaara, R., Nicorici, D., Hautaniemi, S., Alazzouzi, H., Mecklin, J., Järvinen, H., Hemminki, A., Astola, J., Schwartz, S., & Aaltonen, L. A. (2005). Gene-Expression Profiling Predicts Recurrence in Dukes' C Colorectal Cancer. *Gastroenterology*, *129*(3), 874–884. <https://doi.org/https://doi.org/10.1053/j.gastro.2005.06.066>
- Arango, D., Wilson, A. J., Shi, Q., Corner, G. A., Arañes, M. J., Nicholas, C., Lesser, M., Mariadason, J. M., & Augenlicht, L. H. (2004). Molecular mechanisms of action and prediction of response to oxaliplatin in colorectal cancer cells. *British Journal of Cancer*, *91*(11), 1931–1946. <https://doi.org/10.1038/sj.bjc.6602215>

- Arnold, M., Abnet, C. C., Neale, R. E., Vignat, J., Giovannucci, E. L., McGlynn, K. A., & Bray, F. (2020). Global Burden of 5 Major Types of Gastrointestinal Cancer. *Gastroenterology*, 159(1), 335-349.e15. [https://doi.org/https://doi.org/10.1053/j.gastro.2020.02.068](https://doi.org/10.1053/j.gastro.2020.02.068)
- Arshad, N., & Visweswariah, S. S. (2012). The multiple and enigmatic roles of guanylyl cyclase C in intestinal homeostasis. *FEBS Letters*, 586(18), 2835–2840.
- Aslantürk, Ö. S. (2018). In vitro cytotoxicity and cell viability assays: principles, advantages, and disadvantages. *Genotoxicity-A Predictable Risk to Our Actual World*, 2, 64–80.
- B Rubinfield, I Albert, E Porfiri, C Fiol, S Munemitsu, & P Polakis. (1996). Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. *Science*.
- Baeg, G. H., Matsumine, A., Kuroda, T., Bhattacharjee, R. N., Miyashiro, I., Toyoshima, K., & Akiyama, T. (1995). The tumour suppressor gene product APC blocks cell cycle progression from G0/G1 to S phase. *The EMBO Journal*, 14(22), 5618–5625. <https://pubmed.ncbi.nlm.nih.gov/8521819>
- Baghban, R., Roshangar, L., Jahanban-Esfahlan, R., Seidi, K., Ebrahimi-Kalan, A., Jaymand, M., Kolahian, S., Javaheri, T., & Zare, P. (2020). Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Communication and Signaling*, 18(1), 1–19.
- Baldwin, S. A., Beal, P. R., Yao, S. Y. M., King, A. E., Cass, C. E., & Young, J. D. (2004). The equilibrative nucleoside transporter family, SLC29. *Pflügers Archiv*, 447(5), 735–743.
- Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., & Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1), 16878. <https://doi.org/10.1038/s41598-017-17204-5>
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. v., & Sonkin, D. (2012a). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391), 603–607.

- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. v., & Sonkin, D. (2012b). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, *483*(7391), 603–607.
- Bashir, B., Merlino, D. J., Rappaport, J. A., Gnass, E., Palazzo, J. P., Feng, Y., Fearon, E. R., Snook, A. E., & Waldman, S. A. (2019). Silencing the GUCA2A-GUCY2C tumor suppressor axis in CIN, serrated, and MSI colorectal neoplasia. *Human Pathology*, *87*, 103–114. <https://doi.org/https://doi.org/10.1016/j.humpath.2018.11.032>
- Basu, A., Bodycombe, N. E., Cheah, J. H., Price, E. v, Liu, K., Schaefer, G. I., Ebright, R. Y., Stewart, M. L., Ito, D., & Wang, S. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, *154*(5), 1151–1161.
- Battaglin, F., Naseem, M., Lenz, H.-J., & Salem, M. E. (2018). Microsatellite instability in colorectal cancer: overview of its clinical significance and novel perspectives. *Clinical Advances in Hematology & Oncology: H&O*, *16*(11), 735.
- Baxter, N. N., Kennedy, E. B., Bergsland, E., Berlin, J., George, T. J., Gill, S., Gold, P. J., Hantel, A., Jones, L., Lieu, C., Mahmoud, N., Morris, A. M., Ruiz-Garcia, E., You, Y. N., & Meyerhardt, J. A. (2021). Adjuvant Therapy for Stage II Colon Cancer: ASCO Guideline Update. *Journal of Clinical Oncology*, *40*(8), 892–910. <https://doi.org/10.1200/JCO.21.02538>
- Bazzocco, S., Dopeso, H., Carton-Garcia, F., Macaya, I., Andretta, E., Chionh, F., Rodrigues, P., Garrido, M., Alazzouzi, H., Nieto, R., Sanchez, A., Schwartz, S., Bilic, J., Mariadason, J. M., & Arango, D. (2015). Highly expressed genes in rapidly proliferating tumor cells as new targets for colorectal cancer treatment. *Clinical Cancer Research*, *21*(16), 3695–3704. <https://doi.org/10.1158/1078-0432.CCR-14-2457>
- Beidler, D. R., Chang, J.-Y., Zhou, B., & Cheng, Y. (1996). Camptothecin resistance involving steps subsequent to the formation of protein-linked DNA breaks in human camptothecin-resistant KB cell lines. *Cancer Research*, *56*(2), 345–353.

- Berger, S. H., Barbour, K. W., & Berger, F. G. (1988). A naturally occurring variation in thymidylate synthase structure is associated with a reduced response to 5-fluoro-2'-deoxyuridine in a human colon tumor cell line. *Molecular Pharmacology*, *34*(4), 480–484.
- Bernick, P. E., Klimstra, D. S., Shia, J., Minsky, B., Saltz, L., Shi, W., Thaler, H., Guillem, J., Paty, P., & Cohen, A. M. (2004). Neuroendocrine carcinomas of the colon and rectum. *Diseases of the Colon & Rectum*, *47*(2), 163–169.
- Bertagnolli, M. M., Niedzwiecki, D., Compton, C. C., Hahn, H. P., Hall, M., Damas, B., Jewell, S. D., Mayer, R. J., Goldberg, R. M., & Saltz, L. B. (2009). Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon cancer: Cancer and Leukemia Group B Protocol 89803. *Journal of Clinical Oncology*, *27*(11), 1814.
- Bevins, C. L., & Salzman, N. H. (2011). Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nature Reviews Microbiology*, *9*(5), 356–368. <https://doi.org/10.1038/nrmicro2546>
- Bhonde, M. R., Hanski, M., Stehr, J., Jebautzke, B., Peiró-Jordán, R., Fechner, H., Yokoyama, K. K., Lin, W., Zeitz, M., & Hanski, C. (2010). Mismatch repair system decreases cell survival by stabilizing the tetraploid G1 arrest in response to SN-38. *International Journal of Cancer*, *126*(12), 2813–2825.
- Bienz, M., & Hamada, F. (2004). Adenomatous polyposis coli proteins and cell adhesion. *Current Opinion in Cell Biology*, *16*(5), 528–535. <https://doi.org/https://doi.org/10.1016/j.ceb.2004.08.001>
- Bleiberg, H. (1996). Role of chemotherapy for advanced colorectal cancer: new opportunities. *Seminars in Oncology*, *23*(1 Suppl 3), 42–50.
- Blumenthal, G. M., Karuri, S. W., Zhang, H., Zhang, L., Khozin, S., Kazandjian, D., Tang, S., Sridhara, R., Keegan, P., & Pazdur, R. (2015). Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non-small-cell lung cancer: US Food and Drug Administration trial-level and patient-level analyses. *Journal of Clinical Oncology*, *33*(9), 1008.

- Böckelman, C., Engelman, B. E., Kaprio, T., Hansen, T. F., & Glimelius, B. (2015a). Risk of recurrence in patients with colon cancer stage II and III: a systematic review and meta-analysis of recent literature. *Acta Oncologica*, *54*(1), 5–16.
- Böckelman, C., Engelman, B. E., Kaprio, T., Hansen, T. F., & Glimelius, B. (2015b). Risk of recurrence in patients with colon cancer stage II and III: a systematic review and meta-analysis of recent literature. *Acta Oncologica*, *54*(1), 5–16.
- Bode, A. M., & Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. *Nature Reviews Cancer*, *4*(10), 793–805. <https://doi.org/10.1038/nrc1455>
- Bokemeyer, C., Bondarenko, I., Hartmann, J. T., de Braud, F., Schuch, G., Zubel, A., Celik, I., Schlichting, M., & Koralewski, P. (2011). Efficacy according to biomarker status of cetuximab plus FOLFOX-4 as first-line treatment for metastatic colorectal cancer: the OPUS study. *Annals of Oncology*, *22*(7), 1535–1546.
- Boron, W. F., & Boulpaep, E. L. (2016). *Medical Physiology: Elsevier Health Sciences*. 2016. ISBN 9781455733286. https://en.wikipedia.org/wiki/Urinary_bladder, 738.
- Bos, J. L., Rehmann, H., & Wittinghofer, A. (2007). GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell*, *129*(5), 865–877. <https://doi.org/10.1016/j.cell.2007.05.018>
- Braun, M. S., Richman, S. D., Quirke, P., Daly, C., Adlard, J. W., Elliott, F., Barrett, J. H., Selby, P., Meade, A. M., & Stephens, R. J. (2008). Predictive biomarkers of chemotherapy efficacy in colorectal cancer: results from the UK MRC FOCUS trial. *Journal of Clinical Oncology*, *26*(16), 2690–2698.
- Brunen, D., Willems, S., Kellner, U., Midgley, R., Simon, I., & Bernards, R. (2013). TGF- β : an emerging player in drug resistance. *Cell Cycle*, *12*(18), 2960–2968.
- Bruno Sarmiento. (2015). *Concepts and Models for Drug Permeability Studies: Cell and Tissue based In Vitro Culture Models*. Woodhead Publishing.

- Budczies, J., Klauschen, F., Sinn, B. v, Györffy, B., Schmitt, W. D., Darb-Esfahani, S., & Denkert, C. (2012). Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PloS One*, 7(12), e51862.
- Bukowski, K., Kciuk, M., & Kontek, R. (2020). Mechanisms of Multidrug Resistance in Cancer Chemotherapy. *International Journal of Molecular Sciences*, 21(9), 3233. <https://doi.org/10.3390/ijms21093233>
- Cantwell-Dorris, E. R., O’Leary, J. J., & Sheils, O. M. (2011). BRAFV600E: implications for carcinogenesis and molecular therapy. *Molecular Cancer Therapeutics*, 10(3), 385–394.
- Capitain, O., Boisdron-Celle, M., Poirier, A.-L., Abadie-Lacourtoisie, S., Morel, A., & Gamelin, E. (2008). The influence of fluorouracil outcome parameters on tolerance and efficacy in patients with advanced colorectal cancer. *The Pharmacogenomics Journal*, 8(4), 256–267. <https://doi.org/10.1038/sj.tpj.6500476>
- Caputo, F., Santini, C., Bardasi, C., Cerma, K., Casadei-Gardini, A., Spallanzani, A., Andrikou, K., Cascinu, S., & Gelsomino, F. (2019). BRAF-mutated colorectal cancer: clinical and molecular insights. *International Journal of Molecular Sciences*, 20(21), 5369.
- Cassidy, J. W., Caldas, C., & Bruna, A. (2015). Maintaining tumor heterogeneity in patient-derived tumor xenografts. *Cancer Research*, 75(15), 2963–2968.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., & Larsson, E. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, 2(5), 401–404.
- Cervantes, A., Adam, R., Roselló, S., Arnold, D., Normanno, N., Taïeb, J., Seligmann, J., de Baere, T., Osterlund, P., Yoshino, T., & Martinelli, E. (2023). Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up^{☆}; *Annals of Oncology*, 34(1), 10–32. <https://doi.org/10.1016/j.annonc.2022.10.003>
- Chang, K., Creighton, C. J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y. S. N., Chu, A., Chuah, E., Chun, H.-J. E., Dhalla, N., Guin, R., Hirst, M., Hirst, C., Holt, R. A., Jones, S. J. M., ... Center, B. C. R.

- (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*, 45(10), 1113–1120. <https://doi.org/10.1038/ng.2764>
- Chen, Y., Ren, B., Yang, J., Wang, H., Yang, G., Xu, R., You, L., & Zhao, Y. (2020). The role of histone methylation in the development of digestive cancers: a potential direction for cancer management. *Signal Transduction and Targeted Therapy*, 5(1), 1–13.
- Chua, W., Goldstein, D., Lee, C. K., Dhillon, H., Michael, M., Mitchell, P., Clarke, S. J., & Iacopetta, B. (2009). Molecular markers of response and toxicity to FOLFOX chemotherapy in metastatic colorectal cancer. *British Journal of Cancer*, 101(6), 998–1004.
- Ciccolallo, L., Capocaccia, R., Coleman, M. P., Berrino, F., Coebergh, J. W. W., Damhuis, R. A. M., Faivre, J., Martinez-Garcia, C., Møller, H., & de Leon, M. P. (2005). Survival differences between European and US patients with colorectal cancer: role of stage at diagnosis and surgery. *Gut*, 54(2), 268–273.
- Clevers, H. (2013). The intestinal crypt, a prototype stem cell compartment. *Cell*, 154(2), 274–284.
- Compton, C. C. (2003). Colorectal carcinoma: diagnostic, prognostic, and molecular features. *Modern Pathology*, 16(4), 376–388.
- Crooke, H., Kobayashi, M., Mitchell, B., Nwokeji, E., Laurie, M., Kamble, S., McKenna, M., Masood, A., & Korytowsky, B. (2018). Estimating 1- and 5-year relative survival trends in colorectal cancer (CRC) in the United States: 2004 to 2014. *Journal of Clinical Oncology*, 36(4_suppl), 587. https://doi.org/10.1200/JCO.2018.36.4_suppl.587
- Cunningham, D., Humblet, Y., Siena, S., Khayat, D., Bleiberg, H., Santoro, A., Bets, D., Mueser, M., Harstrick, A., & Verslype, C. (2004). Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *New England Journal of Medicine*, 351(4), 337–345.
- Cunningham, D., Sirohi, B., Pluzanska, A., Utracka-Hutka, B., Zaluski, J., Glynne-Jones, R., Koralewski, P., Bridgewater, J., Mainwaring, P., & Wasan, H. (2009). Two different first-

line 5-fluorouracil regimens with or without oxaliplatin in patients with metastatic colorectal cancer. *Annals of Oncology*, 20(2), 244–250.

Curtin, K., Slattery, M. L., & Samowitz, W. S. (2011). CpG Island Methylation in Colorectal Cancer: Past, Present and Future. *Pathology Research International*, 2011, 902674. <https://doi.org/10.4061/2011/902674>

De, A. (2011). Wnt/Ca²⁺ signaling pathway: a brief overview. *Acta Biochim Biophys Sin*, 43(10), 745–756.

de Gramont, A., Vignoud, J., Tournigand, C., Louvet, C., Andre, T., Varette, C., Raymond, E., Moreau, S., le Bail, N., & Krulik, M. (1997). Oxaliplatin with high-dose leucovorin and 5-fluorouracil 48-hour continuous infusion in pretreated metastatic colorectal cancer. *European Journal of Cancer*, 33(2), 214–219.

de Man, F. M., Goey, A. K. L., van Schaik, R. H. N., Mathijssen, R. H. J., & Bins, S. (2018). Individualization of Irinotecan Treatment: A Review of Pharmacokinetics, Pharmacodynamics, and Pharmacogenetics. *Clinical Pharmacokinetics*, 57(10), 1229–1254. <https://doi.org/10.1007/s40262-018-0644-7>

de Roock, W., de Vriendt, V., Normanno, N., Ciardiello, F., & Tejpar, S. (2011). KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *The Lancet Oncology*, 12(6), 594–603.

Diaz, L. A., Le, D. T., Kim, T. W., van Cutsem, E., Geva, R., Jaeger, D., Hara, H., Burge, M. E., O’Neil, B. H., & Kavan, P. (2020). *Pembrolizumab monotherapy for patients with advanced MSI-H colorectal cancer: Longer-term follow-up of the phase II, KEYNOTE-164 study*. American Society of Clinical Oncology.

Douillard, Jy., Cunningham, D., Roth, A. D., Navarro, M., James, R. D., Karasek, P., Jandik, P., Iveson, T., Carmichael, J., & Alakl, M. (2000). Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *The Lancet*, 355(9209), 1041–1047.

- Drexler, H. G. (1998). Review of alterations of the cyclin-dependent kinase inhibitor INK4 family genes p15, p16, p18 and p19 in human leukemia–lymphoma cells. *Leukemia*, *12*(6), 845–859.
- Duffy, M. J., O’Grady, S., Tang, M., & Crown, J. (2021). MYC as a target for cancer treatment. *Cancer Treatment Reviews*, *94*, 102154.
- Edge, S., Byrd, D. R., Compton, C. C., Fritz, A. G., Greene, F., & Trotti, A. (2010). *AJCC Cancer Staging Handbook*.
- Eggenberger, J. C. (2011). Carcinoid and other neuroendocrine tumors of the colon and rectum. *Clinics in Colon and Rectal Surgery*, *24*(03), 129–134.
- Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., & Mooney, M. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer*, *45*(2), 228–247.
- Engelman, J. A. (2009). Targeting PI3K signalling in cancer: opportunities, challenges and limitations. *Nature Reviews Cancer*, *9*(8), 550–562.
- Ergun, Y., Acikgoz, Y., Bal, O., Ucar, G., Dirikoc, M., Caliskan Yildirim, E., Akdeniz, N., & Uncu, D. (2019). KRAS codon 12 and 13 mutations may guide the selection of irinotecan or oxaliplatin in first-line treatment of metastatic colorectal cancer. *Expert Review of Molecular Diagnostics*, *19*(12), 1131–1140. <https://doi.org/10.1080/14737159.2019.1693266>
- Eso, Y., Shimizu, T., Takeda, H., Takai, A., & Marusawa, H. (2020). Microsatellite instability and immune checkpoint inhibitors: toward precision medicine against gastrointestinal and hepatobiliary cancers. *Journal of Gastroenterology*, *55*(1), 15–26. <https://doi.org/10.1007/s00535-019-01620-7>
- Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics*, *8*(4), 286–298. <https://doi.org/10.1038/nrg2005>

- Fallahi-Sichani, M., Honarnejad, S., Heiser, L. M., Gray, J. W., & Sorger, P. K. (2013). Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nature Chemical Biology*, 9(11), 708–714. <https://doi.org/10.1038/nchembio.1337>
- Fallik, D., Borrini, F., Boige, V., Viguier, J., Jacob, S., Miquel, C., Sabourin, J.-C., Ducreux, M., & Praz, F. (2003). Microsatellite instability is a predictive factor of the tumor response to irinotecan in patients with advanced colorectal cancer. *Cancer Research*, 63(18), 5738–5744.
- Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, 61(5), 759–767.
- Fekete, J. T., & Györfy, B. (2019). ROCplot. org: Validating predictive biomarkers of chemotherapy/hormonal therapy/anti-HER2 therapy using transcriptomic data of 3,104 breast cancer patients. *International Journal of Cancer*, 145(11), 3140–3151.
- Fujii, M., & Sato, T. (2021). Somatic cell-derived organoids as prototypes of human epithelial tissues and diseases. *Nature Materials*, 20(2), 156–169.
- Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., & Diaz, L. A. (2019). Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nature Reviews Gastroenterology & Hepatology*, 16(6), 361–375. <https://doi.org/10.1038/s41575-019-0126-x>
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., & Larsson, E. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, 6(269), p11–p11.
- Garcia-Carbonero, N., Martinez-Useros, J., Li, W., Orta, A., Perez, N., Carames, C., Hernandez, T., Moreno, I., Serrano, G., & Garcia-Foncillas, J. (2020). KRAS and BRAF mutations as prognostic and predictive biomarkers for standard chemotherapy response in metastatic colorectal cancer: a single institutional study. *Cells*, 9(1), 219.
- Gertler, R., Rosenberg, R., Stricker, D., Friederichs, J., Hoos, A., Werner, M., Ulm, K., Holzmann, B., Nekarda, H., & Siewert, J.-R. (2004). Telomere length and human

- telomerase reverse transcriptase expression as markers for progression and prognosis of colorectal carcinoma. *Journal of Clinical Oncology*, 22(10), 1807–1814.
- Giacchetti, S., Perpoint, B., Zidani, R., le Bail, N., Faggiuolo, R., Focan, C., Chollet, P., Llory, J. F., Letourneau, Y., & Coudert, B. (2000). Phase III multicenter randomized trial of oxaliplatin added to chronomodulated fluorouracil–leucovorin as first-line treatment of metastatic colorectal cancer. *Journal of Clinical Oncology*, 18(1), 136.
- Giannini, G., Ristori, E., Cerignoli, F., Rinaldi, C., Zani, M., Viel, A., Ottini, L., Crescenzi, M., Martinotti, S., Bignami, M., Frati, L., Screpanti, I., & Gulino, A. (2002). Human MRE11 is inactivated in mismatch repair-deficient cancers. *EMBO Reports*, 3(3), 248–254. <https://doi.org/10.1093/embo-reports/kvf044>
- Gibbons, A. v, Lin, J. E., Kim, G. W., Marszalowicz, G. P., Li, P., Stoecker, B. A., Blomain, E. S., Rattan, S., Snook, A. E., Schulz, S., & Waldman, S. A. (2013). Intestinal GUCY2C Prevents TGF- β Secretion Coordinating Desmoplasia and Hyperproliferation in Colorectal Cancer. *Cancer Research*, 73(22), 6654–6666. <https://doi.org/10.1158/0008-5472.CAN-13-0887>
- Giovanella, B. C., Stehlin, J. S., Wall, M. E., Wani, M. C., Nicholas, A. W., Liu, L. F., Silber, R., & Potmesil, M. (1989). DNA topoisomerase I--targeted chemotherapy of human colon cancer in xenografts. *Science*, 246(4933), 1046–1048.
- Glimelius, B., Garmo, H., Berglund, Å., Fredriksson, L. A., Berglund, M., Kohnke, H., Byström, P., Sørbye, H., & Wadelius, M. (2011). Prediction of irinotecan and 5-fluorouracil toxicity and response in patients with advanced colorectal cancer. *The Pharmacogenomics Journal*, 11(1), 61–71. <https://doi.org/10.1038/tpj.2010.10>
- Gongora, C., Vezzio-Vie, N., Tuduri, S., Denis, V., Causse, A., Auzanneau, C., Collod-Beroud, G., Coquelle, A., Pasero, P., & Pourquier, P. (2011). New Topoisomerase I mutations are associated with resistance to camptothecin. *Molecular Cancer*, 10(1), 1–13.
- Goodspeed, A., Heiser, L. M., Gray, J. W., & Costello, J. C. (2016). Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics Cancer Cell Lines as Pharmacogenomic Models. *Molecular Cancer Research*, 14(1), 3–13.

- Gordon, M. D., & Nusse, R. (2006). Wnt signaling: multiple pathways, multiple receptors, and multiple transcription factors. *Journal of Biological Chemistry*, 281(32), 22429–22433.
- Gotovdorj, T., Lee, E., Lim, Y., Cha, E. J., Kwon, D., Hong, E., Kim, Y., & Oh, M.-Y. (2014). 2, 3, 7, 8-Tetrachlorodibenzo-p-dioxin induced cell-specific drug transporters with acquired cisplatin resistance in cisplatin sensitive cancer cells. *Journal of Korean Medical Science*, 29(9), 1188–1198.
- Grady, W. M., & Carethers, J. M. (2008). Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology*, 135(4), 1079–1099.
- Graf, U., Casanova, E. A., Wyck, S., Dalcher, D., Gatti, M., Vollenweider, E., Okoniewski, M. J., Weber, F. A., Patel, S. S., Schmid, M. W., Li, J., Sharif, J., Wanner, G. A., Koseki, H., Wong, J., Pelczar, P., Penengo, L., Santoro, R., & Cinelli, P. (2017). Prame17 mediates ground-state pluripotency through proteasomal–epigenetic combined pathways. *Nature Cell Biology*, 19(7), 763–773. <https://doi.org/10.1038/ncb3554>
- Graham, B. S., & Lynch, D. T. (2021). Burkitt Lymphoma. In *StatPearls [Internet]*. StatPearls Publishing.
- Greer, E. L., & Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, 13(5), 343–357.
- Group, B. D. W., Atkinson Jr, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., & Schooley, R. T. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69(3), 89–95.
- Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., de Sousa E Melo, F., Missiaglia, E., Ramay, H., Barras, D., ... Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, 21(11), 1350–1356. <https://doi.org/10.1038/nm.3967>
- Häfner, M. F., & Debus, J. (2016). Radiotherapy for colorectal cancer: current standards and future perspectives. *Visceral Medicine*, 32(3), 172–177.

- Haggar, F. A., & Boushey, R. P. (2009). Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clinics in Colon and Rectal Surgery*, 22(4), 191–197. <https://doi.org/10.1055/s-0029-1242458>
- Haibe-Kains, B., El-Hachem, N., Birkbak, N. J., Jin, A. C., Beck, A. H., Aerts, H. J. W. L., & Quackenbush, J. (2013). Inconsistency in large pharmacogenomic studies. *Nature*, 504(7480), 389–393.
- Haites, N. (2001). *Oncogenes* (S. Brenner & J. H. B. T.-E. of G. Miller, Eds.; pp. 1370–1372). Academic Press. <https://doi.org/https://doi.org/10.1006/rwgn.2001.0925>
- Halkidou, K., Gaughan, L., Cook, S., Leung, H. Y., Neal, D. E., & Robson, C. N. (2004). Upregulation and nuclear recruitment of HDAC1 in hormone refractory prostate cancer. *The Prostate*, 59(2), 177–189. <https://doi.org/10.1002/pros.20022>
- Haluska, F. G., Tsujimoto, Y., & Croce, C. M. (1987). The t(8;14) chromosome translocation of the Burkitt lymphoma cell line Daudi occurred during immunoglobulin gene rearrangement and involved the heavy chain diversity region. *Proceedings of the National Academy of Sciences of the United States of America*, 84(19), 6835–6839. <https://doi.org/10.1073/pnas.84.19.6835>
- Hammond, W. A., Swaika, A., & Mody, K. (2016). Pharmacologic resistance in colorectal cancer: A review. In *Therapeutic Advances in Medical Oncology* (Vol. 8, Issue 1, pp. 57–84). <https://doi.org/10.1177/1758834015614530>
- Hampel, H., Frankel, W. L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Nakagawa, H., Sotamaa, K., Prior, T. W., & Westman, J. (2005). Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *New England Journal of Medicine*, 352(18), 1851–1860.
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. In *Cell* (Vol. 144, Issue 5, pp. 646–674). Elsevier. <https://doi.org/10.1016/j.cell.2011.02.013>
- Harris, S. L., & Levine, A. J. (2005). The p53 pathway: positive and negative feedback loops. *Oncogene*, 24(17), 2899–2908. <https://doi.org/10.1038/sj.onc.1208615>

- Hatzis, C., Bedard, P. L., Birkbak, N. J., Beck, A. H., Aerts, H. J. W. L., Stern, D. F., Shi, L., Clarke, R., Quackenbush, J., & Haibe-Kains, B. (2014). Enhancing Reproducibility in Cancer Drug Screening: How Do We Move Forward? Path to Improve Concordance in Cancer Drug Screening. *Cancer Research*, *74*(15), 4016–4023.
- Haupt, S., Berger, M., Goldberg, Z., & Haupt, Y. (2003). Apoptosis-the p53 network. *Journal of Cell Science*, *116*(20), 4077–4085.
- Hector, S., Bolanowska-Higdon, W., Zdanowicz, J., Hitt, S., & Pendyala, L. (2001). In vitro studies on the mechanisms of oxaliplatin resistance. *Cancer Chemotherapy and Pharmacology*, *48*(5), 398–406.
- Hidalgo, M., Amant, F., Biankin, A. v, Budinská, E., Byrne, A. T., Caldas, C., Clarke, R. B., de Jong, S., Jonkers, J., & Mælandsmo, G. M. (2014). Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discovery*, *4*(9), 998–1013.
- Holohan, C., van Schaeybroeck, S., Longley, D. B., & Johnston, P. G. (2013). Cancer drug resistance: an evolving paradigm. *Nature Reviews Cancer*, *13*(10), 714–726. <https://doi.org/10.1038/nrc3599>
- Horisberger, K., Erben, P., Muessle, B., Woernle, C., Stroebel, P., Kaehler, G., Wenz, F., Hochhaus, A., Post, S., & Willeke, F. (2009). Topoisomerase I expression correlates to response to neoadjuvant irinotecan-based chemoradiation in rectal cancer. *Anti-Cancer Drugs*, *20*(6), 519–524.
- Horobin, R., & Kiernan, J. (2020). *Conn's biological stains: a handbook of dyes, stains and fluorochromes for use in biology and medicine*. Taylor & Francis.
- Huang, D., Sun, W., Zhou, Y., Li, P., Chen, F., Chen, H., Xia, D., Xu, E., Lai, M., Wu, Y., & Zhang, H. (2018). Mutations of key driver genes in colorectal cancer progression and metastasis. *Cancer and Metastasis Reviews*, *37*(1), 173–187. <https://doi.org/10.1007/s10555-017-9726-5>
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. <https://doi.org/10.1038/nprot.2008.211>

- Huang, Y., Anderle, P., Bussey, K. J., Barbacioru, C., Shankavaram, U., Dai, Z., Reinhold, W. C., Papp, A., Weinstein, J. N., & Sadée, W. (2004). Membrane transporters and channels: role of the transportome in cancer chemosensitivity and chemoresistance. *Cancer Research*, *64*(12), 4294–4301.
- Huang, Y., Cate, S. P., Battistuzzi, C., Oquendo, M. A., Brent, D., & Mann, J. J. (2004). An association between a functional polymorphism in the monoamine oxidase a gene promoter, impulsive traits and early abuse experiences. *Neuropsychopharmacology*, *29*(8), 1498–1505.
- Huang, Y., Liu, N., Liu, J., Liu, Y., Zhang, C., Long, S., Luo, G., Zhang, L., & Zhang, Y. (2019). Mutant p53 drives cancer chemotherapy resistance due to loss of function on activating transcription of PUMA. *Cell Cycle*, *18*(24), 3442–3455.
- Huch, M., Knoblich, J. A., Lutolf, M. P., & Martinez-Arias, A. (2017). The hope and the hype of organoid research. *Development*, *144*(6), 938–941.
- Huelsken, J., & Behrens, J. (2002). The Wnt signalling pathway. *Journal of Cell Science*, *115*(21), 3977–3978.
- Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., Cokelaer, T., Greninger, P., van Dyk, E., Chang, H., de Silva, H., Heyn, H., Deng, X., Egan, R. K., Liu, Q., ... Garnett, M. J. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, *166*(3), 740–754. <https://doi.org/https://doi.org/10.1016/j.cell.2016.06.017>
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, *4*(2), 249–264. <https://doi.org/10.1093/biostatistics/4.2.249>
- Ishikawa, T., & Ali-Osman, F. (1993). Glutathione-associated cis-diamminedichloroplatinum (II) metabolism and ATP-dependent efflux from leukemia cells. Molecular characterization of glutathione-platinum complex and its biological significance. *Journal of Biological Chemistry*, *268*(27), 20116–20125.

- Issa, J.-P. (2004). CpG island methylator phenotype in cancer. *Nature Reviews Cancer*, 4(12), 988–993.
- J, T. C., Nickolas, P., W, K. K., Bert, V., & Rachel, K. (2016). Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Sciences*, 113(50), 14330–14335. <https://doi.org/10.1073/pnas.1616440113>
- Jasperson, K. W., Tuohy, T. M., Neklason, D. W., & Burt, R. W. (2010). Hereditary and familial colon cancer. *Gastroenterology*, 138(6), 2044–2058.
- Jennifer, C.-F. (2013). Cancer Immunotherapy. *Science*, 342(6165), 1432–1433. <https://doi.org/10.1126/science.342.6165.1432>
- Jiao, S., Peters, U., Berndt, S., Brenner, H., Butterbach, K., Caan, B. J., Carlson, C. S., Chan, A. T., Chang-Claude, J., Chanock, S., Curtis, K. R., Duggan, D., Gong, J., Harrison, T. A., Hayes, R. B., Henderson, B. E., Hoffmeister, M., Kolonel, L. N., le Marchand, L., ... Hsu, L. (2014). Estimating the heritability of colorectal cancer. *Human Molecular Genetics*, 23(14), 3898–3905. <https://doi.org/10.1093/hmg/ddu087>
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241–254.
- Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1), 118–127.
- Johnston, P. G., & Kaye, S. (2001). Capecitabine: a novel agent for the treatment of solid tumors. *Anti-Cancer Drugs*, 12(8), 639–646.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
- Jover, R., Nguyen, T., Pérez–Carbonell, L., Zapater, P., Payá, A., Alenda, C., Rojas, E., Cubiella, J., Balaguer, F., & Morillas, J. D. (2011). 5-Fluorouracil adjuvant chemotherapy does not increase survival in patients with CpG island methylator phenotype colorectal cancer. *Gastroenterology*, 140(4), 1174–1181.
- Joyce, C., Rayi, A., & Kasi, A. (2021). *Tumor-Suppressor Genes*.

- Kang, H., O'Connell, J. B., Leonardi, M. J., Maggard, M. A., McGory, M. L., & Ko, C. Y. (2007). Rare tumors of the colon and rectum: a national review. *International Journal of Colorectal Disease*, 22(2), 183–189. <https://doi.org/10.1007/s00384-006-0145-2>
- Karakas, B., Bachman, K. E., & Park, B. H. (2006). Mutation of the PIK3CA oncogene in human cancers. *British Journal of Cancer*, 94(4), 455–459.
- Kelland, L. R. (1993). New platinum antitumor complexes. *Critical Reviews in Oncology/Hematology*, 15(3), 191–219.
- Keum, N. N., & Giovannucci, E. (2019). Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. In *Nature Reviews Gastroenterology and Hepatology* (Vol. 16, Issue 12). <https://doi.org/10.1038/s41575-019-0189-8>
- Kim, J. E., Hong, Y. S., Ryu, M., Lee, J., Chang, H. M., Lim, S., Kim, J. H., Jang, S., Kim, M., & Yu, C. S. (2011). Association between deficient mismatch repair system and efficacy to irinotecan-containing chemotherapy in metastatic colon cancer. *Cancer Science*, 102(9), 1706–1711.
- Kim, Y. S., & Deng, G. (2007). Epigenetic changes (aberrant DNA methylation) in colorectal neoplasia. *Gut and Liver*, 1(1), 1.
- Klingbiel, D., Saridaki, Z., Roth, A. D., Bosman, F. T., Delorenzi, M., & Tejpar, S. (2015). Prognosis of stage II and III colon cancer treated with adjuvant 5-fluorouracil or FOLFIRI in relation to microsatellite status: Results of the PETACC-3 trial. *Annals of Oncology*, 26(1), 126–132. <https://doi.org/10.1093/annonc/mdu499>
- Knudson Jr, A. G. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences*, 68(4), 820–823.
- Koncina, E., Haan, S., Rauh, S., & Letellier, E. (2020a). Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges. *Cancers*, 12(2), 319.
- Koncina, E., Haan, S., Rauh, S., & Letellier, E. (2020b). Prognostic and Predictive Molecular Biomarkers for Colorectal Cancer: Updates and Challenges. *Cancers*, 12(2), 319. <https://doi.org/10.3390/cancers12020319>

- Koopman, M., Knijn, N., Richman, S., Seymour, M., Quirke, P., van Tinteren, H., van Krieken, J., Punt, C. J. A., & Nagtegaal, I. D. (2009). 6003 The correlation between Topoisomerase-I (Topo1) expression and outcome of treatment with capecitabine and irinotecan in advanced colorectal cancer (ACC) patients (pts) treated in the CAIRO study of the Dutch Colorectal Cancer Group (DCCG). *EJC Supplements*, 2(7), 321–322.
- Kuipers, E. J., Grady, W. M., Lieberman, D., Seufferlein, T., Sung, J. J., Boelens, P. G., van de Velde, C. J. H., & Watanabe, T. (2015). Colorectal cancer. *Nature Reviews Disease Primers*, 1(1), 15065. <https://doi.org/10.1038/nrdp.2015.65>
- Kunicka, T., Prochazka, P., Krus, I., Bendova, P., Protivova, M., Susova, S., Hlavac, V., Liska, V., Novak, P., Schneiderova, M., Pitule, P., Bruha, J., Vycital, O., Vodicka, P., & Soucek, P. (2016). Molecular profile of 5-fluorouracil pathway genes in colorectal carcinoma. *BMC Cancer*, 16(1), 795. <https://doi.org/10.1186/s12885-016-2826-8>
- Lander, E. S. (1999). Array of hope. *Nature Genetics*, 21(1), 3–4.
- Lao, V. V., & Grady, W. M. (2011). Epigenetics and colorectal cancer. *Nature Reviews. Gastroenterology & Hepatology*, 8(12), 686–700. <https://doi.org/10.1038/nrgastro.2011.173>
- Lecomte, T., Ferraz, J.-M., Zinzindohoué, F., Lorient, M.-A., Tregouet, D.-A., Landi, B., Berger, A., Cugnenc, P.-H., Jian, R., & Beaune, P. (2004). Thymidylate synthase gene polymorphism predicts toxicity in colorectal cancer patients receiving 5-fluorouracil-based chemotherapy. *Clinical Cancer Research*, 10(17), 5880–5888.
- Lee, A. H., Shannon, C. P., Amenyogbe, N., Bennike, T. B., Diray-Arce, J., Idoko, O. T., Gill, E. E., Ben-Othman, R., Pomat, W. S., & van Haren, S. D. (2019a). Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. *Nature Communications*, 10(1), 1–14.
- Lee, A. H., Shannon, C. P., Amenyogbe, N., Bennike, T. B., Diray-Arce, J., Idoko, O. T., Gill, E. E., Ben-Othman, R., Pomat, W. S., & van Haren, S. D. (2019b). Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. *Nature Communications*, 10(1), 1–14.

- Lee, J. W., Devanarayan, V., Barrett, Y. C., Weiner, R., Allinson, J., Fountain, S., Keller, S., Weinryb, I., Green, M., Duan, L., Rogers, J. A., Millham, R., O'Brien, P. J., Sailstad, J., Khan, M., Ray, C., & Wagner, J. A. (2006). Fit-for-Purpose Method Development and Validation for Successful Biomarker Measurement. *Pharmaceutical Research*, 23(2), 312–328. <https://doi.org/10.1007/s11095-005-9045-3>
- Li, L., Schaid, D. J., Fridley, B. L., Kalari, K. R., Jenkins, G. D., Abo, R. P., Batzler, A., Moon, I., Pelleymounter, L., & Eckloff, B. W. (2012). Gemcitabine metabolic pathway genetic polymorphisms and response in non-small cell lung cancer patients. *Pharmacogenetics and Genomics*, 22(2), 105.
- Li, P., Fang, Y. J., Li, F., Ou, Q. J., Chen, G., & Ma, G. (2013). ERCC1, defective mismatch repair status as predictive biomarkers of survival for stage III colon cancer patients receiving oxaliplatin-based adjuvant chemotherapy. *British Journal of Cancer*, 108(6), 1238–1244. <https://doi.org/10.1038/bjc.2013.83>
- Li, P., Schulz, S., Bombonati, A., Palazzo, J. P., Hyslop, T. M., Xu, Y., Baran, A. A., Siracusa, L. D., Pitari, G. M., & Waldman, S. A. (2007). Guanylyl Cyclase C Suppresses Intestinal Tumorigenesis by Restricting Proliferation and Maintaining Genomic Integrity. *Gastroenterology*, 133(2), 599–607. <https://doi.org/10.1053/j.gastro.2007.05.052>
- LI, X., Haluska Jr, P., HSIANG, Y., Bharti, A., Kufe, D. W., & Rubin, E. H. (1996). Identification of Topoisomerase I Mutations Affecting Both DNA Cleavage and Interaction with Camptothecin a. *Annals of the New York Academy of Sciences*, 803(1), 111–127.
- Lieu, C., Kennedy, E. B., Bergsland, E., Berlin, J., George, T. J., Gill, S., Gold, P. J., Hantel, A., Jones, L., Mahmoud, N., Meyerhardt, J., Morris, A. M., Ruíz-García, E., You, Y. N., & Baxter, N. (2019). Duration of Oxaliplatin-Containing Adjuvant Therapy for Stage III Colon Cancer: ASCO Clinical Practice Guideline. *Journal of Clinical Oncology*, 37(16), 1436–1447. <https://doi.org/10.1200/JCO.19.00281>
- Lin, J. E., Li, P., Pitari, G. M., Schulz, S., & Waldman, S. A. (2009). *Guanylyl cyclase C in colorectal cancer: susceptibility gene and potential therapeutic target*.

- Liu, R., Chen, Y., Liu, G., Li, C., Song, Y., Cao, Z., Li, W., Hu, J., Lu, C., & Liu, Y. (2020). PI3K/AKT pathway as a key link modulates the multidrug resistance of cancers. *Cell Death & Disease*, *11*(9), 797. <https://doi.org/10.1038/s41419-020-02998-6>
- Liu, Y., Cheng, G., Qian, J., Ju, H., Zhu, Y., Stefano, M., Keilholz, U., & Li, D. (2017). Expression of guanylyl cyclase C in tissue samples and the circulation of rectal cancer patients. *Oncotarget*, *8*(24), 38841.
- Longley, D. B., Harkin, D. P., & Johnston, P. G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer*, *3*(5), 330–338.
- Loukola, A., Eklin, K., Laiho, P., Salovaara, R., Kristo, P., Järvinen, H., Mecklin, J.-P., Launonen, V., & Aaltonen, L. A. (2001). Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). *Cancer Research*, *61*(11), 4545–4549.
- Loupakis, F., Ruzzo, A., Cremolini, C., Vincenzi, B., Salvatore, L., Santini, D., Masi, G., Stasi, I., Canestrari, E., Rulli, E., Floriani, I., Bencardino, K., Galluccio, N., Catalano, V., Tonini, G., Magnani, M., Fontanini, G., Basolo, F., Falcone, A., & Graziano, F. (2009). KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. *British Journal of Cancer*, *101*(4), 715–721. <https://doi.org/10.1038/sj.bjc.6605177>
- Luqmani, Y. A. (2005). Mechanisms of Drug Resistance in Cancer Chemotherapy. *Medical Principles and Practice*, *14*(suppl 1)(Suppl. 1), 35–48. <https://doi.org/10.1159/000086183>
- Machover, D. (1997). A comprehensive review of 5-fluorouracil and leucovorin in patients with metastatic colorectal carcinoma. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *80*(7), 1179–1187.
- Magee, M. S., Abraham, T. S., Baybutt, T. R., Flickinger, J. C., Ridge, N. A., Marszalowicz, G. P., Prajapati, P., Hersperger, A. R., Waldman, S. A., & Snook, A. E. (2018). Human GUCY2C-Targeted Chimeric Antigen Receptor (CAR)-Expressing T Cells Eliminate Colorectal Cancer Metastases CAR-T Cells Target Human GUCY2C+ Colorectal Tumors. *Cancer Immunology Research*, *6*(5), 509–516.

- Magee, M. S., Kraft, C. L., Abraham, T. S., Baybutt, T. R., Marszalowicz, G. P., Li, P., Waldman, S. A., & Snook, A. E. (2016). GUCY2C-directed CAR-T cells oppose colorectal cancer metastases without autoimmunity. *Oncoimmunology*, *5*(10), e1227897.
- Mahli, A., Saugspier, M., Koch, A., Sommer, J., Dietrich, P., Lee, S., Thasler, R., Schulze-Luehrmann, J., Luehrmann, A., & Thasler, W. E. (2018). ERK activation and autophagy impairment are central mediators of irinotecan-induced steatohepatitis. *Gut*, *67*(4), 746–756.
- Mantovani, A., Marchesi, F., Malesci, A., Laghi, L., & Allavena, P. (2017). Tumour-associated macrophages as treatment targets in oncology. *Nature Reviews Clinical Oncology*, *14*(7), 399–416.
- Mantovani, F., Collavin, L., & del Sal, G. (2019). Mutant p53 as a guardian of the cancer cell. *Cell Death & Differentiation*, *26*(2), 199–212.
- Marcus, L., Lemery, S. J., Keegan, P., & Pazdur, R. (2019). FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid TumorsFDA Approval Summary: Pembrolizumab for MSI-H Solid Tumors. *Clinical Cancer Research*, *25*(13), 3753–3758.
- Martinelli, E., Morgillo, F., Troiani, T., & Ciardiello, F. (2017). Cancer resistance to therapies against the EGFR-RAS-RAF pathway: the role of MEK. *Cancer Treatment Reviews*, *53*, 61–69.
- Martinez-Balibrea, E., Martínez-Cardús, A., Ginés, A., Ruiz de Porras, V., Moutinho, C., Layos, L., Manzano, J. L., Bugés, C., Bystrup, S., Esteller, M., & Abad, A. (2015). Tumor-Related Molecular Mechanisms of Oxaliplatin Resistance. *Molecular Cancer Therapeutics*, *14*(8), 1767–1776. <https://doi.org/10.1158/1535-7163.MCT-14-0636>
- Massagué, J. (2012). TGF β signalling in context. *Nature Reviews. Molecular Cell Biology*, *13*(10), 616–630. <https://doi.org/10.1038/nrm3434>
- Masucci, G. v, Cesano, A., Hawtin, R., Janetzki, S., Zhang, J., Kirsch, I., Dobbin, K. K., Alvarez, J., Robbins, P. B., Selvan, S. R., Streicher, H. Z., Butterfield, L. H., & Thurin, M. (2016). Validation of biomarkers to predict response to immunotherapy in cancer:

Volume I — pre-analytical and analytical validation. *Journal for ImmunoTherapy of Cancer*, 4(1), 76. <https://doi.org/10.1186/s40425-016-0178-1>

Mazard, T., Causse, A., Simony, J., Leconet, W., Vezzio-Vie, N., Torro, A., Jarlier, M., Evrard, A., del Rio, M., & Assenat, E. (2013). Sorafenib Overcomes Irinotecan Resistance in Colorectal Cancer by Inhibiting the ABCG2 Drug-Efflux Pump Sorafenib Inhibits ABCG2 and Overcomes Irinotecan Resistance. *Molecular Cancer Therapeutics*, 12(10), 2121–2134.

Medico, E., Russo, M., Picco, G., Cancelliere, C., Valtorta, E., Corti, G., Buscarino, M., Isella, C., Lamba, S., Martinoglio, B., Veronese, S., Siena, S., Sartore-Bianchi, A., Beccuti, M., Mottolese, M., Linnebacher, M., Cordero, F., di Nicolantonio, F., & Bardelli, A. (2015). The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. *Nature Communications*, 6(1), 7002. <https://doi.org/10.1038/ncomms8002>

Meijer, C., Mulder, N. H., Timmer-Bosscha, H., Sluiter, W. J., Meersma, G. J., & de Vries, E. G. E. (1992). Relationship of cellular glutathione to the cytotoxicity and resistance of seven platinum compounds. *Cancer Research*, 52(24), 6885–6889.

Metsalu, T., & Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research*, 43(W1), W566–W570.

Milazzo, G., Mercatelli, D., Di Muzio, G., Triboli, L., De Rosa, P., Perini, G., & Giorgi, F. M. (2020). Histone Deacetylases (HDACs): Evolution, Specificity, Role in Transcriptional Complexes, and Pharmacological Actionability. In *Genes* (Vol. 11, Issue 5). <https://doi.org/10.3390/genes11050556>

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., Jemal, A., Kramer, J. L., & Siegel, R. L. (2019). Cancer treatment and survivorship statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(5), 363–385.

Mirabelli, P., Coppola, L., & Salvatore, M. (2019). Cancer cell lines are useful model systems for medical research. *Cancers*, 11(8), 1098.

- Mishra, L., Shetty, K., Tang, Y., Stuart, A., & Byers, S. W. (2005). The role of TGF- β and Wnt signaling in gastrointestinal stem cells and cancer. *Oncogene*, *24*(37), 5775–5789.
- Miyaki, M., Iijima, T., Konishi, M., Sakai, K., Ishii, A., Yasuno, M., Hishima, T., Koike, M., Shitara, N., & Iwama, T. (1999). Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. *Oncogene*, *18*(20), 3098–3103.
- Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *Journal of Immunological Methods*, *65*(1–2), 55–63.
- Mouradov, D., Sloggett, C., Jorissen, R. N., Love, C. G., Li, S., Burgess, A. W., Arango, D., Strausberg, R. L., Buchanan, D., Wormald, S., O'Connor, L., Wilding, J. L., Bicknell, D., Tomlinson, I. P. M., Bodmer, W. F., Mariadason, J. M., & Sieber, O. M. (2014). Colorectal Cancer Cell Lines Are Representative Models of the Main Molecular Subtypes of Primary Cancer. *Cancer Research*, *74*(12), 3238. <https://doi.org/10.1158/0008-5472.CAN-14-0013>
- Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., Kovar, C. L., Lewis, L. R., Morgan, M. B., Newsham, I. F., Reid, J. G., Santibanez, J., Shinbrot, E., Trevino, L. R., Wu, Y.-Q., Wang, M., Gunaratne, P., Donehower, L. A., Creighton, C. J., ... group, T. source sites and disease working. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, *487*(7407), 330–337. <https://doi.org/10.1038/nature11252>
- Nassif, N. T., Lobo, G. P., Wu, X., Henderson, C. J. A., Morrison, C. D., Eng, C., Jalaludin, B., & Segelov, E. (2004). PTEN mutations are common in sporadic microsatellite stable colorectal cancer. *Oncogene*, *23*(2), 617–628. <https://doi.org/10.1038/sj.onc.1207059>
- Nazemalhosseini Mojarad, E., Kuppen, P. J., Aghdaei, H. A., & Zali, M. R. (2013). The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterology and Hepatology from Bed to Bench*, *6*(3), 120–128. <https://pubmed.ncbi.nlm.nih.gov/24834258>
- NCI, N. (2015). *Advances in Colorectal Cancer Research*.

- Neubig, R. R., Spedding, M., Kenakin, T., & Christopoulos, A. (2003). International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on terms and symbols in quantitative pharmacology. *Pharmacological Reviews*, 55(4), 597–606.
- Neufeld, K. L., Zhang, F., Cullen, B. R., & White, R. L. (2000). APC-mediated downregulation of beta-catenin activity involves nuclear sequestration and nuclear export. *EMBO Reports*, 1(6), 519–523. <https://doi.org/10.1093/embo-reports/kvd117>
- Ng, P. K.-S., Li, J., Jeong, K. J., Shao, S., Chen, H., Tsang, Y. H., Sengupta, S., Wang, Z., Bhavana, V. H., Tran, R., Soewito, S., Minussi, D. C., Moreno, D., Kong, K., Dogruluk, T., Lu, H., Gao, J., Tokheim, C., Zhou, D. C., ... Mills, G. B. (2018). Systematic Functional Annotation of Somatic Mutations in Cancer. *Cancer Cell*, 33(3), 450-462.e10. <https://doi.org/https://doi.org/10.1016/j.ccell.2018.01.021>
- Nielsen, F. (2016). Hierarchical Clustering. In F. Nielsen (Ed.), *Introduction to HPC with MPI for Data Science* (pp. 195–211). Springer International Publishing. https://doi.org/10.1007/978-3-319-21903-5_8
- Ogino, S., Cantor, M., Kawasaki, T., Brahmandam, M., Kirkner, G. J., Weisenberger, D. J., Campan, M., Laird, P. W., Loda, M., & Fuchs, C. S. (2006). CpG island methylator phenotype (CIMP) of colorectal cancer is best characterised by quantitative DNA methylation analysis and prospective cohort studies. *Gut*, 55(7), 1000–1006.
- Ogino, S., Chan, A. T., Fuchs, C. S., & Giovannucci, E. (2011). Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut*, 60(3), 397–411.
- Oliveira, A. F., Bretes, L., & Furtado, I. (2019). Review of PD-1/PD-L1 Inhibitors in Metastatic dMMR/MSI-H Colorectal Cancer. *Frontiers in Oncology*, 9, 396. <https://www.frontiersin.org/article/10.3389/fonc.2019.00396>
- Overman, M. J., Ernstoff, M. S., & Morse, M. A. (2018). Where We Stand With Immunotherapy in Colorectal Cancer: Deficient Mismatch Repair, Proficient Mismatch

Repair, and Toxicity Management. *American Society of Clinical Oncology Educational Book*, 38, 239–247. https://doi.org/10.1200/EDBK_200821

Overman, M. J., Lonardi, S., Wong, K. Y. M., Lenz, H.-J., Gelsomino, F., Aglietta, M., Morse, M. A., van Cutsem, E., McDermott, R., & Hill, A. (2018). *Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-deficient/microsatellite instability-high metastatic colorectal cancer.*

Overman, M. J., McDermott, R., Leach, J. L., Lonardi, S., Lenz, H.-J., Morse, M. A., Desai, J., Hill, A., Axelson, M., & Moss, R. A. (2017). Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *The Lancet Oncology*, 18(9), 1182–1191.

Oxnard, G. R., Wilcox, K. H., Gonen, M., Polotsky, M., Hirsch, B. R., & Schwartz, L. H. (2016). Response rate as a regulatory end point in single-arm studies of advanced solid tumors. *JAMA Oncology*, 2(6), 772–779. <https://doi.org/10.1001/jamaoncol.2015.6315>

Palshof, J. A., Høgdall, E. V. S., Poulsen, T. S., Linnemann, D., Jensen, B. V., Pfeiffer, P., Tarpgaard, L. S., Brüner, N., Stenvang, J., Yilmaz, M., & Nielsen, D. L. (2017). Topoisomerase I copy number alterations as biomarker for irinotecan efficacy in metastatic colorectal cancer. *BMC Cancer*, 17(1), 48. <https://doi.org/10.1186/s12885-016-3001-y>

Paré, L., Marcuello, E., Altés, A., del Río, E., Sedano, L., Salazar, J., Cortés, A., Barnadas, A., & Baiget, M. (2008). Pharmacogenetic prediction of clinical outcome in advanced colorectal cancer patients receiving oxaliplatin/5-fluorouracil as first-line chemotherapy. *British Journal of Cancer*, 99(7), 1050–1055. <https://doi.org/10.1038/sj.bjc.6604671>

Petitprez, A., & K Larsen, A. (2013). Irinotecan resistance is accompanied by upregulation of EGFR and Src signaling in human cancer models. *Current Pharmaceutical Design*, 19(5), 958–964.

Phua, L. C., Mal, M., Koh, P. K., Cheah, P. Y., Chan, E. C. Y., & Ho, H. K. (2013). Investigating the role of nucleoside transporters in the resistance of colorectal cancer to

- 5-fluorouracil therapy. *Cancer Chemotherapy and Pharmacology*, 71(3), 817–823. <https://doi.org/10.1007/s00280-012-2054-0>
- Pino, M. S., & Chung, D. C. (2010). The chromosomal instability pathway in colon cancer. *Gastroenterology*, 138(6), 2059–2072. <https://doi.org/10.1053/j.gastro.2009.12.065>
- Polakis, P. (2000). Wnt signaling and cancer. *Genes & Development*, 14(15), 1837–1851.
- Popat, S., Hubner, R., & Houlston, R. S. (2005). Systematic review of microsatellite instability and colorectal cancer prognosis. *Journal of Clinical Oncology*, 23(3), 609–618.
- Porter, E. M., Bevins, C. L., Ghosh, D., & Ganz, T. (2002). The multifaceted Paneth cell. *Cellular and Molecular Life Sciences CMLS*, 59(1), 156–170.
- Potter, L. R. (2011). Guanylyl cyclase structure, function and regulation. *Cellular Signalling*, 23(12), 1921–1926.
- Prior, I. A., Lewis, P. D., & Mattos, C. (2012). A comprehensive survey of Ras mutations in cancer. *Cancer Research*, 72(10), 2457–2467. <https://doi.org/10.1158/0008-5472.CAN-11-2612>
- Qiu, L., Tang, Q., Bai, J., Qian, X., Li, R., Liu, B., & Zheng, M. (2008). Predictive value of thymidylate synthase expression in advanced colorectal cancer patients receiving fluoropyrimidine-based chemotherapy: evidence from 24 studies. *International Journal of Cancer*, 123(10), 2384–2389.
- Quayle, F. J., & Lowney, J. K. (2006). Colorectal lymphoma. *Clinics in Colon and Rectal Surgery*, 19(02), 49–53.
- Raymond, E., Chaney, S. G., Taamma, A., & Cvitkovic, E. (1998). Oxaliplatin: a review of preclinical and clinical studies. *Annals of Oncology*, 9(10), 1053–1071.
- Raymond, E., Faivre, S., Woynarowski, J. M., & Chaney, S. G. (1998). Oxaliplatin: mechanism of action and antineoplastic activity. *Seminars in Oncology*, 25(2 Suppl 5), 4–12.
- Rees, M. G., Seashore-Ludlow, B., Cheah, J. H., Adams, D. J., Price, E. v, Gill, S., Javaid, S., Coletti, M. E., Jones, V. L., Bodycombe, N. E., Soule, C. K., Alexander, B., Li, A.,

- Montgomery, P., Kotz, J. D., Hon, C. S.-Y., Munoz, B., Liefeld, T., Dančik, V., ... Schreiber, S. L. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nature Chemical Biology*, *12*(2), 109–116. <https://doi.org/10.1038/nchembio.1986>
- Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., & Mesirov, J. P. (2006). GenePattern 2.0. *Nature Genetics*, *38*(5), 500–501.
- Richman, S. D., Seymour, M. T., Chambers, P., Elliott, F., Daly, C. L., Meade, A. M., Taylor, G., Barrett, J. H., & Quirke, P. (2009). KRAS and BRAF Mutations in Advanced Colorectal Cancer Are Associated With Poor Prognosis but Do Not Preclude Benefit From Oxaliplatin or Irinotecan: Results From the MRC FOCUS Trial. *Journal of Clinical Oncology*, *27*(35), 5931–5937. <https://doi.org/10.1200/JCO.2009.22.4295>
- Rojas, A., Gonzalez, I., & Figueroa, H. (2008). Cell line cross-contamination in biomedical research: a call to prevent unawareness. *Acta Pharmacologica Sinica*, *29*(7), 877–880.
- Rottenberg, S., Disler, C., & Perego, P. (2021). The rediscovery of platinum-based cancer therapy. *Nature Reviews Cancer*, *21*(1), 37–50.
- Rougier, P., Bugat, R., Douillard, J. Y., Culine, S., Suc, E., Brunet, P., Becouarn, Y., Ychou, M., Marty, M., & Extra, J. M. (1997). Phase II study of irinotecan in the treatment of advanced colorectal cancer in chemotherapy-naïve patients and patients pretreated with fluorouracil-based chemotherapy. *Journal of Clinical Oncology*, *15*(1), 251–260.
- Roumeliotis, T. I., Williams, S. P., Gonçalves, E., Alsinet, C., del Castillo Velasco-Herrera, M., Aben, N., Ghavidel, F. Z., Michaut, M., Schubert, M., Price, S., Wright, J. C., Yu, L., Yang, M., Dienstmann, R., Guinney, J., Beltrao, P., Brazma, A., Pardo, M., Stegle, O., ... Choudhary, J. S. (2017). Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. *Cell Reports*, *20*(9), 2201–2214. <https://doi.org/https://doi.org/10.1016/j.celrep.2017.08.010>
- Rowan, A. J., Lamlum, H., Ilyas, M., Wheeler, J., Straub, J., Papadopoulou, A., Bicknell, D., Bodmer, W. F., & Tomlinson, I. P. (2000). APC mutations in sporadic colorectal tumors: A mutational “hotspot” and interdependence of the “two hits.” *Proceedings of the*

National Academy of Sciences of the United States of America, 97(7), 3352–3357.
<https://doi.org/10.1073/pnas.97.7.3352>

Saha, M. N., Qiu, L., & Chang, H. (2013). Targeting p53 by small molecules in hematological malignancies. *Journal of Hematology & Oncology*, 6(1), 1–9.

Sakowicz-Burkiewicz, M., Przybyla, T., Wesslering, M., Bielarczyk, H., Maciejewska, I., & Pawelczyk, T. (2016). Suppression of TWIST1 enhances the sensitivity of colon cancer cells to 5-fluorouracil. *The International Journal of Biochemistry & Cell Biology*, 78, 268–278.

Salz, T., Deng, C., Pampo, C., Siemann, D., Qiu, Y., Brown, K., & Huang, S. (2015). Histone methyltransferase hSETD1A is a novel regulator of metastasis in breast cancer. *Molecular Cancer Research*, 13(3), 461–469.

Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., Yan, H., Gazdar, A., Powell, S. M., & Riggins, G. J. (2004). High frequency of mutations of the PIK3CA gene in human cancers. *Science*, 304(5670), 554.

Sargent, D. J., Marsoni, S., Monges, G., Thibodeau, S. N., Labianca, R., Hamilton, S. R., French, A. J., Kabat, B., Foster, N. R., Torri, V., Ribic, C., Grothey, A., Moore, M., Zaniboni, A., Seitz, J.-F., Sinicrope, F., & Gallinger, S. (2010). Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 28(20), 3219–3226. <https://doi.org/10.1200/JCO.2009.27.1825>

Sato, T., van Es, J. H., Snippert, H. J., Stange, D. E., Vries, R. G., van den Born, M., Barker, N., Shroyer, N. F., van de Wetering, M., & Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature*, 469(7330), 415–418.

Schirmacher, V. (2019). From chemotherapy to biological therapy: A review of novel concepts to reduce the side effects of systemic cancer treatment. *International Journal of Oncology*, 54(2), 407–419.

Schneikert, J., & Behrens, J. (2007). The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut*, 56(3), 417–425.

- Seashore-Ludlow, B., Rees, M. G., Cheah, J. H., Cokol, M., Price, E. v, Coletti, M. E., Jones, V., Bodycombe, N. E., Soule, C. K., & Gould, J. (2015). Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discovery*, *5*(11), 1210–1223.
- Seyfried, N. T., Dammer, E. B., Swarup, V., Nandakumar, D., Duong, D. M., Yin, L., Deng, Q., Nguyen, T., Hales, C. M., & Wingo, T. (2017). A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer’s disease. *Cell Systems*, *4*(1), 60–72.
- Shao, J., Xu, Z., Peng, X., Chen, M., Zhu, Y., Xu, L., Zhu, H., Yang, B., Luo, P., & He, Q. (2016). Gefitinib synergizes with irinotecan to suppress hepatocellular carcinoma via antagonizing Rad51-mediated DNA-repair. *PLoS One*, *11*(1), e0146968.
- Shaukat, A., Kahi, C. J., Burke, C. A., Rabeneck, L., Sauer, B. G., & Rex, D. K. (2021). ACG Clinical Guidelines: Colorectal Cancer Screening 2021. *Official Journal of the American College of Gastroenterology / ACG*, *116*(3).
- Sherr, C. J. (2004). Principles of tumor suppression. *Cell*, *116*(2), 235–246. [https://doi.org/10.1016/s0092-8674\(03\)01075-4](https://doi.org/10.1016/s0092-8674(03)01075-4)
- Shi, X.-H., Tan, Z., Shi, J., Gu, Y., Zuo, Q., Xu, X.-W., Tang, L.-J., & Yu, E.-D. (2017). Expression of guanylyl cyclase C and its prognostic value in different developing stages of colorectal cancer patients. *Int J Clin Exp Pathol*, *10*(2), 1918–1927.
- Shiovitz, S., Bertagnolli, M. M., Renfro, L. A., Nam, E., Foster, N. R., Dzieciatkowski, S., Luo, Y., Lao, V. V., Monnat Jr, R. J., & Emond, M. J. (2014). CpG island methylator phenotype is associated with response to adjuvant irinotecan-based therapy for stage III colon cancer. *Gastroenterology*, *147*(3), 637–645.
- Shirota, Y., Stoehlmacher, J., Brabender, J., Xiong, Y.-P., Uetake, H., Danenberg, K. D., Groshen, S., Tsao-Wei, D. D., Danenberg, P. v, & Lenz, H.-J. (2001). ERCC1 and Thymidylate Synthase mRNA Levels Predict Survival for Colorectal Cancer Patients Receiving Combination Oxaliplatin and Fluorouracil Chemotherapy. *Journal of Clinical Oncology*, *19*(23), 4298–4304. <https://doi.org/10.1200/JCO.2001.19.23.4298>

- Shoemaker, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer*, 6(10), 813–823.
- Siegel, R. L., Torre, L. A., Soerjomataram, I., Hayes, R. B., Bray, F., Weber, T. K., & Jemal, A. (2019). Global patterns and trends in colorectal cancer incidence in young adults. *Gut*, 68(12), 2179–2185. <https://doi.org/10.1136/gutjnl-2019-319511>
- Skehan, P., Storeng, R., Scudiero, D., Monks, A., McMahon, J., Vistica, D., Warren, J. T., Bokesch, H., Kenney, S., & Boyd, M. R. (1990). New colorimetric cytotoxicity assay for anticancer-drug screening. *JNCI: Journal of the National Cancer Institute*, 82(13), 1107–1112.
- Smit, W. L., Spaan, C. N., Johannes de Boer, R., Ramesh, P., Martins Garcia, T., Meijer, B. J., Vermeulen, J. L. M., Lezzerini, M., MacInnes, A. W., & Koster, J. (2020). Driver mutations of the adenoma-carcinoma sequence govern the intestinal epithelial global translational capacity. *Proceedings of the National Academy of Sciences*, 117(41), 25560–25570.
- Spindler, K. G., Appelt, A. L., Pallisgaard, N., Andersen, R. F., & Jakobsen, A. (2013). KRAS-mutated plasma DNA as predictor of outcome from irinotecan monotherapy in metastatic colorectal cancer. *British Journal of Cancer*, 109(12), 3067–3072. <https://doi.org/10.1038/bjc.2013.633>
- Stacey E. Mills. (2007). *Histology for Pathologists: 3rd (third) Edition*.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Tejpar, S., Peeters, M., Humblet, Y., Vermorken, J. B., de Hertogh, G., de Roock, W., Nippgen, J., von Heydebreck, A., Stroh, C., & van Cutsem, E. (2008). Relationship of efficacy with KRAS status (wild type versus mutant) in patients with irinotecan-refractory metastatic colorectal cancer (mCRC), treated with irinotecan (q2w) and escalating doses of

cetuximab (q1w): The EVEREST experience (preliminary data). *Journal of Clinical Oncology*, 26(15_suppl), 4001. https://doi.org/10.1200/jco.2008.26.15_suppl.4001

The Editors of Encyclopaedia. (2022, September 15). *Small intestine*. Encyclopedia Britannica.

Therasse, P., Arbuck, S. G., Eisenhauer, E. A., Wanders, J., Kaplan, R. S., Rubinstein, L., Verweij, J., van Glabbeke, M., van Oosterom, A. T., & Christian, M. C. (2000). New guidelines to evaluate the response to treatment in solid tumors. *Journal of the National Cancer Institute*, 92(3), 205–216.

Troiani, T., Napolitano, S., & Normanno, N. (n.d.). *RAS in colorectal cancer: ESMO biomarker factsheet*.

v B ASTLER, & F A COLLER. (1954). *The prognostic significance of direct extension of carcinoma of the colon and rectum*. <https://doi.org/10.1097/00000658-195406000-00015>

van Cutsem, E., Cervantes, A., Adam, R., Sobrero, A., van Krieken, J. H., Aderka, D., Aguilar, E. A., Bardelli, A., Benson, A., & Bodoky, G. (2016). ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Annals of Oncology*, 27(8), 1386–1422.

van Cutsem, E., Köhne, C.-H., Láng, I., Folprecht, G., Nowacki, M. P., Cascinu, S., Shchepotin, I., Maurel, J., Cunningham, D., & Tejpar, S. (2011). Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of overall survival according to tumor KRAS and BRAF mutation status. *J Clin Oncol*, 29(15), 2011–2019.

van Parijs, L., & Abbas, A. K. (1998). Homeostasis and self-tolerance in the immune system: turning lymphocytes off. *Science*, 280(5361), 243–248.

Vichai, V., & Kirtikara, K. (2006). Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nature Protocols*, 1(3), 1112–1116.

Vilar, E., Scaltriti, M., Balmaña, J., Saura, C., Guzman, M., Arribas, J., Baselga, J., & Tabernero, J. (2008). Microsatellite instability due to hMLH1 deficiency is associated

- with increased cytotoxicity to irinotecan in human colorectal cancer cell lines. *British Journal of Cancer*, 99(10), 1607–1612. <https://doi.org/10.1038/sj.bjc.6604691>
- Vinay, D. S., Ryan, E. P., Pawelec, G., Talib, W. H., Stagg, J., Elkord, E., Lichtor, T., Decker, W. K., Whelan, R. L., Kumara, H. M. C. S., Signori, E., Honoki, K., Georgakilas, A. G., Amin, A., Helferich, W. G., Boosani, C. S., Guha, G., Ciriolo, M. R., Chen, S., ... Kwon, B. S. (2015). Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Seminars in Cancer Biology*, 35, S185–S198. <https://doi.org/https://doi.org/10.1016/j.semcancer.2015.03.004>
- Waldman, S. A., Hyslop, T., Schulz, S., Barkun, A., Nielsen, K., Haaf, J., Bonaccorso, C., Li, Y., & Weinberg, D. S. (2009). Association of GUCY2C expression in lymph nodes with time to recurrence and disease-free survival in pN0 colorectal cancer. *Jama*, 301(7), 745–752.
- Wang, D., Zou, L., Jin, Q., Hou, J., Ge, G., & Yang, L. (2018). Human carboxylesterases: a comprehensive review. *Acta Pharmaceutica Sinica B*, 8(5), 699–712. <https://doi.org/https://doi.org/10.1016/j.apsb.2018.05.005>
- Wang, J., Mouradov, D., Wang, X., Jorissen, R. N., Chambers, M. C., Zimmerman, L. J., Vasaikar, S., Love, C. G., Li, S., Lowes, K., Leuchowius, K. J., Jousset, H., Weinstock, J., Yau, C., Mariadason, J., Shi, Z., Ban, Y., Chen, X., Coffey, R. J. C., ... Sieber, O. M. (2017). Colorectal Cancer Cell Line Proteomes Are Representative of Primary Tumors and Predict Drug Sensitivity. *Gastroenterology*, 153(4), 1082–1095. <https://doi.org/10.1053/j.gastro.2017.06.008>
- Wang, L.-H., Wu, C.-F., Rajasekaran, N., & Shin, Y. K. (2018). Loss of Tumor Suppressor Gene Function in Human Cancer: An Overview. *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*, 51(6), 2647–2693. <https://doi.org/10.1159/000495956>
- Wang, W., Kandimalla, R., Huang, H., Zhu, L., Li, Y., Gao, F., Goel, A., & Wang, X. (2019). Molecular subtyping of colorectal cancer: Recent progress, new challenges and emerging opportunities. *Seminars in Cancer Biology*, 55, 37–52.

- Wang, Y. (2009). Wnt/Planar cell polarity signaling: A new paradigm for cancer therapy. *Molecular Cancer Therapeutics*, 8(8), 2103–2109. <https://doi.org/10.1158/1535-7163.MCT-09-0282>
- Wolpin, B. M., & Mayer, R. J. (2008). Systemic treatment of colorectal cancer. *Gastroenterology*, 134(5), 1296–1310. <https://doi.org/10.1053/j.gastro.2008.02.098>
- Wu, D. T., Bitzer, M., Ju, W., Mundel, P., & Böttinger, E. P. (2005). TGF- β concentration specifies differential signaling profiles of growth arrest/differentiation and apoptosis in podocytes. *Journal of the American Society of Nephrology*, 16(11), 3211–3221.
- Xie, Y.-H., Chen, Y.-X., & Fang, J.-Y. (2020). Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduction and Targeted Therapy*, 5(1), 22. <https://doi.org/10.1038/s41392-020-0116-z>
- Xu, Y., & Pasche, B. (2007). TGF-beta signaling alterations and susceptibility to colorectal cancer. *Human Molecular Genetics*, 16 Spec No 1(SPEC), R14–R20. <https://doi.org/10.1093/hmg/ddl486>
- Xu, Y., & Villalona-Calero, M. A. (2002). Irinotecan: mechanisms of tumor resistance and novel strategies for modulating its activity. *Annals of Oncology*, 13(12), 1841–1851.
- Yaghoubi, N., Soltani, A., Ghazvini, K., Hassanian, S. M., & Hashemy, S. I. (2019). PD-1/PD-L1 blockade as a novel treatment for colorectal cancer. *Biomedicine & Pharmacotherapy*, 110, 312–318.
- Yanagisawa, Y., Maruta, F., Iinuma, N., Ishizone, S., Koide, N., Nakayama, J., & Miyagawa, S. (2007). Modified Irinotecan/5FU/Leucovorin therapy in advanced colorectal cancer and predicting therapeutic efficacy by expression of tumor-related enzymes. *Scandinavian Journal of Gastroenterology*, 42(4), 477–484.
- Yang, H., Sun, L., Liu, M., & Mao, Y. (2018). Patient-derived organoids: A promising model for personalized cancer treatment. In *Gastroenterology report* (Vol. 6, Issue 4, pp. 243–245). Oxford University Press.

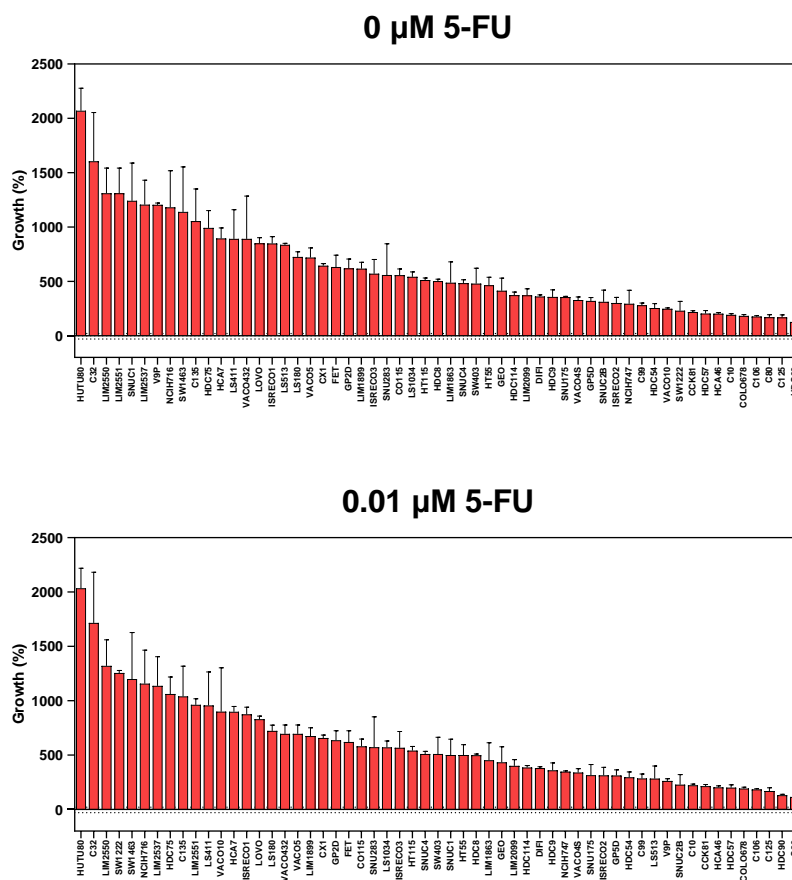
- Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J. A., & Thompson, I. R. (2012). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, *41*(D1), D955–D961.
- Yao, J., Zang, W., Ge, Y., Weygant, N., Yu, P., Li, L., Rao, G., Jiang, Z., Yan, R., He, L., Yu, Y., Jin, M., Cheng, G., & An, G. (2018). RAS/BRAF Circulating Tumor DNA Mutations as a Predictor of Response to First-Line Chemotherapy in Metastatic Colorectal Cancer Patients. *Canadian Journal of Gastroenterology and Hepatology*, *2018*, 4248971. <https://doi.org/10.1155/2018/4248971>
- Yashiro, M., Qiu, H., Hasegawa, T., Zhang, X., Matsuzaki, T., & Hirakawa, K. (2011). An EGFR inhibitor enhances the efficacy of SN38, an active metabolite of irinotecan, in SN38-refractory gastric carcinoma cells. *British Journal of Cancer*, *105*(10), 1522–1532. <https://doi.org/10.1038/bjc.2011.397>
- Yu, J., Wu, W. K. K., Li, X., He, J., Li, X. X., Ng, S. S. M., Yu, C., Gao, Z., Yang, J., Li, M., Wang, Q., Liang, Q., Pan, Y., Tong, J. H., To, K. F., Wong, N., Zhang, N., Chen, J., Lu, Y., ... Sung, J. J. Y. (2015). Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. *Gut*, *64*(4), 636–645. <https://doi.org/10.1136/gutjnl-2013-306620>
- Yuki, S., Gamoh, M., Denda, T., Takashima, A., Takahashi, S., Nakamura, M., Ohori, H., Yamaguchi, T., Kobayashi, Y., Baba, H., Kotake, M., Amagai, K., Kondo, H., Shimada, K., Sato, A., Ishioka, C., Komine, K., Ouchi, K., Morita, S., & Komatsu, Y. (2020). Analysis of consensus molecular subtypes (CMS) classification in the TRICOLORE trial: A randomized phase III trial of S-1 and irinotecan (IRI) plus bevacizumab (Bmab) versus mFOLFOX6 or CapeOX plus Bmab as first-line treatment for metastatic colorectal cancer (mCRC). *Journal of Clinical Oncology*, *38*(4_suppl), 169. https://doi.org/10.1200/JCO.2020.38.4_suppl.169
- Zhan, T., Rindtorff, N., & Boutros, M. (2017). Wnt signaling in cancer. *Oncogene*, *36*(11), 1461–1473. <https://doi.org/10.1038/onc.2016.304>

- Zhang, N., Yin, Y., Xu, S.-J., & Chen, W.-S. (2008). 5-Fluorouracil: mechanisms of resistance and reversal strategies. *Molecules*, *13*(8), 1551–1569.
- Zhang, W., & Liu, H. T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Research*, *12*(1), 9–18.
- Zhang, Y., & Zhang, Z. (2020). The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cellular & Molecular Immunology*, *17*(8), 807–821. <https://doi.org/10.1038/s41423-020-0488-6>
- Zhao, Z., & Shilatifard, A. (2019). Epigenetic modifications of histones in cancer. *Genome Biology*, *20*(1), 1–16.
- Zhu, G., Pei, L., Xia, H., Tang, Q., & Bi, F. (2021). Role of oncogenic KRAS in the prognosis, diagnosis and treatment of colorectal cancer. *Molecular Cancer*, *20*(1), 1–17.

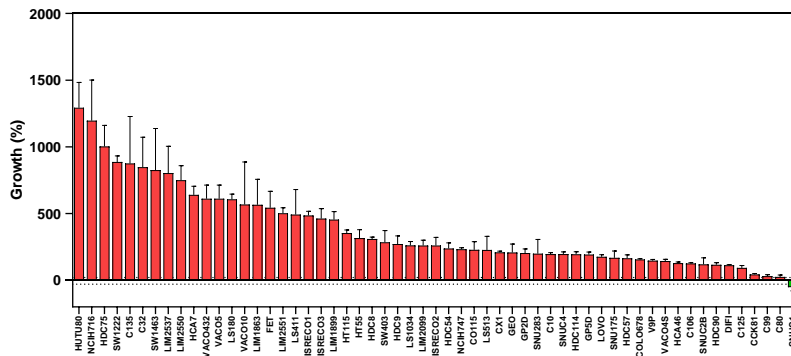
APPENDIX

Appendix 1

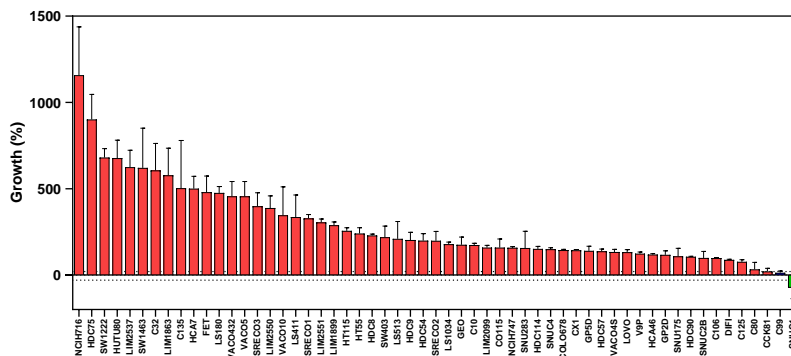
Appendix figure 1: The growth rate of colorectal cancer lines upon all 5-FU doses tested. The absolute growth of each cell line upon 72 hours of treatment with 5-FU was calculated compared to the beginning of the treatment for the 11 doses and the non-treated cells. Columns show the mean (+/- SEM) from 3 independent experiments. The color in the graph represents the cciRECIST categories (red: progressive disease; blue; stable disease; light green: partial response; dark green: complete response).



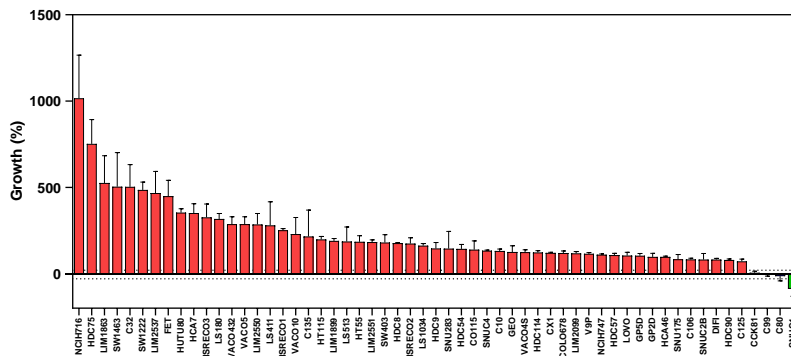
2.5 μ M 5-FU



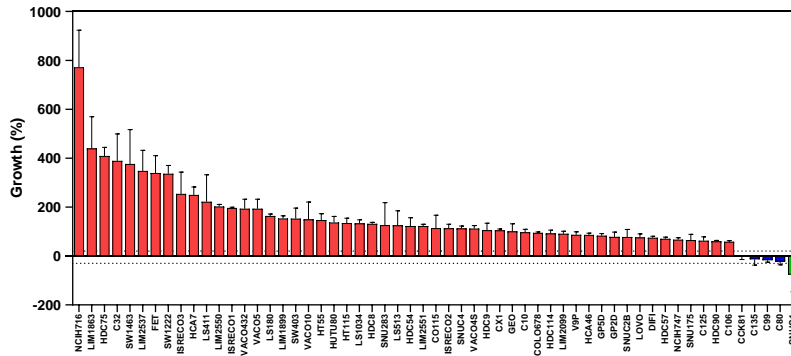
5 μ M 5-FU



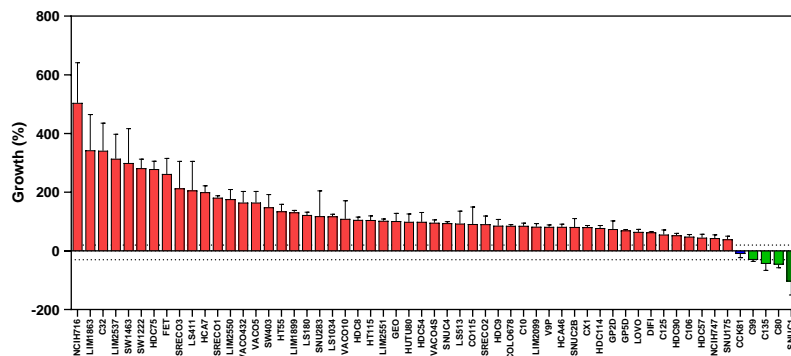
10 μ M 5-FU



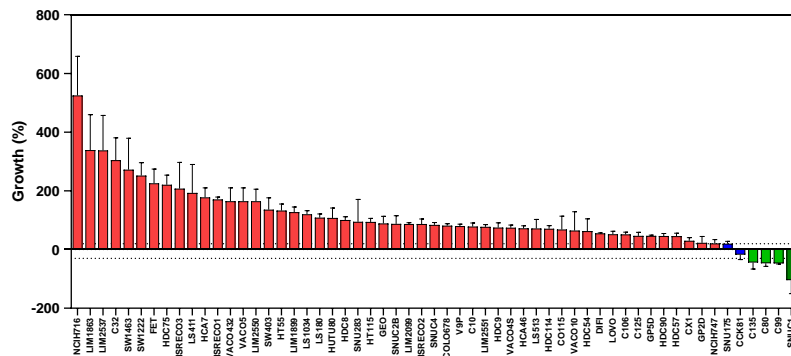
25 μ M 5-FU



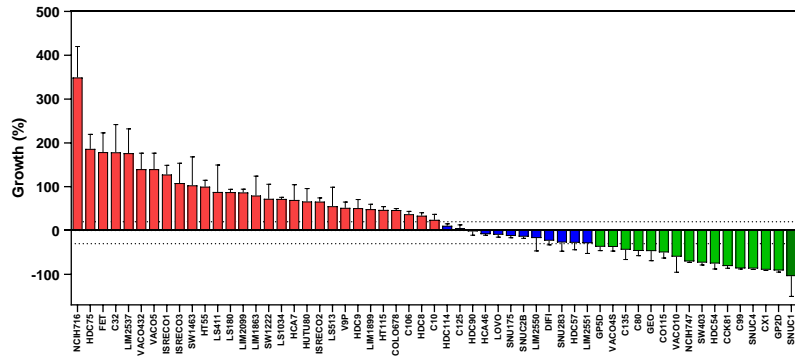
50 μ M 5-FU



100 μ M 5-FU

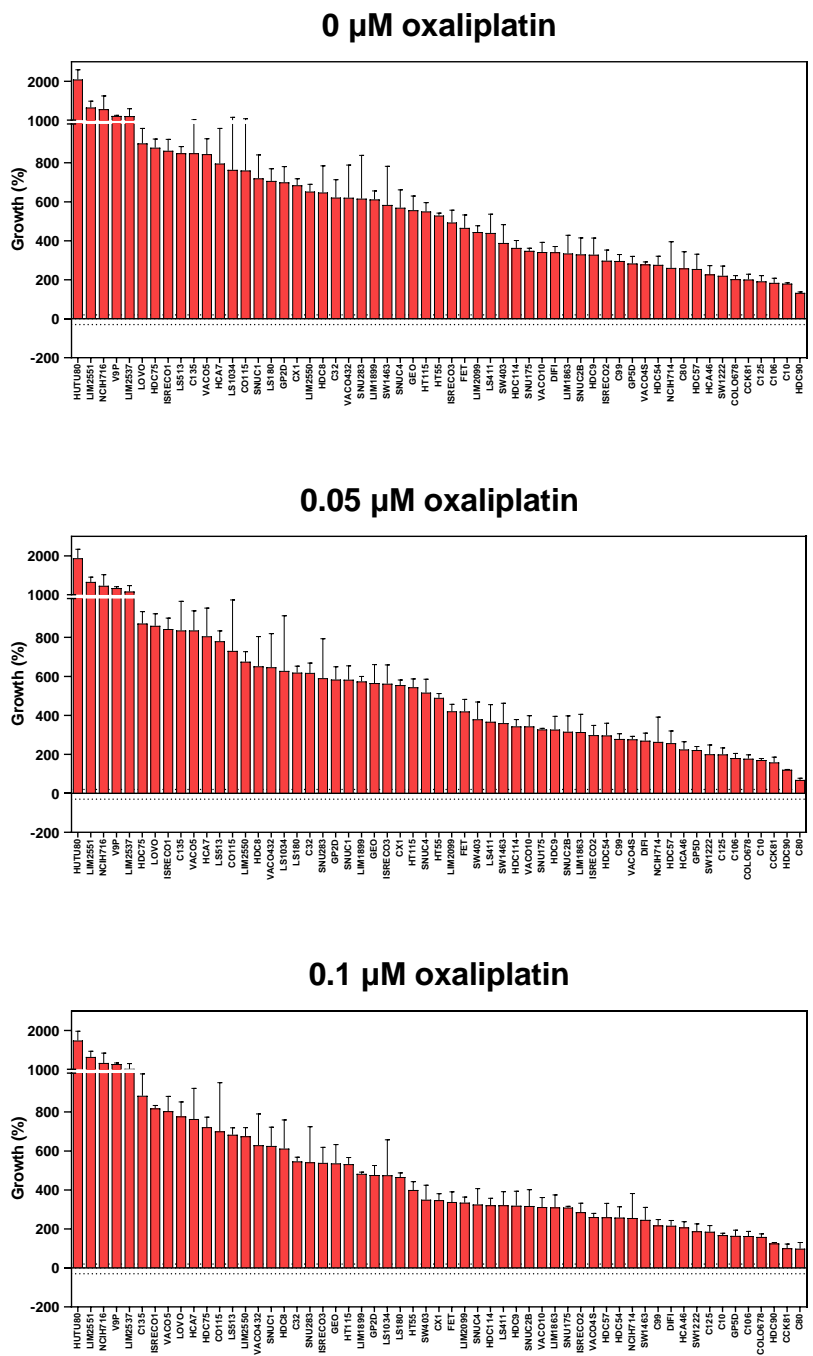


500 μ M 5-FU

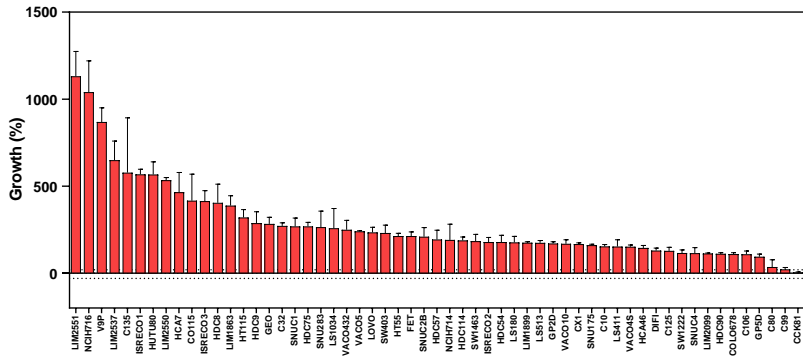


Appendix 2

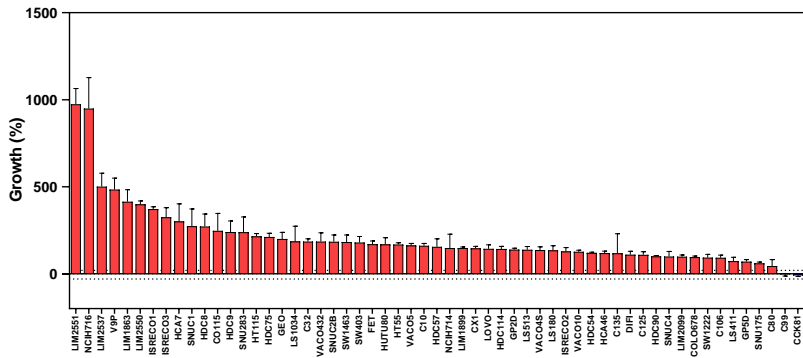
Appendix figure 2: The growth rate of colorectal cancer lines upon all oxaliplatin doses tested. The absolute growth of each cell line upon 72 hours of treatment with oxaliplatin was calculated compared to the beginning of the treatment for the 11 doses and the non-treated cells. Columns show the mean (\pm SEM) from 3 independent experiments. The color in the graph represents the cclRECIST categories (red: progressive disease; blue; stable disease; light green: partial response; dark green: complete response).



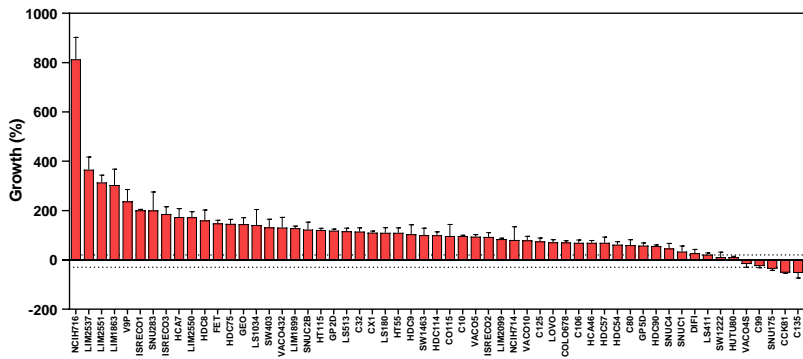
0.5 μ M oxaliplatin



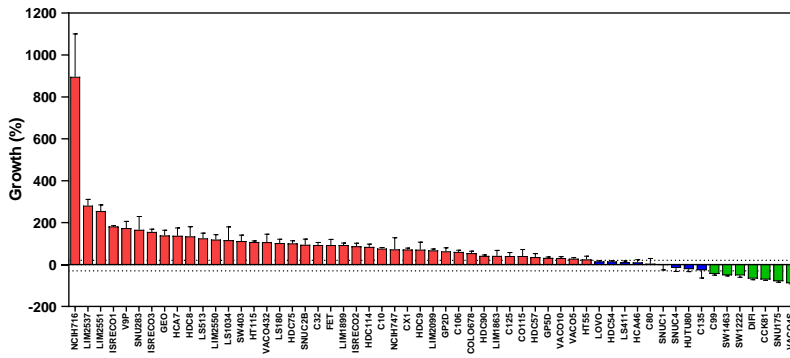
1 μ M oxaliplatin



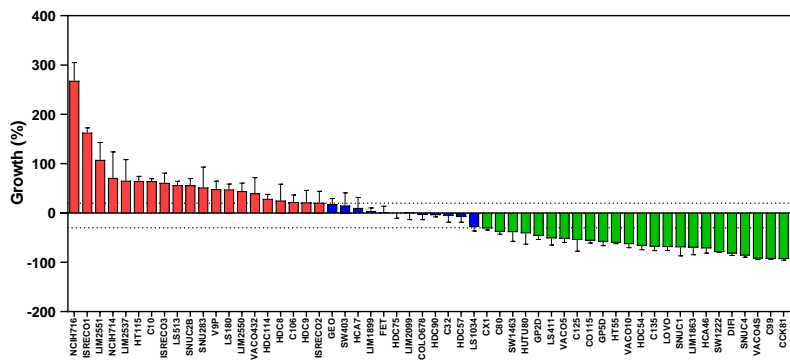
5 μ M oxaliplatin



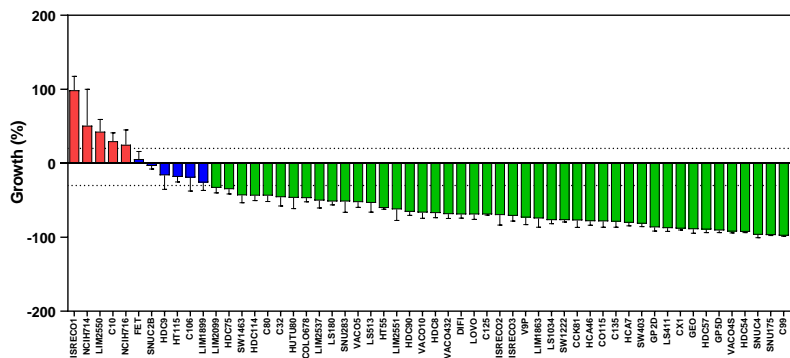
10 μ M oxaliplatin



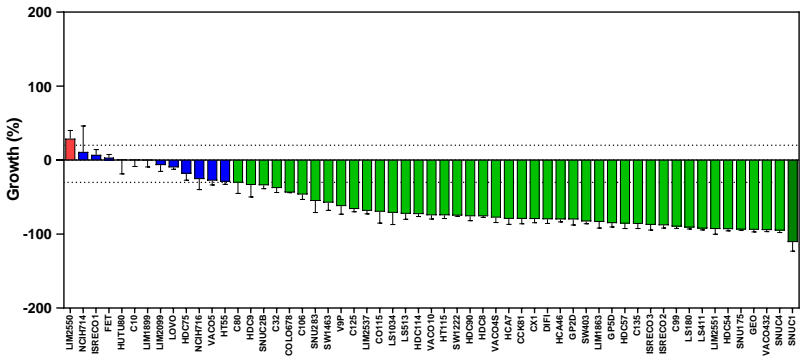
25 μ M oxaliplatin



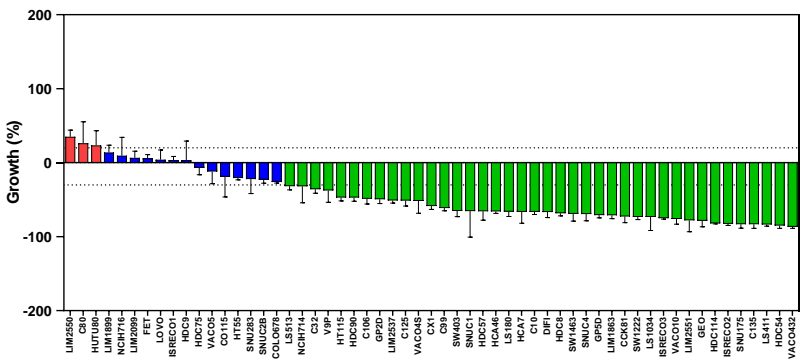
50 μ M oxaliplatin



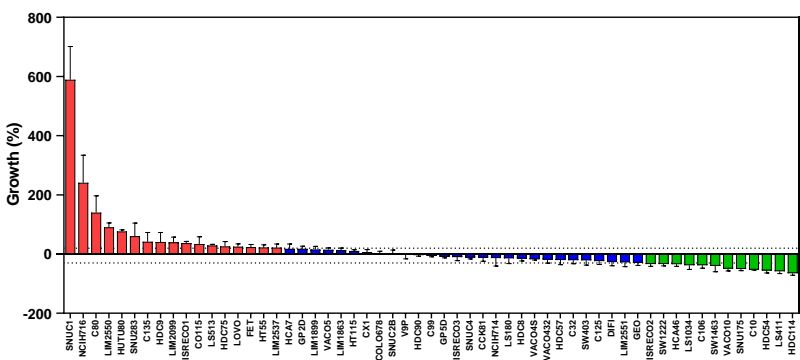
100 μM oxaliplatin



200 μM oxaliplatin

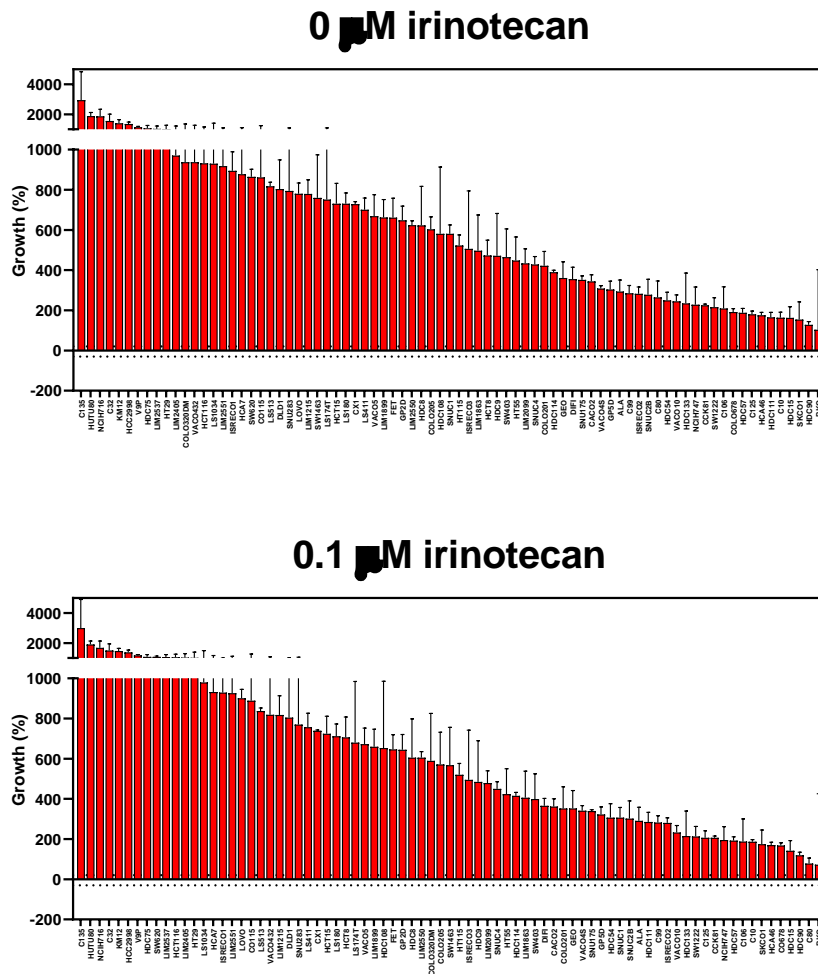


500 μM oxaliplatin

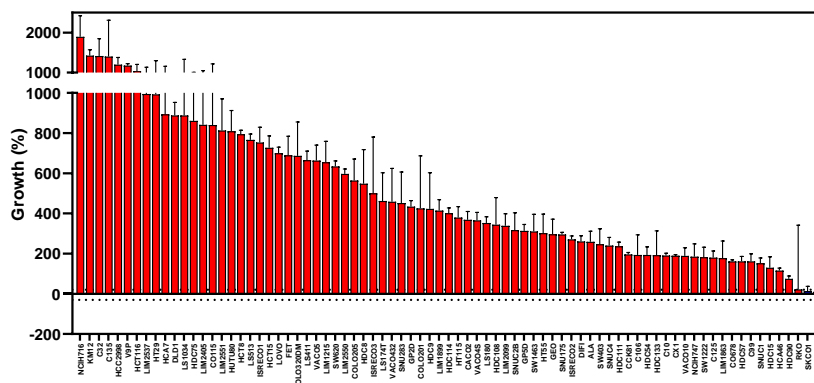


Appendix 3

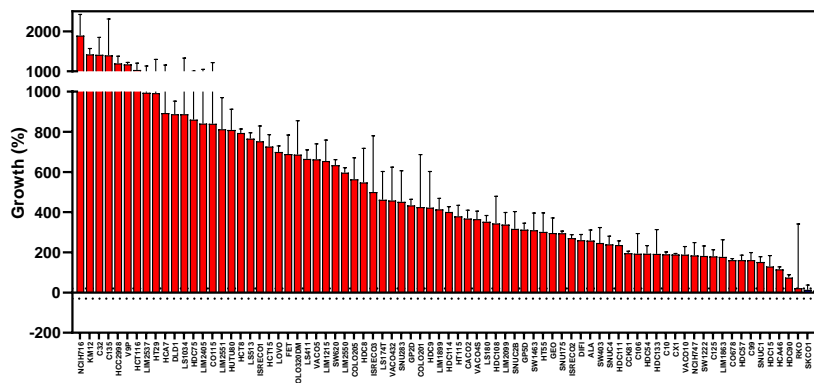
Appendix figure 3: The growth rate of colorectal cancer lines upon all irinotecan doses tested. The absolute growth of each cell line upon 72 hours of treatment with irinotecan was calculated compared to the beginning of the treatment for the 11 doses and the non-treated cells. Columns show the mean (\pm SEM) from 3 independent experiments. The color in the graph represents the cclRECIST categories (red: progressive disease; blue; stable disease; light green: partial response; dark green: complete response).



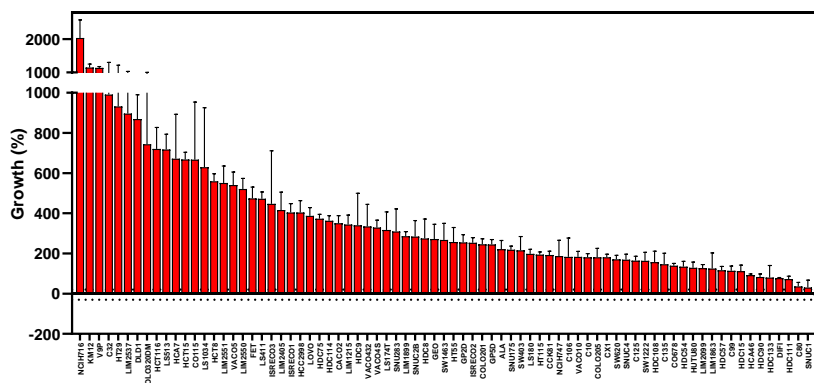
0.5 μ M irinotecan



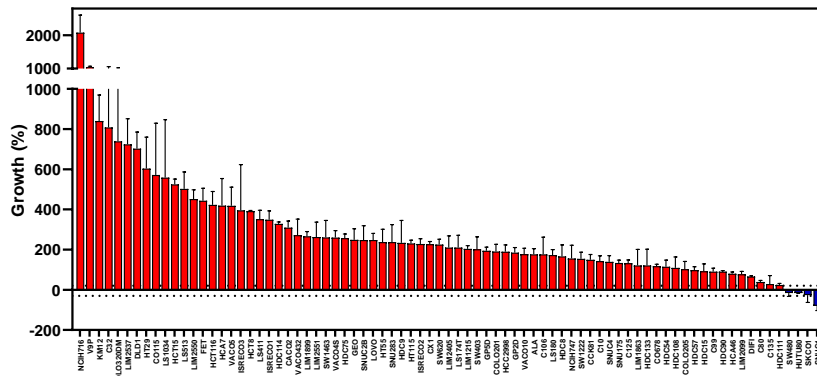
1 μ M irinotecan



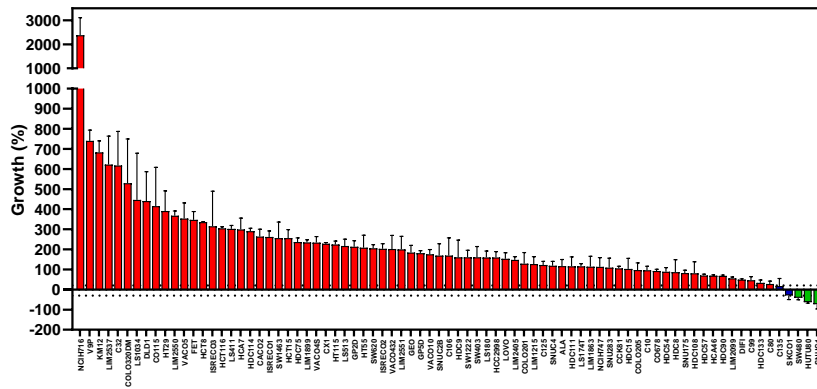
2.5 μ M irinotecan



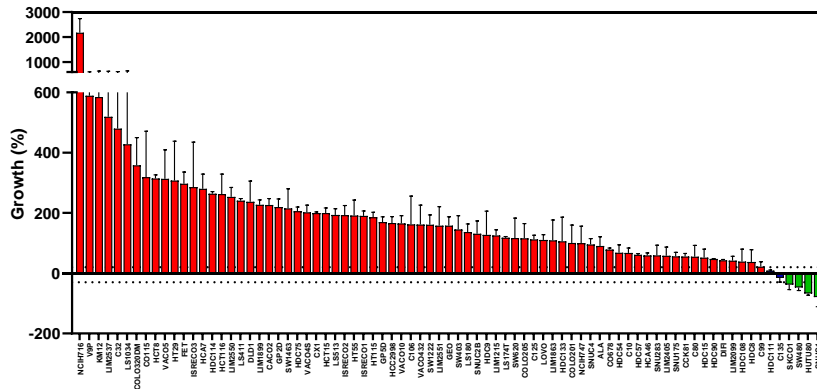
5 μ M irinotecan



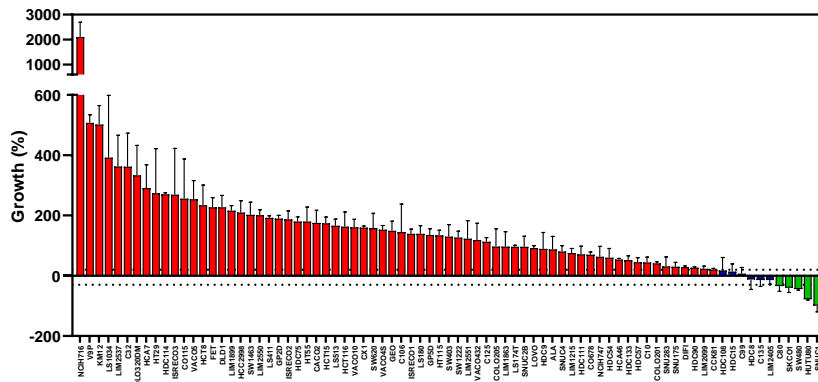
10 μ M irinotecan



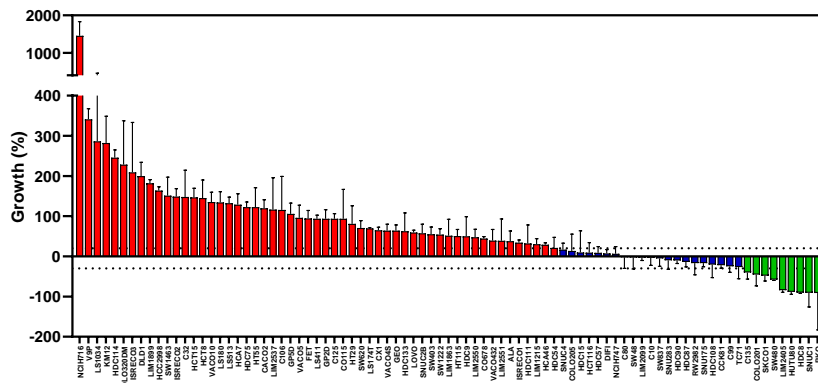
15 μ M irinotecan



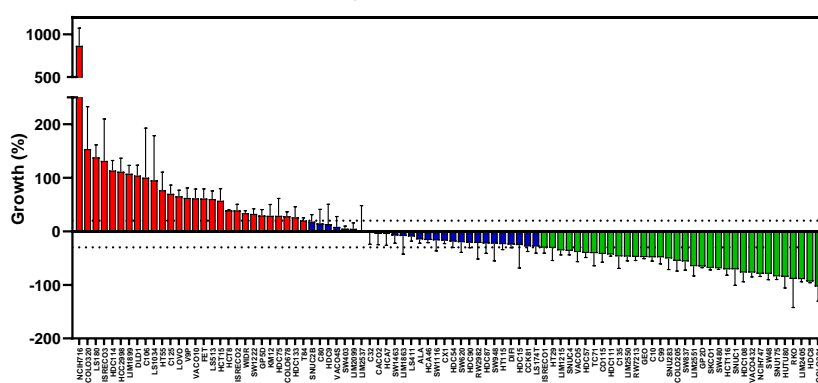
20 μ M irinotecan



40 μ M irinotecan



80 μ M irinotecan



Appendix 4

1 Table 1. List of enriched cellular components using differentially expressed genes in 5-FU sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005759	mitochondrial matrix	839	389	2.65E-05	0.016147	35
GO:0005739	mitochondrion	839	1437	7.87E-05	0.023967	89

2 Table 2. List of enriched molecular functions using differentially expressed genes in 5-FU sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005515	protein binding	826	12554	9.83E-06	0.01049	606

3 Table 3. List of enriched cellular components using differentially expressed proteins in 5-FU sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005739	mitochondrion	188	1437	5.79E-13	1.87E-10	45
GO:0005789	endoplasmic reticulum membrane	188	1054	2.61E-08	3.17E-06	31
GO:0005829	cytosol	188	5510	2.95E-08	3.17E-06	86
GO:0016020	membrane	188	2547	2.03E-07	1.64E-05	50
GO:0005743	mitochondrial inner membrane	188	465	1.29E-06	8.33E-05	18
GO:0005759	mitochondrial matrix	188	389	2.43E-04	0.011548	13
GO:0008250	oligosaccharyltransferase complex	188	14	2.50E-04	0.011548	4

4 Table 4. List of enriched molecular functions using differentially expressed proteins in 5-FU sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0003723	RNA binding	190	1471	7.04E-15	2.86E-12	51
GO:0005515	protein binding	190	12554	3.58E-08	7.29E-06	160

5 Table 5. List of enriched KEGG pathways using differentially expressed proteins in 5-FU sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
hsa00100	Metabolic pathways	105	1540	2.25E-04	0.044148	36

6 Table 6. List of enriched cellular components using differentially expressed proteins in oxaliplatin-sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005829	cytosol	1047	5510	1.60E-12	1.18E-09	382
GO:0005654	nucleoplasm	1047	3963	1.83E-08	6.73E-06	274
GO:0005925	focal adhesion	1047	425	4.97E-07	1.22E-04	48
GO:0005737	cytoplasm	1047	5576	7.33E-07	1.35E-04	354
GO:0005634	nucleus	1047	5984	3.75E-06	5.53E-04	371
GO:0030054	cell junction	1047	232	5.42E-05	0.006657	28
GO:0016020	membrane	1047	2547	2.90E-04	0.030559	168
GO:0032587	ruffle membrane	1047	99	4.77E-04	0.043922	15
GO:0070062	extracellular exosome	1047	2218	6.09E-04	0.049868	147

7 Table 7. List of enriched molecular functions using differentially expressed genes in oxaliplatin sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005515	protein binding	1045	12554	1.64E-13	1.94E-10	801801

8 Table 8. List of enriched cellular components using differentially expressed proteins in oxaliplatin-sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005829	cytosol	263	5510	7.87E-22	3.08E-19	144
GO:0070062	extracellular exosome	263	2218	1.55E-17	3.04E-15	79
GO:0005737	cytoplasm	263	5576	4.27E-12	5.57E-10	124
GO:0005925	focal adhesion	263	425	4.36E-11	4.26E-09	27
GO:0016020	membrane	263	2547	3.78E-09	2.95E-07	68
GO:0048471	perinuclear region of cytoplasm	263	764	1.67E-05	0.001088	26
GO:0032991	macromolecular complex	263	703	3.68E-05	0.002057	24
GO:0005654	nucleoplasm	263	3963	4.82E-05	0.002356	78
GO:0015629	actin cytoskeleton	263	258	1.25E-04	0.005443	13
GO:0032587	ruffle membrane	263	99	2.76E-04	0.010808	8
GO:0005777	peroxisome	263	112	5.85E-04	0.01987	8
GO:0005856	cytoskeleton	263	541	6.10E-04	0.01987	18

GO:0005794	Golgi apparatus	263	1141	7.06E-04	0.021243	29
GO:0099524	postsynaptic cytosol	263	17	0.001218	0.034029	4
GO:0016363	nuclear matrix	263	129	0.001345	0.035053	8
GO:0031982	vesicle	263	174	0.001798	0.042922	9
GO:0001726	ruffle	263	101	0.001866	0.042922	7
GO:0005789	endoplasmic reticulum membrane	263	1054	0.002153	0.046763	26

9 Table 9. List of enriched molecular functions using differentially expressed proteins in oxaliplatin sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0045296	cadherin binding	261	316	1.39E-11	7.83E-09	25
GO:0005515	protein binding	261	12554	1.94E-08	5.49E-06	214
GO:0003779	actin binding	261	339	1.19E-06	2.25E-04	19
GO:0098641	cadherin binding involved in cell-cell adhesion	261	18	3.54E-06	5.00E-04	6
GO:0005524	ATP binding	261	1545	3.64E-05	0.004116	42
GO:0003723	RNA binding	261	1471	5.91E-05	0.005562	40
GO:0042802	identical protein binding	261	1714	1.73E-04	0.013988	43
GO:0051015	actin filament binding	261	226	3.13E-04	0.022086	12

10 Table 10. List of enriched KEGG pathways using differentially expressed proteins in oxaliplatin-sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
hsa01100	Metabolic pathways	131	1540	5.60E-06	0.0014	47

11 Table 11. List of enriched cellular components category using differentially expressed genes in irinotecan sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005654	nucleoplasm	1185	3963	1.96E-12	1.48E-09	325
GO:0005813	centrosome	1185	548	1.85E-09	6.96E-07	69
GO:0005829	cytosol	1185	5510	2.88E-09	7.24E-07	407
GO:0005634	nucleus	1185	5984	6.68E-09	1.26E-06	434
GO:0016020	membrane	1185	2547	9.06E-07	0.000136	203
GO:0005737	cytoplasm	1185	5576	7.62E-06	0.000821	388
GO:0016323	basolateral plasma membrane	1185	239	7.63E-06	0.000821	33
GO:0005814	centriole	1185	160	0.000131	0.012284	23

12 Table 12. List of enriched molecular functions using differentially expressed genes in irinotecan sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005515	protein binding	1170	12554	9.35E-16	1.13E-12	900
GO:0005524	ATP binding	1170	1545	3.36E-07	2.03E-04	145
GO:0004842	ubiquitin-protein transferase activity	1170	253	1.71E-05	0.006888	35
GO:0004674	protein serine/threonine kinase activity	1170	392	9.79E-05	0.029549	45
GO:0016887	ATPase activity	1170	351	1.46E-04	0.035317	41
GO:0003682	chromatin binding	1170	468	2.27E-04	0.045619	50

13 Table 13. List of enriched biological processes using differentially expressed proteins in irinotecan sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0001732	formation of cytoplasmic translation initiation complex	394	17	5.68E-10	1.26E-06	9
GO:0015031	protein transport	394	439	1.27E-06	0.001415389	27
GO:1905907	negative regulation of amyloid fibril formation	394	14	5.86E-06	0.004339027	6
GO:0006413	translational initiation	394	53	1.09E-05	0.006062305	9
GO:0006631	fatty acid metabolic process	394	95	2.36E-05	0.010490489	11
GO:0050821	protein stabilization	394	220	4.77E-05	0.015611891	16
GO:0006635	fatty acid beta-oxidation	394	48	4.92E-05	0.015611891	8

14 Table 14. List of enriched cellular components using differentially expressed proteins in irinotecan sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005829	cytosol	409	5510	7.30E-27	3.32E-24	211
GO:0070062	extracellular exosome	409	2218	1.54E-23	3.50E-21	117
GO:0005739	mitochondrion	409	1437	1.03E-19	1.56E-17	85
GO:0005737	cytoplasm	409	5576	1.51E-12	1.72E-10	177
GO:0005759	mitochondrial matrix	409	389	2.49E-12	2.27E-10	34
GO:0016020	membrane	409	2547	8.87E-12	6.72E-10	101
GO:0033290	eukaryotic 48S preinitiation complex	409	16	2.51E-10	1.43E-08	9
GO:0005852	eukaryotic translation initiation factor 3 complex	409	16	2.51E-10	1.43E-08	9
GO:0016282	eukaryotic 43S preinitiation complex	409	18	8.25E-10	4.17E-08	9
GO:0005925	focal adhesion	409	425	2.17E-09	9.87E-08	31
GO:0005654	nucleoplasm	409	3963	3.73E-08	1.54E-06	125
GO:0016272	prefoldin complex	409	7	6.10E-08	2.31E-06	6
GO:0071541	eukaryotic translation initiation factor 3 complex, eIF3m	409	7	5.10E-06	1.78E-04	5
GO:0016323	basolateral plasma membrane	409	239	6.30E-06	2.05E-04	18
GO:0015629	actin cytoskeleton	409	258	1.71E-05	5.20E-04	18

GO:0005912	adherens junction	409	175	4.93E-05	0.001401	14
GO:0005743	mitochondrial inner membrane	409	465	6.16E-05	0.001649	24
GO:0030054	cell junction	409	232	6.61E-05	0.001665	16
GO:0016324	apical plasma membrane	409	376	6.95E-05	0.001665	21
GO:0032991	macromolecular complex	409	703	7.66E-05	0.001742	31
GO:0030864	cortical actin cytoskeleton	409	60	1.80E-04	0.003891	8
GO:0005768	endosome	409	317	2.18E-04	0.004502	18
GO:0034774	secretory granule lumen	409	115	4.82E-04	0.00953	10
GO:0016328	lateral plasma membrane	409	73	6.08E-04	0.01118	8
GO:0005782	peroxisomal matrix	409	53	6.14E-04	0.01118	7
GO:1904813	ficolin-1-rich granule lumen	409	124	8.32E-04	0.014559	10
GO:0001726	ruffle	409	101	9.09E-04	0.01532	9
GO:0071818	BAT3 complex	409	3	0.001163	0.018896	3
GO:0031901	early endosome membrane	409	186	0.001266	0.019863	12
GO:0005769	early endosome	409	308	0.001338	0.020299	16
GO:0005903	brush border	409	62	0.001417	0.020791	7
GO:0030027	lamellipodium	409	192	0.001632	0.023199	12
GO:0043231	intracellular membrane-bounded organelle	409	984	0.002381	0.031943	34
GO:0005783	endoplasmic reticulum	409	1142	0.002387	0.031943	38
GO:0005884	actin filament	409	94	0.002679	0.034832	8
GO:0002102	podosome	409	30	0.002783	0.035169	5
GO:0048471	perinuclear region of cytoplasm	409	764	0.002871	0.035307	28

15 Table 15. List of enriched molecular functions using differentially expressed proteins in irinotecan sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
GO:0005515	protein binding	407	12554	8.12E-17	5.76E-14	344
GO:0045296	cadherin binding	407	316	5.50E-11	1.95E-08	30
GO:0003723	RNA binding	407	1471	1.07E-07	2.52E-05	64
GO:0042802	identical protein binding	407	1714	2.56E-07	4.54E-05	70
GO:0003779	actin binding	407	339	1.35E-06	1.92E-04	24
GO:0019904	protein domain specific binding	407	263	4.58E-06	5.41E-04	20
GO:0003857	3-hydroxyacyl-CoA dehydrogenase activity	407	7	7.02E-06	7.11E-04	5
GO:0016491	oxidoreductase activity	407	239	1.78E-05	0.001579	18
GO:0003743	translation initiation factor activity	407	64	6.62E-05	0.005215	9
GO:0005524	ATP binding	407	1545	7.89E-05	0.005595	57
GO:0051015	actin filament binding	407	226	1.20E-04	0.007722	16
GO:0051537	2 iron, 2 sulfur cluster binding	407	26	2.04E-04	0.012043	6
GO:0098641	cadherin binding involved in cell-cell adhesion	407	18	5.09E-04	0.027744	5
GO:0042803	protein homodimerization activity	407	716	8.67E-04	0.043919	30

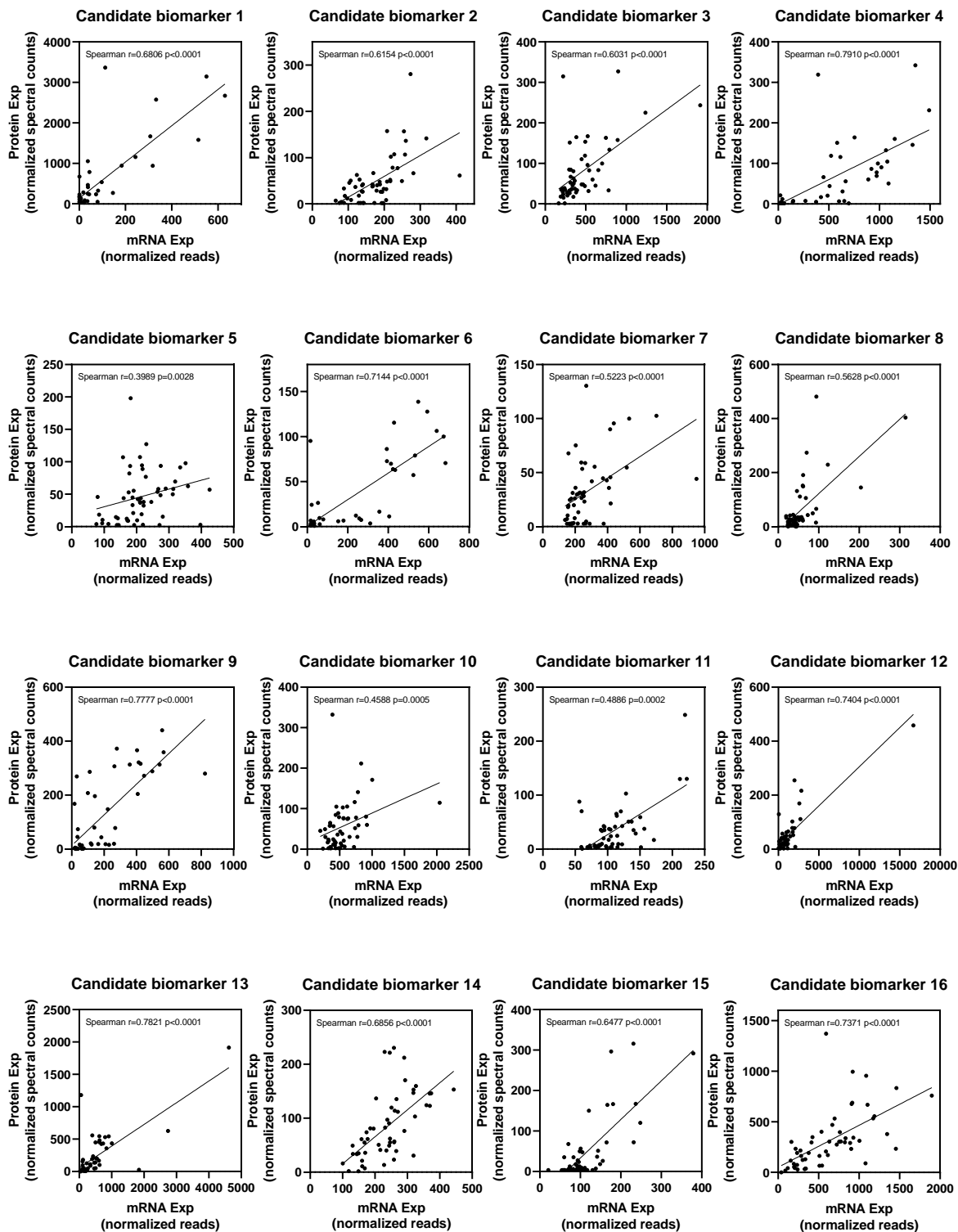
16 Table 16. List of enriched KEGG pathways using differentially expressed proteins in irinotecan sensitive and resistant cell lines

ID	Description	List Total	Pop Hits	P-value	P.adjust	Count
hsa01100	Metabolic pathways	225	1540	3.11E-09	8.36E-07	80
hsa00280	Valine, leucine and isoleucine degradation	225	48	5.19E-07	6.98E-05	11
hsa01200	Carbon metabolism	225	115	1.45E-05	0.0013	14
hsa00640	Propanoate metabolism	225	32	2.00E-05	0.001344	8

Notes: List Total: number of genes in the query list mapped to any gene set in this ontology/KEGG pathway; Pop Hits: number of genes annotated to this gene set on the background list; P-value: Fisher Exact P-value; P adjust: Benjamini-Hochberg corrected P-value.

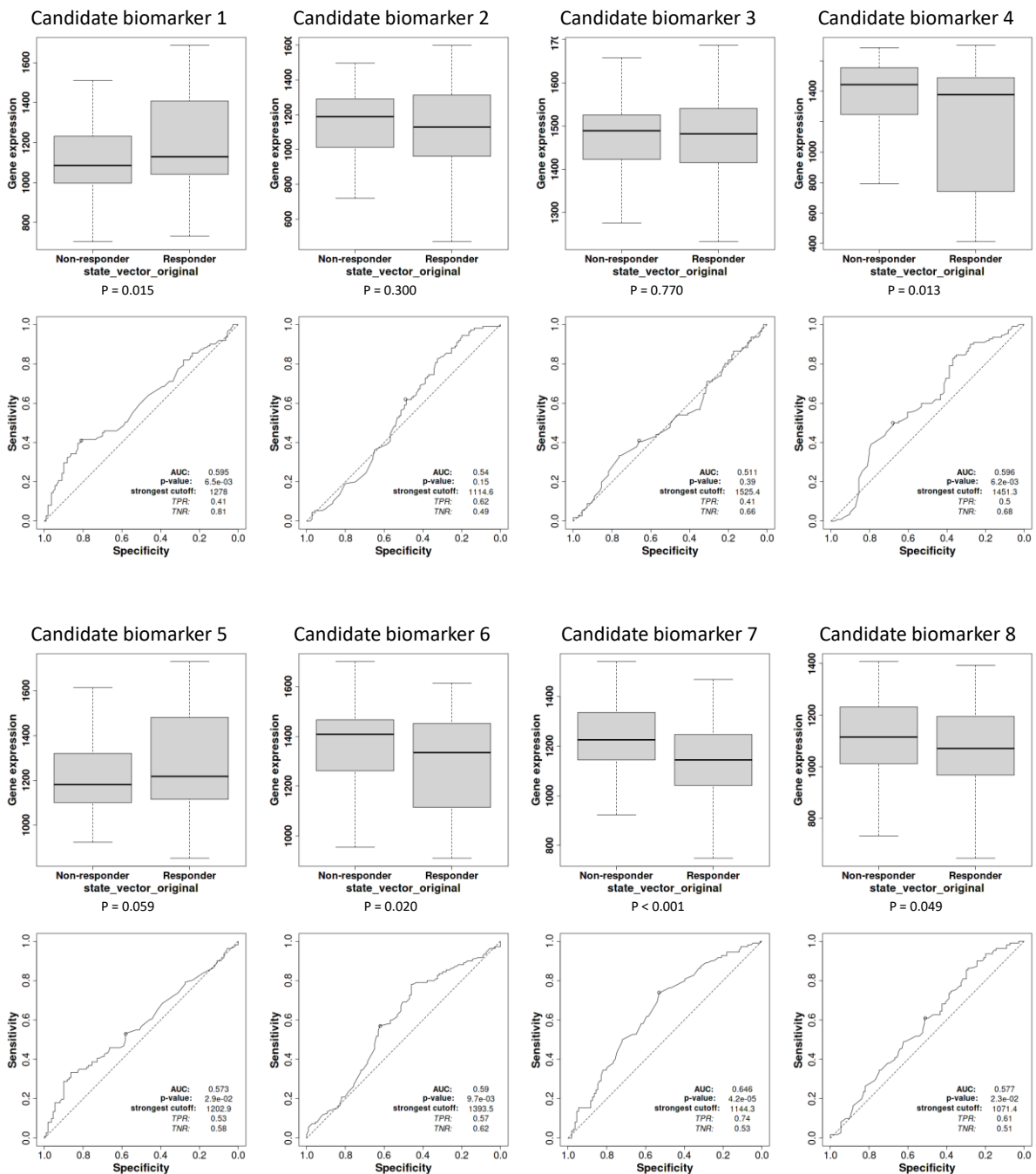
Appendix 5

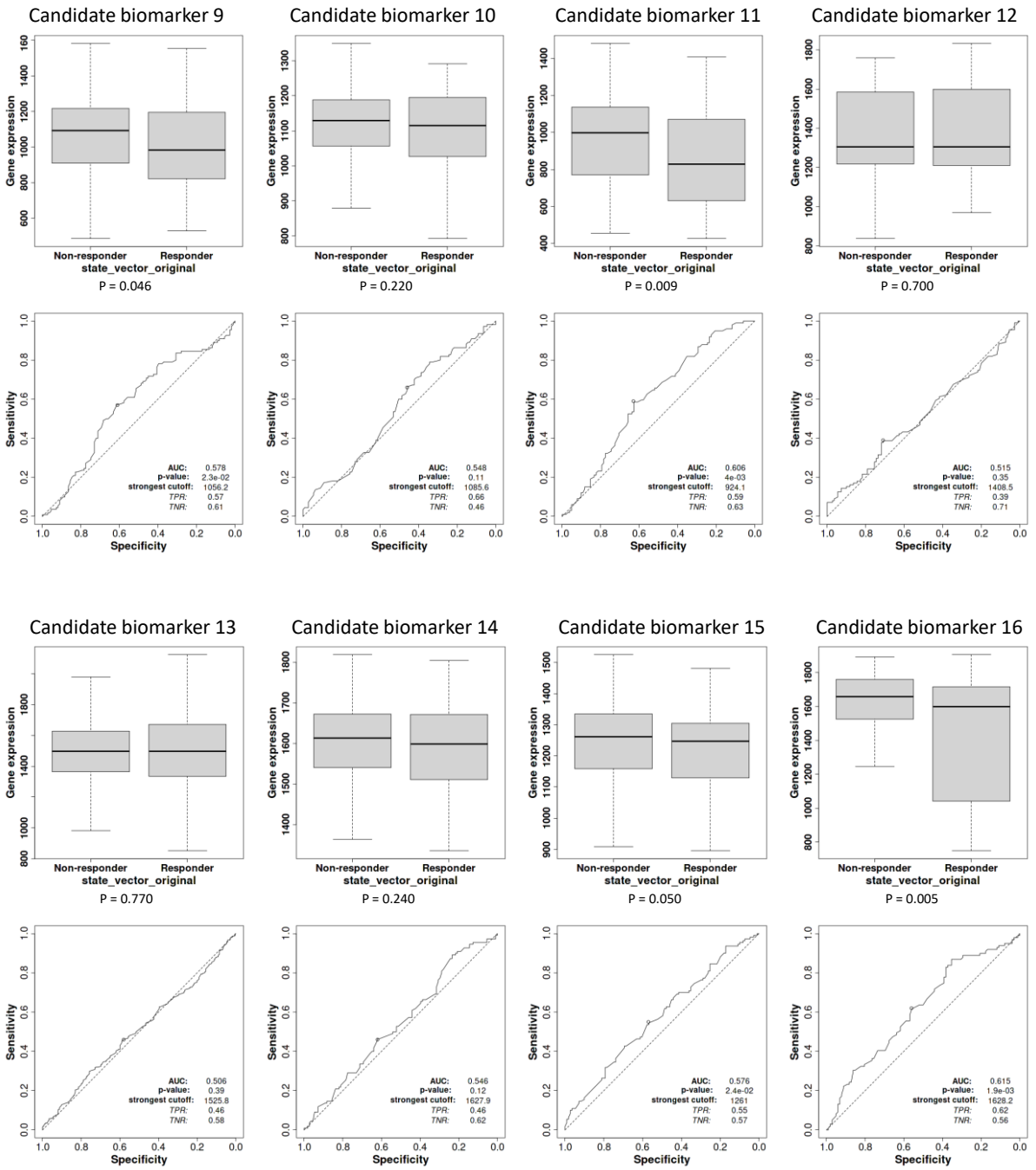
Appendix figure 5: Correlation between mRNA and protein expression of 16 potential biomarkers in a panel of colorectal cancer cell lines. Spearman's correlation coefficient was used in the comparison.



Appendix 6

Appendix figure 6: Validation of 16 potential biomarkers' predictive power in an external cohort. The mRNA expression of biomarkers in irinotecan responder and non-responder patients in an external cohort of 221 metastatic colorectal patients treated with irinotecan was shown in the upper graph. ROC curve (receiver operating characteristic curve) reveals the ability of this model to classify the patients in this cohort as responders and non-responders at all expression thresholds that can be used to dichotomize patients as responders and non-responders, which was shown in the lower graph.





Acknowledgments

I want to thank all the people who guided, helped, and supported me during the last five years. Doing a Ph.D. can always be challenging, especially in a country about 10,000 kilometers away from home. Without you, it is impossible to finish this arduous task.

First, I feel grateful to have **Dr. Diego Arango del Corro** and **Dra. Águeda Martínez Barriocanal** as my directors.

Diego, thank you for giving me the opportunity to start a new journey in Barcelona and guiding me for five years. I can always learn a lot from the meeting and conversations with you. Also, thank you for your consideration and support in my life.

Águeda, thank you for your dedicated and invaluable teaching. I feel so lucky to have you as my director. You taught me how to conduct assays and analyses as rigorously as possible and encouraged me to stay enthusiastic about science during hard times.

I also would like to thank the members of the monitoring commission **Dr. Josep Roma**, **Dr. Miguel Francisco** and **Dr. Raul Herance**, for your contribution, questions and discussions about this study. Thanks to my tutor **Dra. Ana Mesegure** and coordinator **Dr. Jordi**, for your kind support.

Estefanía Anguita, thank you for your kind and generous help. You can always bring happiness to the lab and love to help everyone when they are in trouble in the experiment. I hope you are doing well in the new position.

Priscila Guimarães, you are the first one to teach me those basic techniques like SRB assay, freezing cells, etc. Thank you for your patience and for sharing the possible difficulties and challenges in the thesis.

Rocio Nieto, **Valentina Maggio**, and **Elia Garcia**, thank you for always helping and supporting me in learning new techniques back in VHIR.

Nuria Vivancos, **Bruno Brotons**, **Mario Cazalla**, **Laia Beà**, and **Mariona Dalmau**, thank you for bringing so many laughs to the lab. Also, thank you for your kind help and suggestions inside and outside the lab.

Thank you, **Josehiginio, Lizbeth, Irati, and Fernando**, for giving me valuable advice. Also, thank you, **María Cuende, María Alfonso, and Alberto**, for your help during our time together.

Also, I want to thank the staff of the UAB biochemistry department, VHIR and IRBLleida for your help during my 5-year Ph.D. life.

最后，我还要感谢我的爸爸妈妈。在这五年期间，虽然远隔万里，你们无微不至的关怀及鼓励，批评与鞭策帮助我顺利走到了今天。感谢我最好的朋友们，**钱相君，刘平**，与你们分享生活和学习上的酸甜苦辣也是我生活中不可分割的一部分。最重要的，我要感谢我的妻子**宋扬**，你不仅是最爱的人之一，也是我生活上的伙伴，科研中的战友。无论遇到了何种困难与挑战，你都会毫无保留地帮助我，鼓励我。希望与你在接下来的人生长河里并肩战斗，共同追求美好的生活。