

Impact of 5-FU treatment on tumor cell plasticity in colorectal cancer

UOC

A scRNAseq analysis

**Universitat
Oberta
de Catalunya**



**UNIVERSITAT DE
BARCELONA**

Miquel Gratacós i Aurich

MU Bioinf. i Bioest.

Àrea de treball final

Tutor/a de TFM

Laia Bassaganyas Bars

Professor/a responsable de l'assignatura

David Merino Arranz

Data Lliurament: 18/06/2024





Aquesta obra està subjecta a una llicència de
[Reconeixement-NoComercial-](#)
[SenseObraDerivada 3.0 Espanya de Creative](#)
[Commons](#)

**Llicències alternatives (triar alguna de les següents i substituir la de la
pàgina anterior)**

A) Creative Commons:



Aquesta obra està subjecta a una llicència de
[Reconeixement-NoComercial-
SenseObraDerivada 3.0 Espanya de Creative
Commons](#)



Aquesta obra està subjecta a una llicència de
[Reconeixement-NoComercial-Compartirlqual
3.0 Espanya de Creative Commons](#)



Aquesta obra està subjecta a una llicència de
[Reconeixement-NoComercial 3.0 Espanya de
Creative Commons](#)



Aquesta obra està subjecta a una llicència de
[Reconeixement-SenseObraDerivada 3.0
Espanya de Creative Commons](#)



Aquesta obra està subjecta a una llicència de
[Reconeixement-CompartirIgual 3.0 Espanya de Creative Commons](#)



Aquesta obra està subjecta a una llicència de
[Reconeixement 3.0 Espanya de Creative Commons](#)

B) GNU Free Documentation License (GNU FDL)

Copyright © 2024 Miquel Gratacós i Aurich.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

C) Copyright

© (Miquel Gratacós i Aurich)

Reservats tots els drets. Està prohibit la reproducció total o parcial d'aquesta obra per qualsevol mitjà o procediment, compresos la impressió, la reprografia, el microfilm, el tractament informàtic o qualsevol altre sistema, així com la distribució d'exemplars mitjançant lloguer i préstec, sense l'autorització escrita de l'autor o dels límits que autoritzi la Llei de Propietat Intel·lectual.

FITXA DEL TREBALL FINAL

Títol del treball:	<i>Impact of 5-FU treatment on tumor cell plasticity in colorectal cancer</i>
Nom de l'autor:	<i>Miquel Gratacós i Aurich</i>
Nom del director/a:	<i>Laia Bassaganyas Bars</i>
Nom del PRA:	<i>David Merino Arranz</i>
Data de lliurament (mm/aaaa):	<i>06/2024</i>
Titulació o programa:	<i>Màster Universitari en Bioinformàtica i Bioestadística</i>
Àrea del Treball Final:	<i>M0.207 – TFM Anàlisis de dades Òmiques</i>
Idioma del treball:	<i>anglès</i>
Paraules clau	<i>Single-cell RNA sequencing, Colorectal Cancer (CRC), Seurat</i>

Resum del Treball

El càncer colorectal (CCR) és el tercer càncer més diagnosticat a tot el món, afectant tant homes com dones. Malgrat els avenços en les estratègies diagnòstiques i terapèutiques, aproximadament la meitat dels pacients amb CRC experimenten una recurrència dins dels cinc anys posteriors a la cirurgia i la quimioteràpia. La plasticitat cel·lular, especialment dins de les cèl·lules mare del càncer (cancer stem cells o CSCs, en angles), juga un paper crític en la tolerància al tractament, la progressió del tumor, la metàstasi i la resistència als fàrmacs, contribuint a la recaiguda del tumor.

En aquest estudi, hem utilitzat la tecnologia de seqüenciació d'ARN de cèl·lula única (scRNA-seq) per analitzar la dinàmica fenotípica de les poblacions de cèl·lules CCR en resposta a la teràpia amb 5-fluorouracil (5-FU) *in vitro*. Els

nostres resultats revelen canvis en la dinàmica del cicle cel·lular i en l'estat funcional de les cel·lules canceroses després del tractament amb 5-FU. Per una banda, vam observar que les cèl·lules resistentes semblen entrar majoritariament en un estadi d'aturada del cicle cel·lular (fase G1), fet es va revertint progressivament amb el temps . Al dia 14, el nombre de cel·lules torna a ser similar al punt d'abans del tractament i s'observa de nou un nombre significant de cel·lules en estat proliferatiu (S/G2M). Per altra banda, l'anàlisi de l'estat funcional utilitzant les signatures de CancerSEA, ens indica que les cel·lules resistentes que entren en aturada del cicle cel·lular podrien adquirir majoritariament un fenotip 'metastatic' i de reparació de dany genomic. Aquest fenotip aniria evolucionant cap a un estat d'*stemness*, destacant la resposta adaptativa del tumor. L'anàlisi de la comunicació cel·lular va mostrar interaccions actives que involucren cèl·lules metastàtiques, hipòxia, inflamació i *stemness*, reflectint un estat tumoral flexible i adaptable.

Aquest estudi subratlla la importància d'entendre l'alta plasticitat i heterogeneïtat de les cèl·lules tumorals en resposta al tractament i en l'aparició de la recaiguda. L'enfocament cap a les vies de *stemness* i metàstasi podria ser crucial per entendre la resistència i prevenir la recaiguda, millorant els resultats del tractament per als pacients amb CRC.

Abstract

Colorectal cancer (CRC) is the third most diagnosed cancer worldwide, affecting both men and women. Despite advancements in diagnostic and therapeutic strategies, approximately half of CRC patients experience a recurrence within five years post-surgery and chemotherapy. The cell plasticity, specially within cancer stem cells (CSCs), plays a key role in treatment tolerance, tumor progression, metastasis and the drug resistance, contributing to tumor relapse.

In this study, we employed single-cell RNA sequencing (scRNA-seq) technology to analyze the phenotypic dynamics of CRC cell populations in response to 5-fluorouracil (5-FU) therapy *in vitro*. Our findings reveal changes in cell cycle dynamics and functional state of cancer cells post 5-FU treatment. On one hand, we observed that the resistant cells seem to enter mostly in a

cell cycle arrest stage (G1 phase), a fact that is progressively reversed over time. On day 14, the number of cells returns to the pre-treatment point and a significant number of cells in a proliferative state (S/G2M) is again observed. Moreover, analysis of the functional status using CancerSEA signatures indicates that resistant cells entering cell cycle arrest could acquire a mostly metastatic phenotype and genomic damage repair. This phenotype would evolve into a state of stemness, highlighting the tumor's adaptive response. Analysis of cellular communication showed active interactions involving metastatic cells, hypoxia, inflammation and stemness, reflecting a flexible and adaptive tumor state.

This study underscores the importance of understanding the high plasticity and heterogeneity of tumor cells in response to treatment and in relapse apparition. Targeting stemness and metastasis pathways could prove crucial to understand resistance and prevent relapse, improving treatment outcomes for CRC patients.

Index

1.	Introduction	1
1.1.	Context and Justification of the Work.....	1
1.2.	Objectives	3
1.1.1.	General objectives	3
1.1.2.	Specific objectives	4
1.3.	Impact on sustainability, ethical-social and diversity	4
1.4.	Approach and method followed.....	5
1.5.	Planning	6
1.5.1	Tasks.....	6
1.5.2.	Callendar	8
1.6.	Brief summary of products obtained.....	9
1.7.	Brief description of the other chapters of the report	9
2.	Material and Methods.....	12
2.1.	By Time points analysis	12
2.1.1.	Quality Control and normalization of Raw matrix.	12
2.1.2.	Dimensionality reduction.....	13
2.1.3.	Cluster the cells and UMAP.	13
2.1.4.	Cell-cell interaction and LR analysis	14
2.2.	Integrative Analysis	14
2.2.1.	Perform Integration and Cluster identification	14
2.2.2.	Identify conserved cell type markers.....	15
2.2.3.	Cell-cell interaction analysis.....	15
3.	Results	16
3.1.	Analysis by time points.....	16

3.1.1.	Day 0 Cell-cell interactions	20
3.1.2.	Day 2 Cell-cell interactions	21
3.1.3.	Day 7 Cell-cell interactions	23
3.1.4.	Day 14 Cell-cell interactions.....	25
3.2.	Integrative Analysis	27
3.2.1.	Cell-cell interactions.....	30
4.	Conclusion and Discussion	33
4.1.	By Time points Analysis	33
4.1.1.	Clustering and Dynamics	33
4.1.2.	LR interactions	36
4.2.	Integrative Analysis	37
4.2.1.	Clustering and Dynamics	37
4.2.2.	LR Interactions.....	38
4.3.	Biological significance and Final conclusions.....	38
4.3.1.	Dynamics from Seurat Clusters	38
4.3.2.	Cell Cycle Regulation.....	38
4.3.3.	5-FU therapy adaptability	39
4.3.4.	Biological Implications and Future Directions	39
5.	Glossary	41
6.	Bibliografia	43
7.	Annex	47

List of Figures

Figure 1. Experimental model summarizing 5-FU induced tolerance. By analyzing cell viability, three phases have been identified: (1) the sensitive phase, characterized by slowing of cell proliferation causing the death of cells sensitive to 5-FU. An increased expression of pluripotent-related genes is also observed; (2) the tolerance phase during which cells no longer proliferate but exhibits strong transcriptional activity; and finally (3) the recurrence phase, characterized by the resumption of proliferation and mimicking tumor recurrence.....	3
Figure 2. Gantt Diagram of the followed timeline during the project.....	8
Figure 3. UMAP from scRNAseq analysis by time points. Clustering, defined by Seurat v5 Algorithm A) Represents the initial time point of the experiment; before the 5-FU treatment, 6 different clusters are identified. B) Time point 2 of the experiment: HCT116 cells after the 5-FU treatment, 6 different clusters represented. C) 7 days of experiment with 6 different clusters. D) Day 14 of the experiment with 5 cell populations.....	16
Figure 4 Proportion Plot of the cells proportion at each time point based on the Seurat clustering.....	Error! Marcador no definido.
Figure 5. UMAP from scRNAseq analysis by time points. Clustering defined by Cell Cycle Ranked, G2M, G1 and S phase are labeled. A) Initial time point of the experiment, before the 5-FU treatment. B) Time point 2 of the experiment, HCT116 cells after the 5-FU treatment. C) 7 days of experiment. D) Day 14 of the experiment.....	17
Figure 6. Proportion Plot of the cells proportion at each time point based on Cell Cycle scores.....	18
Figure 7. UMAP from scRNAseq analysis by time points. Clustering is defined by the CancerSEA signatures. A) Initial time point of the experiment, before the 5-FU treatment. B) Time point 2 of the experiment: HCT116 cells after the 5-FU treatment.....	19
Figure 8. Proportion Plot of the cell populations at each time point based on the signatures dataset, CancerSEA.....	19

Figure 9. Heatmap of Cell-cell interactions at day 0 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.	20
Figure 10. Most relevant cell-cell interactions for Day 0 on CancerSEA signatures.....	21
Figure 11. Heatmap of Cell-cell interactions at day 2 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.	22
Figure 12. Most relevant cell-cell interactions for Day 2 on CancerSEA signatures.....	23
Figure 13. Heatmap of Cell-cell interactions at day 7 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.	23
Figure 14. Most relevant cell-cell interactions for Day 7 on CancerSEA signatures.....	25
Figure 15. Heatmap of Cell-cell interactions at day 14 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.	25
Figure 16. Heatmap of Cell-cell interactions at day 14 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.	26
Figure 17. UMAP merging all Time points, representing the unintegrated analysis. Each timepoint is represented in different colors as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).	27
Figure 18. UMAP merging all data times matrices without integrating methodology, represented split by days (J0, J2, J7 and J17), a total of 9 different Clusters are identified.....	28
Figure 19. UMAP merging all Time points, representing the integrated analysis using Harmony as Integration methodology. Each timepoint is represented in different colors as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).....	29
Figure 20. UMAP merging all data times matrices using Harmony as Integration methodology, represented split by days (J0, J2, J7 and J17), a total of 8 different Clusters are identified.....	29

Figure 21. <i>UMAP merging all data times matrices using Harmony as Integration methodology, represented split by days (J0, J2, J7 and J17), the CancerSEA Cancer Cell signatures were represented.</i>	30
Figure 22. <i>Heatmap of Cell-cell interactions for the integrated model between CancerSEA signatures, interaction frequency are represented by the intensity of the color.</i>	31
Figure 23. <i>Most relevant cell-cell interactions for the integrative analysis on CancerSEA signatures.</i>	32

List of Tables

Table 1. Deadlines and duration of the project planification.	8
Table 2. <i>Top 10 cell-cell interactions for Day 0 CancerSEA signatures.</i>	20
Table 3. <i>Top 10 cell-cell interactions for Day 2 CancerSEA signatures.</i>	22
Table 4. <i>Top 10 cell-cell interactions for Day 7 CancerSEA signatures.</i>	24
Table 5. <i>Top 10 cell-cell interactions for Day 14 CancerSEA signatures.</i>	26
Table 6. <i>Integrative analysis top 10 cell-cell interactions of CancerSEA signatures.</i>	31

1. Introduction

1.1. Context and Justification of the Work

Colorectal cancer (CRC) is the third most commonly diagnosed cancer, affecting both men and women worldwide. Despite the advancements in diagnostic and therapeutic strategies, approximately half of CRC patients experience recurrence within 5 years post-surgery and chemotherapy¹⁻³.

Transcriptomic and epigenetic studies have highlighted cellular plasticity as a hallmark of treatment tolerance^{4,5}. Cancer cells are heterogeneous in morphology and function and within this diverse populations, cancer stem cells (CSCs) represent a small population of cells capable of self-renewal and tumor regeneration². CSC are increasingly recognized for their role in driving tumor plasticity, which is crucial for cancer progression and metastasis, and in acquiring drug resistance⁶, ultimately contributing to tumor relapse. Despite significant recent advances, many questions remain about how tumors adapt to therapeutic regimens. Addressing these questions is essential to improve outcomes for cancer patients, including those with CRC².

In recent years, single-cell RNA sequencing (scRNA-seq) technology has revolutionized the field by providing a powerful technique for defining molecular signatures of cell types and subtypes, as well as for data analysis, visualization, and mining of scRNA-seq datasets, exemplified by software such as Seurat^{7,8}. This technology allows the measurement of the expression levels of all genes across thousands to millions of individual cells and has become crucial for understanding transcriptome and cell population dynamics⁹. Specifically, scRNA-seq permits whole-transcriptome profiling of individual cells, which supports identification of rare cell subpopulations, inference of complex gene-gene regulatory factors, and tracing of developmental lineages. Additionally, scRNA-seq facilitates the inference of cell-cell communication (CCC) by measuring the expression of genes encoding corresponding ligands, receptors, intermediate signaling proteins, and intracellular targets across interacting cell types and under homeostatic and diseased conditions¹⁰.

As many cancer types, CRC tumors are complex heterocellular systems comprising mutated cancer cells, stromal fibroblasts, and multiple immune cells within the tumor microenvironment (TME). Colonic cancer stem cells (CSCs) have the capacity to replenish the murine colonic epithelium in 3-5 days. Proliferative CSCs (proCSCs) are transcriptionally similar to healthy CSCs, but unlike healthy CSCs, oncogenic proCSCs are trapped in a mitotic state and fail to differentiate into the secretory and absorptive cells of the healthy colon^{10,11}.

Despite the importance of somatic mutations in driving CRC, evidence of nongenetic plasticity has recently been shown to underpin metastasis, immune evasion, and therapy resistance across multiple cancer types¹². Additionally, a new type of stem cells that helps resist treatment has been described, known as revival stem cells (revCSCs). During CRC progression, these cells polarize between proCSC and revCSC phenotypes to establish primary tumors. While revCSCs are crucial for early tumor formation, they cannot expand due to their slow cell-cycle activity. CRC manages the complementary phenotypes of proCSCs and revCSCs to establish and maintain primary tumors¹¹.

The study presented here is part of an on-going collaborative project developed between the Institut de Génomique Fonctionnelle (IGF) of Montpellier and the Lyon Cancer Research Center (CRCL) that seeks to improve our understanding of the role of tumor plasticity in response to chemotherapy in CRC. Preliminary *in-vitro* data have shown that 5-fluorouracil (5-FU)-based chemotherapy, the first-line of treatment for CRC, induces increased translational activity during the “tolerance phase”, when cells stop dividing, followed by a significant resumption of cell division (Figure 1). In addition, this increased protein synthesis would correlate with the emergence of several tumor subpopulations, including cells exhibiting characteristics of pluripotent cells (i.e., CSC).

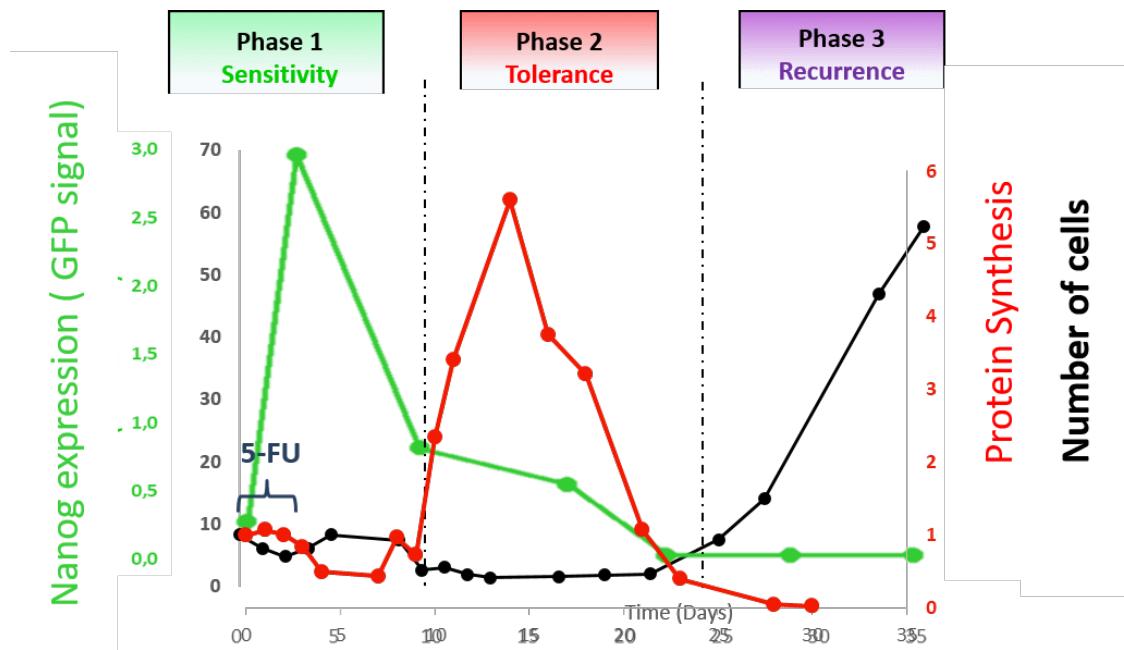


Figure 1. Experimental model summarizing 5-FU induced tolerance. By analyzing cell viability, three phases have been identified: (1) the **sensitive phase**, characterized by slowing of cell proliferation causing the death of cells sensitive to 5-FU. An increased expression of pluripotent-related genes is also observed; (2) the **tolerance phase** during which cells no longer proliferate but exhibits strong transcriptional activity; and finally (3) the **recurrence phase**, characterized by the resumption of proliferation and mimicking tumor recurrence.

In light of these data, the project progressed to the next step, aiming to demonstrate that translational control is a key molecular mechanism in cellular reprogramming in response to 5-FU. To this end, scRNA-seq was conducted at various time points using the *in vitro* experimental model developed by the team to study the effect of this drug in cultured CRC cells. The goal is to go more in depth the phenotypic dynamics of cancer cell populations in response to treatment and to elucidate the potential intercellular communication networks between the different subpopulations of cells.

1.2. Objectives

1.1.1. General objectives

As mentioned above, this study is part of a project from the IGF and CRCL, which aims to achieve several key objectives:

- 1. Characterize relevant cancer cell subpopulations appearing in response to 5-FU.***
- 2. Establish CCC networks between cell subpopulations.***
3. Demonstrate that translational control is one of the molecular mechanisms responsible for cellular reprogramming in response to 5-FU.
4. Determine the exact role of these subpopulations in treatment tolerance and tumor recurrence.

* Specifically, this study will focus on achieving objectives 1 and 2.

1.1.2. Specific objectives

Based on the primary objectives, we have established a detailed scheme to create the workflow for this master's thesis project:

1. Employ and optimize scRNA-seq workflow using Seurat Package in R to identify relevant cellular subpopulations.
2. Perform detailed analysis of these cell subpopulations
 - 2.1. Compare integrative analysis with non-integrative analysis.
 - 2.2. Phenotypically characterize the cellular subpopulations
 - 2.3. Identify relevant cell markers.
3. Establish CCC networks between the defined subpopulations
 - 3.1. Analyze ligand-receptor (LR) interactions within the different clusters obtained.

1.3. Impact on sustainability, ethical-social and diversity

Here, I describe the points related to the guide of sustainability, ethical-social and diversity of UOC¹³, that this project considers:

(1) Sustainability:

a. Economic:

- i. Cost savings: Effective treatments lower overall healthcare costs and reduce hospitalization costs.

ii. Long-term Savings: Early interventions provide long-term economic benefits for healthcare systems.

b. Social:

i. Improved Quality of Life: Effective management enhances patients' quality of life and social contributions.

(2) Ethical-Social Considerations:

a. Ethical Research:

i. Transparency: Maintaining clear, reproducible research methods.

(3) Embracing Diversity:

a. Equity:

i. Reducing Disparities: Developing treatments effective for all populations.

b. Workforce Diversity:

i. Collaboration: Encouraging interdisciplinary and diverse team collaborations.

1.4. Approach and method followed

As mentioned above, this project was conducted in collaboration between the IGF and CRCL to identify and phenotypically characterize relevant cellular subpopulations in a longitudinal experiment of 5-FU therapy. In this context, an experiment was performed on a CRC cancer cell line (HCT116), which was treated with 5-FU, and the evolution of cell populations was observed over 20 days. Single-cell RNA from the colonies was extracted and sequenced at different time points (days 0, 2, 7, and 14). Treatment was administrated just before day 2. This approach allows us to obtain a map of the evolution of different cell populations and proportions in response to 5-FU treatment.

The analysis by time points allows us to identify the specific cellular composition at each moment (0, 2, 7 and 14), providing an idea of different relevant subpopulations, cell markers and specific interactions without losing potentially

important biological variability. This analysis may also enable a better comparison with the preliminary results mentioned above.

On the other hand, integrative analysis, although losing some resolution, enhances the visualization of global changes. This integrative approach forces the data to find similarities between cells and establishes clusters by combining all the data, allowing for a better visualization of cell population dynamics.

Moreover, for each analysis, we performed three different clustering methodologies; (1) clustering based on Seurat algorithms; (2) clustering based on cell cycle status; and (3) clustering based on the 14 cancer-related functional states from CancerSEA signatures using SCINA^{14,15}.

After examining the different clustering results, we selected the best option to decipher relevant CCC players between cancer cell subpopulations. To do so, we performed an analysis of LR using the database for intra- and intercellular signaling knowledge, OmniPath¹⁶, employing the LIANA package¹⁷.

Finally, once the significant markers and cell-cell interactions were addressed, a biological analysis based on the results of the study's results was performed to assess the mechanisms behind the cell population changes throughout the 5-FU treatment.

1.5. Planning

1.5.1 Tasks

The tasks developed during this project are described in Figure 2, resulting in the following schedule:

Literature Review and Preliminary Analysis:

Duration: 2 weeks

Task: Conduct a comprehensive review of literature related to translational control in CRC and perform preliminary analysis of existing scRNASeq data.

Workflow Optimization:

Duration: 3 weeks

Task: Optimize the scRNAseq workflow using the Seurat package from R, including parameter tuning and integration of multiple datasets.

Quality Control:

Duration: 2 weeks

Task:

- Data Preprocessing: Import and initially check raw scRNA-seq data for quality.
- Filtering: Remove low-quality cells based on UMI counts, gene detection, and mitochondrial expression.
- Feature Selection: Identify highly variable genes for analysis.
- Dimensionality Reduction: Use PCA for reducing data complexity.
- Quality Check Visualization: Generate UMAP/t-SNE plots to visualize data quality.

Cluster markers analysis and Phenotypic Characterization

Duration: 5 weeks

Task:

- Identification of cell subpopulations and cluster markers.
- Final identification and characterization of Seurat Clusters. Phenotypically characterize cellular subpopulations through Cancer-Cell Markers.

CCC Analysis

Duration: 2 weeks

Task: Studying CCC and LR interaction of the different cell types or clusters using LIANA.

Biological description:

Duration: 2 weeks

Task: Evaluate the contribution of the identified cellular subpopulations to 5-FU treatment tolerance and tumor recurrence.

Preparation of the Manuscript:

Duration: 4 weeks.

Task: Writing and developing the conclusions and discussion of the project.

1.5.2. Callendar

The time schedule for this project was defined to achieve all the objectives of the study.

Table 1. Deadlines and duration of the project planification.

Task	Start date	Duration (days)	End date
Literature Review and Preliminary Analysis	29/02/2024	14	14/03/2024
PAC01 - Definition and Work Plan	12/03/2024	0	12/03/2024
Workflow Optimization	15/03/2024	21	05/04/2024
Quality Control	22/03/2024	21	12/04/2024
PAC02 - Development of the work Phase 1	09/04/2024	1	09/04/2024
Cluster markers analysis	01/04/2024	35	13/05/2024
CCC Analysis	22/04/2024	14	06/05/2024
Biological description	06/05/2024	14	20/05/2024
PEC03 - Development of the work Phase 2	14/05/2024	1	14/05/2024
Preparation of the Manuscript	14/05/2024	32	15/06/2024

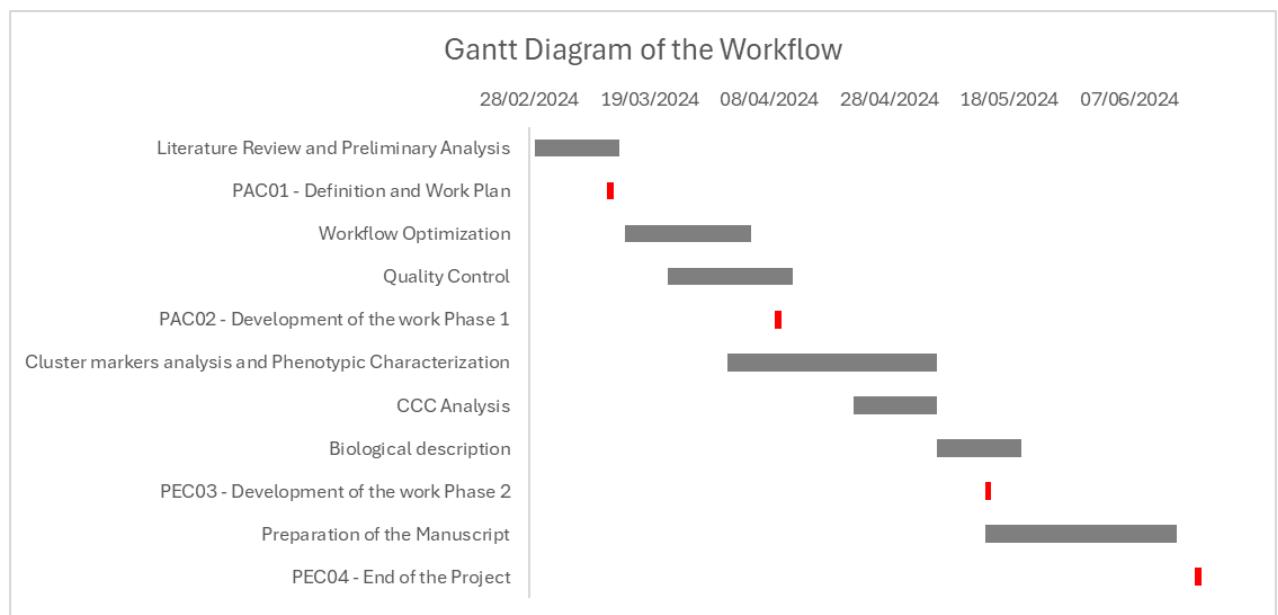


Figure 2. Gantt Diagram of the followed timeline during the project.

1.6. Brief summary of products obtained.

- UMAP for each time point (0, 2, 7 and 14), and table of the most significant markers, based on p-adj and Log2 FC values, from each time point, on 2 different clustering methodologies:
 - o Clustering by Seurat v5 Algorithm.
 - Proportion plot based on Seurat Clusters.
 - o Clustering with labels based on CancerSEA database of Cancer Cell Markers, thanks to the Cell Cycle identification of each cell phase (G2M, G1 and S).
 - Proportion plot based on this Cell Markers, for the by time points analysis.
- Integrative analysis of the experiment.
 - o UMAP and table of the most significant markers.
- R-scripts for Integration and time point analysis for Seurat v5 scRNA-seq analysis.
- CCC Analysis for the by time points analysis and Integrative:
 - o LR plots for each time point for the Seurat and CancerSEA clustering, showing interactions between clusters.
 - o LR Table of the top 50 most concordant interactions for each Time point and Integrative analysis, and dotplot representing its interactions.

1.7. Brief description of the other chapters of the report

2. Material and methods: This chapter provides an in-depth explanation of the analysis performed using R-studio and the methodology followed. It is subdivided into the following categories:
 - 2.1. By time points analysis: This part describes the analysis of data from different time points.
 - 2.1.1. Quality Control and normalization of Raw matrix: This section details the initial analysis of the scRNA-seq data, including quality control measures.

- 2.1.2. Dimensionality reduction: This section explains the process of reducing the dimensionality of the scRNA-seq data for analysis.
 - 2.1.3. Cluster the cells and UMAP: After dimensionality reduction, the clusters of each matrix are identified, and their potential cell markers are described.
 - 2.1.4. Cell-cell interaction analysis: This part focuses on analyzing LR interactions between cells in the defined clusters:
- 2.2. Integrative analysis: Following the time points analysis, this section describes the methodology for integrating all data points into a single data matrix.
- 2.2.1. Perform Integration and Cluster identification
 - 2.2.2. Identify conserved cell type markers
 - 2.2.3. Cell-cell interaction analysis
3. Results: In this chapter, the results from the methodologies described in the previous chapter, divided into the following subsections:
 - 3.1. Results by time points:
 - 3.1.1. Day 0 cell-cell interactions
 - 3.1.2. Day 2 cell-cell interactions
 - 3.1.3. Day 7 cell-cell interactions
 - 3.1.4. Day 14 cell-cell interactions
 - 3.2. Results integrative analysis:
 - 3.2.1. Cell-cell interactions
4. Conclusion and discussion: This chapter explains the biological implications of the results and provides the final conclusions of the project.
- 4.1. By Time points Analysis:
 - 4.1.1. Clustering and Dynamics: Discusses the clustering of cells over different time points.

4.1.2. LR Interactions: Investigates ligand-receptor interactions activated at each time-point..

4.2. Integrative Analysis:

4.2.1. Clustering and Dynamics: Describes the clustering in the integrated analysis.

4.2.2. LR Interactions: Investigates ligand-receptor interactions activated in the integrated dataset.

4.3. Biological significance and Final conclusions

4.3.1. Dynamics from Seurat Clusters

4.3.2. Cell Cycle Regulation

4.3.3. 5-FU therapy adaptability

4.3.4. Biological Implications and Future Directions

2. Material and Methods

As mentioned in the previous sections, the experiment consists of a longitudinal scRNA-seq cancer cell lines (HCT116), from different time points: 0, 2, 7, and day 14. Treatment was administrated on day 2. At each time point, RNA was extracted and sequenced at the single-cell level using the Chromium technology (10X Genomics, Pleasanton, CA, USA). Alignment and gene expression quantification were performed using the Cell Ranger single cell software suite from 10X Genomics. Gene count matrices were used as the starting point for this study. Data processing and visualization were performed using the Seurat v5 package in R. Additional packages, such as ggplot2, SCINA and LIANA, were also employed.

As mentioned above, two different approaches were taken: one was to analyze the datasets separately by time points to observe specific functional states of cancer cell populations at each time-point. The other approach was an integrative analysis, combining all datasets into a single dataset to observe the evolutionary dynamics of the functional states over time.

Based on these methodologies, we developed and optimized 2 R scripts for each analysis, described in this section.

2.1. By Time points analysis

For each matrix, we created a Seurat object, which stores all the raw information about the experiment, such as metadata, assay, counts matrix, and project name, with a set of designed parameters.

2.1.1. Quality Control and normalization of Raw matrix.

First, we explored and visualized several QC metrics, such as the amounts of mitochondrial and ribosomal genes, and we defined the thresholds to filter out those cells and genes that could negatively impact the downstream analysis. (Annex).

Once the undesired cells and genes were removed, normalization and scaling of each matrix were performed. The normalization method used was “LogNormalize”, which normalizes feature expression measurements from each

cell by the total expression, multiplies by a scale factor, and log-transforms the result. For future downstream analysis, the 5000 most variable features were selected.

Next, we performed a linear transformation-based scaling to reduce dimensionality, and then re-scaled the data by cell cycle difference (S/G2M vs G1).

2.1.2. Dimensionality reduction.

Next, PCA was performed on the scaled data. Determining the true dimensionality of a dataset can be challenging, so multiple approaches were used. An Elbow plot and a Dimensional Reduction heatmap were generated to determine the dataset's dimensionality.

2.1.3. Cluster the cells and UMAP.

Once the dimensions are correctly selected, clusters were calculated. The identified clusters were measured using a shared nearest neighbor (SNN) modularity optimization-based clustering¹⁸. For visualization, we applied UMAP, a non-linear dimensional reduction technique, to the datasets. UMAP places similar cells together in low-dimensional space. Although UMAP is valuable for exploring datasets, it has limitations, so we avoided drawing biological conclusions solely based on visualization techniques¹⁹.

For each cluster and time point (day 0, 2, 7, and 14), differentially expressed genes (adj p-value < 0.05 & avg log2FC > 1) were identified and saved in a table for further cell type identification based on their marker function. This allowed us to add Differential Gene Expression (DGE) markers for each Seurat cluster and predict their cancer signatures.

Finally, we used the SCINA package, which employs an automatic cell type detection and assignment algorithm for scRNA-seq. SCINA is capable of assigning cell type identities to a pool of cells profiled by scRNA-seq data with prior knowledge of signatures, such as CancerSEA signatures database, a cancer single-cell functional state atlas, involving 14 functional states (stemness, invasion, metastasis, proliferation, EMT, angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, hypoxia, inflammation and

quiescence) of 41900 cancer single cells from 25 cancer types¹⁵. Based on SCINA results, we created a proportion plot representing the phenotypic dynamics of cancer cells after treatment and over time.

2.1.4. Cell-cell interaction and LR analysis

To study CCC, we used the LIANA framework²⁰, which generates a consensus prediction of activated ligand-receptors after applying various existing methods. For the by-time point analysis, we performed CCC analysis specifically on CancerSEA signatures, and interaction between clusters were recognized. We identified the top 50 most relevant interactions by LIANA's consensus rank aggregate of magnitude LR scores, which will be useful for further studies.

2.2. Integrative Analysis

Integration of single-cell datasets across experimental batches, donors, or conditions may be useful in some scRNA-seq workflows. Integrative analysis helps match shared cell types and states across datasets, enhancing statistical power and enabling accurate comparative analysis. In this project, the different time points (day 0, 2, 7, and 14) represent the experimental batches.

2.2.1. Perform Integration and Cluster identification

Before starting the integrative analysis in Seurat, we merged the rescaled Seurat objects from each time points into a single object, ensuring that Quality Control and removal of low-quality cells were already performed. Seurat v5 allows storing data in layers, which can hold raw, unnormalized counts, normalized data, or z-scored/variance-stabilized data. Our data included 12 layers (a counts layer and a scale data layer of each time point).

Once data was uploaded, normalization and scaling of the count's matrix were performed. We visualized the results of a standard analysis without integration, defining clusters by cell types defined by Seurat algorithm.

Our goal was to integrate data from the different conditions, so cells of the same type/subpopulations would cluster together. The integration method used was Harmony, a well-established technique that projects cells into a shared embedding in which cells group by cell type rather than dataset-specific

conditions²¹. Harmony simultaneously accounts for multiple experimental and biological factors.

After integrating the scRNA-seq data using the Harmony method, we identified clusters in the integrated dataset and generated UMAPs and DimPlots, where the dimensionality of the dataset was determined using elbow plots. Cell clustering was performed using the FindNeighbours and FindCluster functions in Seurat²².

2.2.2. Identify conserved cell type markers

To identify canonical cell type markers, we used the FindAllMarkers() function to find those that differ between identified clusters (differentially expressed genes). We set a log fold change threshold of 0.15 and specified for only positive ones (up-regulated genes). Additionally, we used the min.pct = 0.2 option to exclude genes that are infrequently expressed and set an adjusted p-value threshold of 0.05 to select the most significant cell markers. The test used was the Wilcoxon rank sum test.

Additionally, we identified CancerSEA signatures for the integrative analysis and created UMAPs.

2.2.3. Cell-cell interaction analysis

As with the by-time points study, we performed cell-cell interaction analysis on the integrative dataset. We identified the most concordant interactions on the clustering for CancerSEA signatures, based on a consensus that integrates the predictions of individual methods. We identified the top 50 most relevant interactions by LIANA's consensus rank aggregate of magnitude LR scores, which will be useful for further studies.

3. Results

3.1. Analysis by time points

Based on the different time points (day 0, 2, 7 and 14), we observed a varying number of clusters for the different time points. At days 0, 2 and 7 6 clusters were identified (Figure 3). The proportion of the different clusters remained similar over time; however, by day 14, cluster 6 was no longer present. On day 0, clusters 0 and 1 were the most abundant. On day 2, cells were predominantly found in cluster 0. By day 7, there was a reduction in the percentage of cells in cluster 2, which returned to its initial proportion by day 14, like day 0. Additionally, cluster 3 was highly abundant on day 14 (Figure 4).

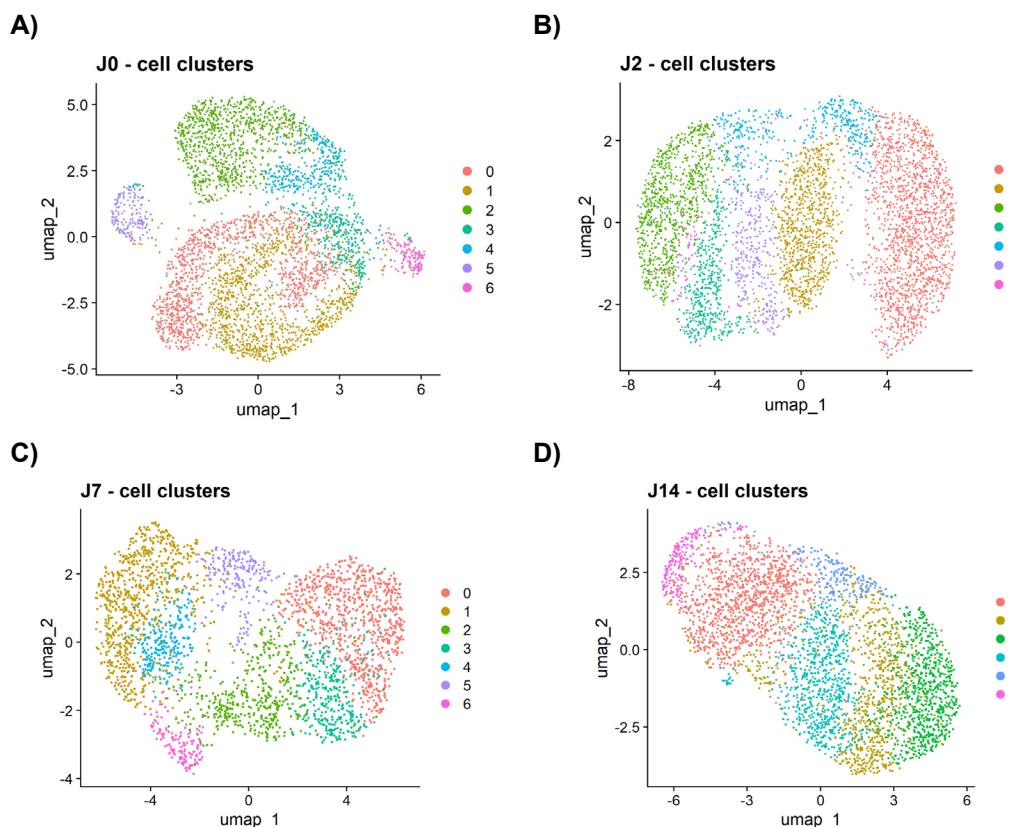


Figure 3. UMAP from scRNAseq analysis by time points. Clustering, defined by Seurat v5 Algorithm **A)** Represents the initial time point of the experiment; before the 5-FU treatment, 6 different clusters are identified. **B)** Time point 2 of the experiment: HCT116 cells after the 5-FU treatment, 6 different clusters represented. **C)** 7 days of experiment with 6 different clusters. **D)** Day 14 of the experiment with 5 cell populations.

The top 50 most relevant cell markers of each Seurat Cluster for all time points are distinguished on a table (Annex)

On the other hand, cell cycle scoring analysis showed that on day 0, most cells, although in different cell cycling phases, are spatially represented in a similar 2D (Figure 5). In contrast, for the other time points, especially day 2, the spatial difference between cell cycles is clearly represented.

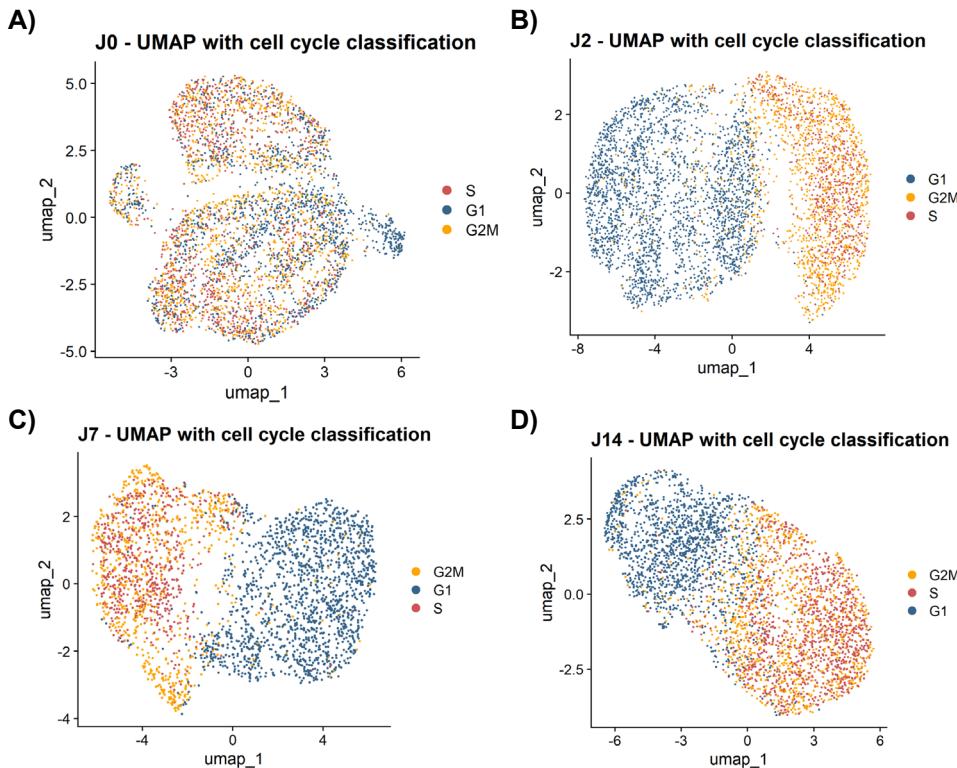


Figure 4. UMAP from scRNAseq analysis by time points. Clustering defined by Cell Cycle Ranked, G2M, G1 and S phase are labeled. **A)** Initial time point of the experiment, before the 5-FU treatment. **B)** Time point 2 of the experiment, HCT116 cells after the 5-FU treatment. **C)** 7 days of experiment. **D)** Day 14 of the experiment.

In terms of cell proportion, on day 0, the most abundant phases are G2M and G1, with difference from the S phase being relatively small (25% approximately). On day 2, however, the S phase is clearly decreased, and most cells are in the G1 phase. This difference in cell proportion between the S and G1 phases recovers by day 14, where the number of S phase and G1 phase cells is similar (approximately 30%) (Figure 6).

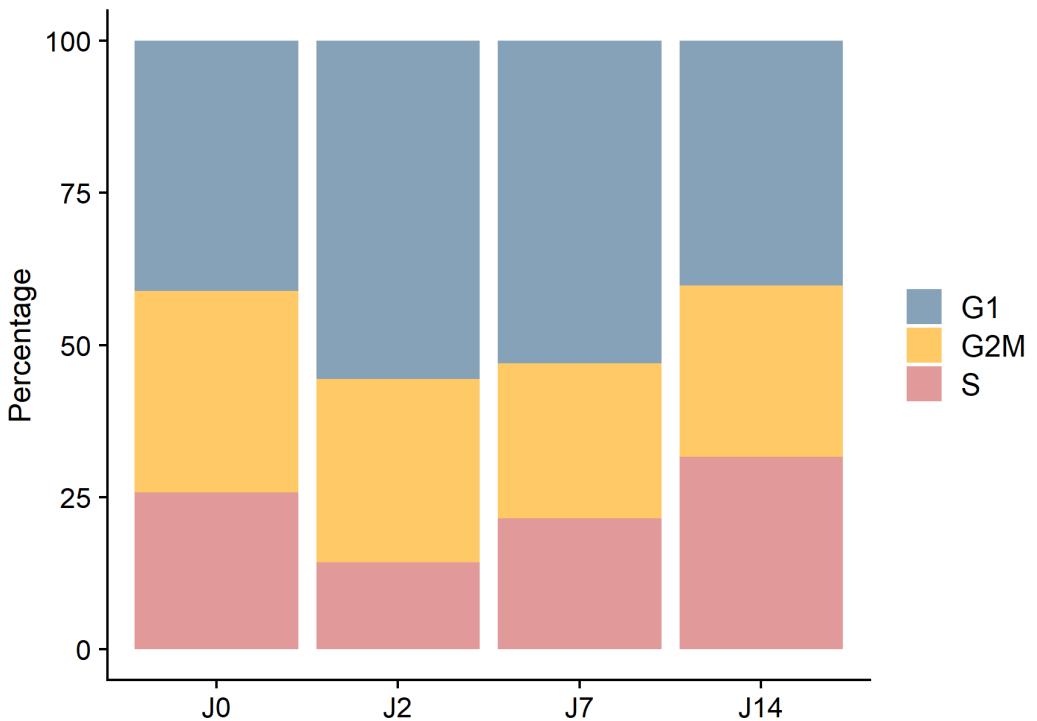


Figure 5. Proportion Plot of the cells proportion at each time point based on Cell Cycle scores.

When annotating by the CancerSEA signatures, we observed that metastasis cells were less abundant on days 0 and 14 compared to days 2 and 7. Cells in stemness state were not present on days 0 and 14 compared to days 2 and 7. Cells in stemness state were not present on days 0 and 2, but their numbers increased on days 7 and 14, while proliferative cells decreased over time. Additionally, apoptotic cells were only present on days 0 and 2. The number of cells experiencing DNA damage was reduced on day 7. Furthermore, cells involved in DNA repair were not present on day 14 (Figure 7 and 8).

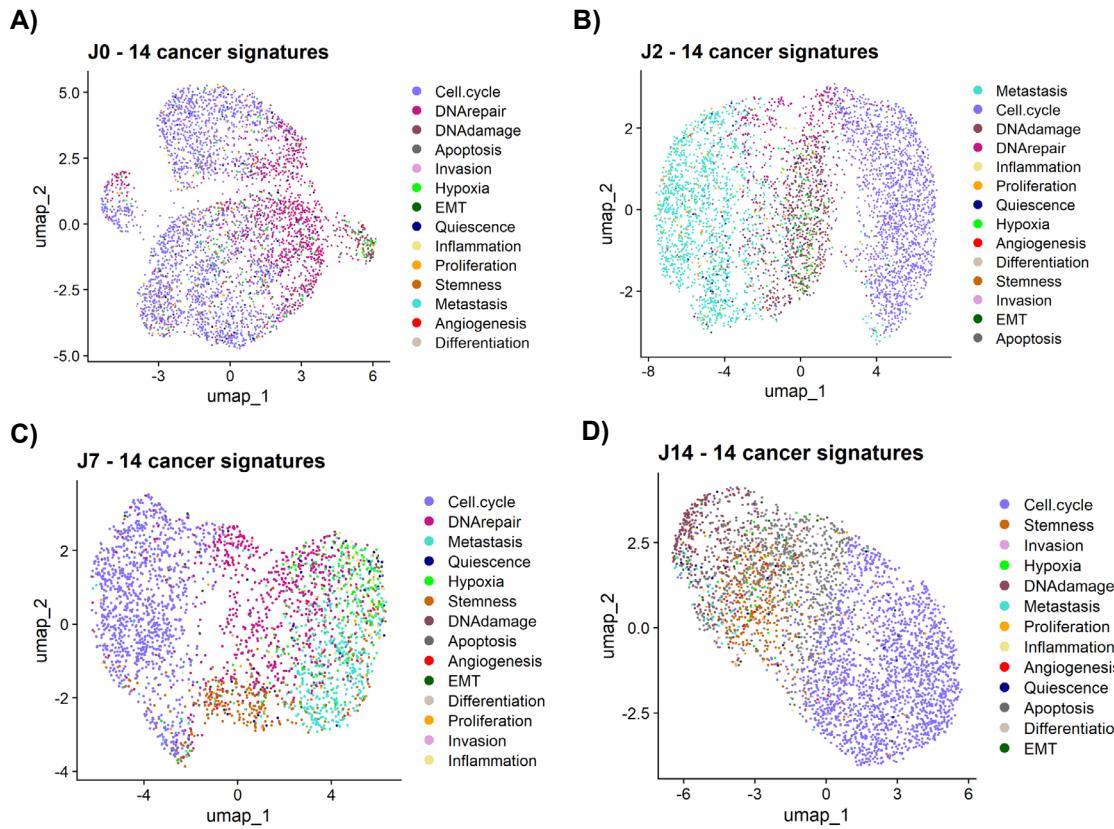


Figure 6. UMAP from scRNAseq analysis by time points. Clustering is defined by the CancerSEA signatures. A) Initial time point of the experiment, before the 5-FU treatment. B) Time point 2 of the experiment: HCT116 cells after the 5-FU treatment.

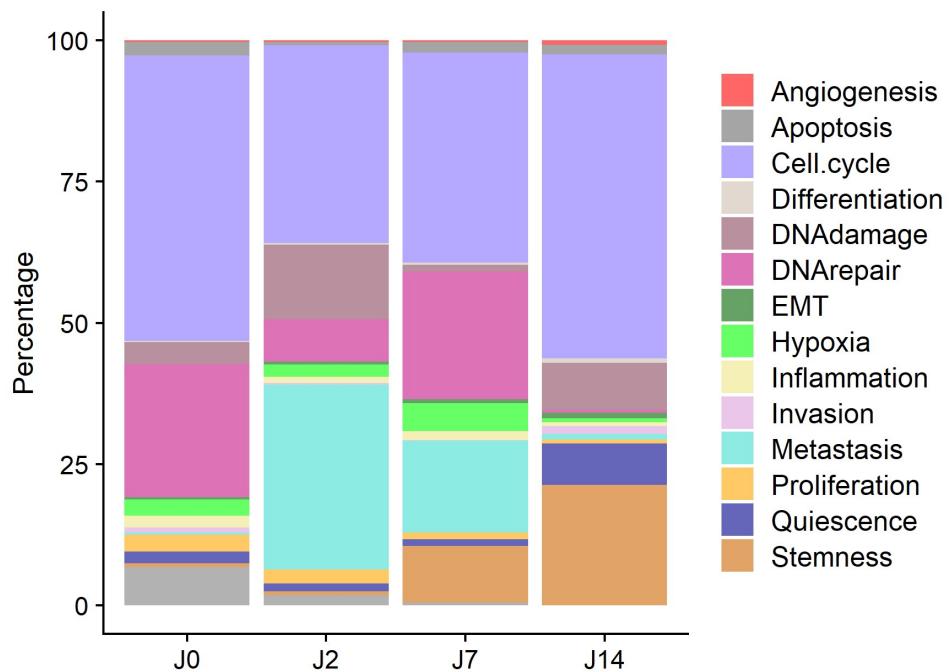


Figure 7. Proportion Plot of the cell populations at each time point based on the signatures dataset, CancerSEA.

3.1.1. Day 0 Cell-cell interactions

For the CancerSEA data, the most frequency interactions take place between cells from DNAdamage and Hypoxia to Cell Cycle cells, followed by proliferation cluster interactions. Moreover, most of cells from DNArepair cluster doesn't interact and have low abundance (Figure 9).

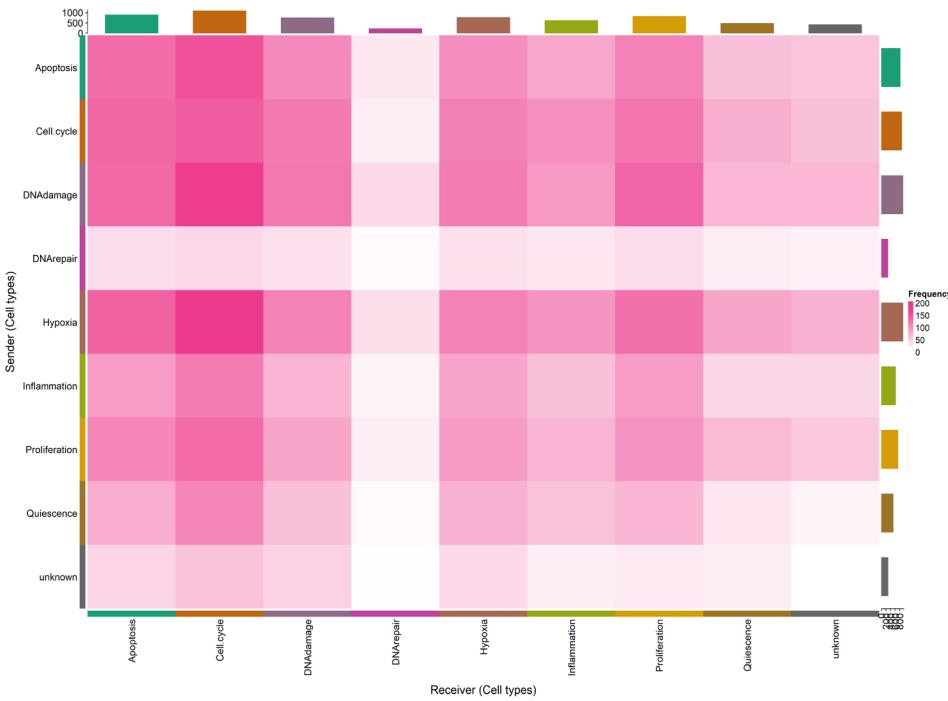


Figure 8. Heatmap of Cell-cell interactions at day 0 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.

The most significant interaction are the ones between CXCL8 and SDC3 for the Inflammation cluster to DNA damage and Cell Cycle clusters, also this exhibits a high interaction specificity. The cell interaction from Proliferation from the ligand FGF7 to FGFR4 to the clusters Cell Cycle, DNA damage, Hypoxia, Inflammation and Quiescence also show a high expression (Table 2 and Figure 10).

Table 2. Top 10 cell-cell interactions for Day 0 CancerSEA signatures.

Ligand	Receptor	ligand.complex	receptor.complex	Aggregate rank
Inflammation	DNAdamage	CXCL8	SDC3	4.476E-06
	Cell.cycle			5.3265E-06
Proliferation	Cell.cycle	FGF7	FGFR1	5.4509E-06

Hypoxia	Inflammation	LGALS1	ITGB1	5.6384E-06
DNA damage	Quiescence	ALDH1A3	CRABP2_RXRA	5.9982E-06
			CRABP2_RARA	6.5808E-06
			CRABP2_RARG	6.9235E-06
Inflammation	unknown	CXCL8	SDC3	7.2491E-06
Hypoxia	Cell.cycle	LGALS1	ITGB1	7.2949E-06
DNA repair	Hypoxia	MIF	CD44	7.2949E-06

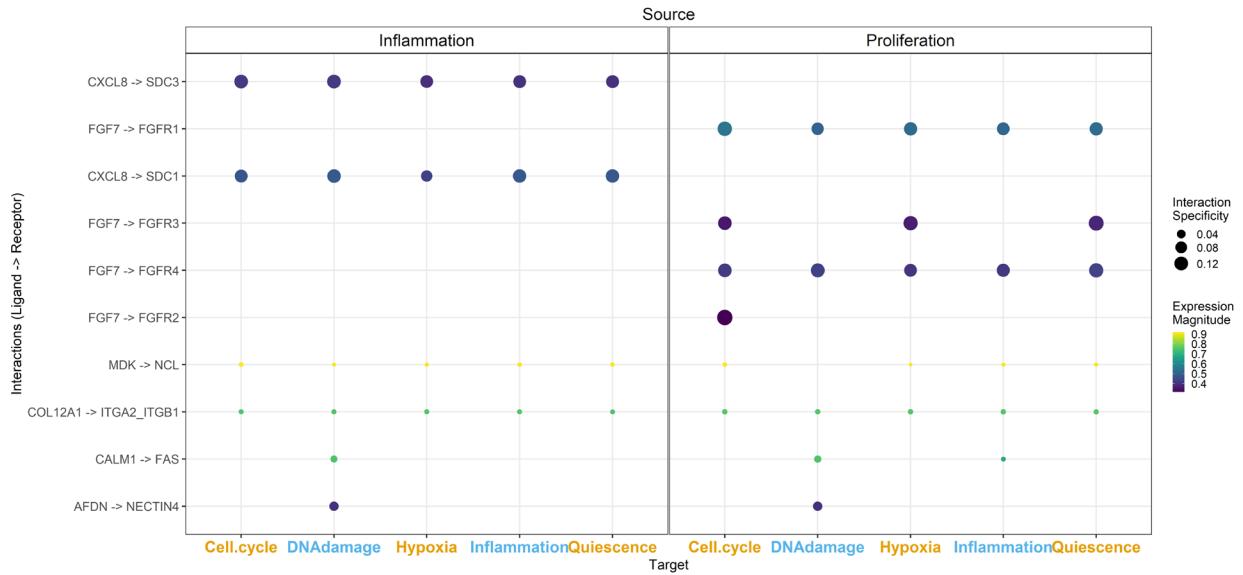


Figure 9. Most relevant cell-cell interactions for Day 0 on CancerSEA signatures.

3.1.2. Day 2 Cell-cell interactions

For the CancerSEA clusters at day 2, we can identify that most of the cell-cell interactions are between DNA damage to DNA damage, and to Hypoxia or Metastasis cell clusters. Also, Hypoxia is a cluster that has a lot of cell interactions as source, mostly to DNA damage cluster. On the other hand, DNA repair is the cluster with less cell abundance and less cell interactions (Figure 11).

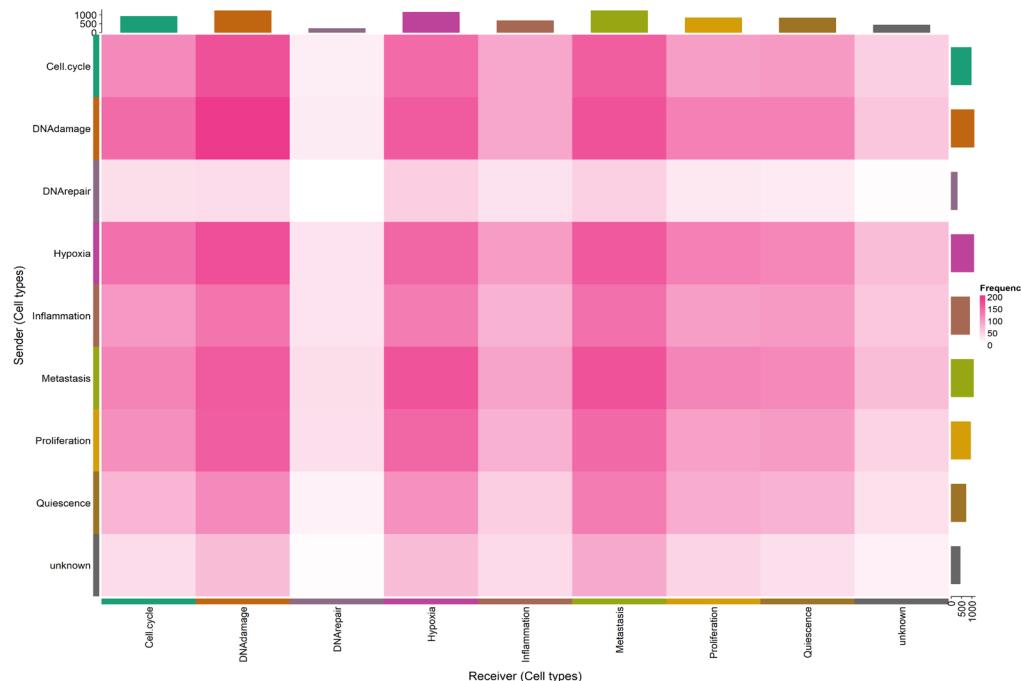


Figure 10. Heatmap of Cell-cell interactions at day 2 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.

At day 2 the main cluster ligand is the Inflammation cells and CXCL8, due to all top 10 interactions from day 2 is from this source and ligand, to SDC1/3 to target clusters such as Inflammation itself, Cell Cycle, DNA damage or Quiescent, also this are highly specify interactions (Table 3 and Figure 12).

Table 3. Top 10 cell-cell interactions for Day 2 CancerSEA signatures.

Ligand	Receptor	ligand.complex	receptor.complex	Aggregate rank
Inflammation	Inflammation	CXCL8	SDC1	3.0531E-08
	Cell.cycle			3.3073E-08
	DNAdamage			3.5752E-08
	Quiescence		SDC3	3.8572E-08
	Cell.cycle			4.1536E-08
	Quiescence			4.4649E-08
	Hypoxia		SDC1	4.7913E-08
	unknown		SDC3	5.1332E-08
				5.491E-08
			SDC1	5.8651E-08

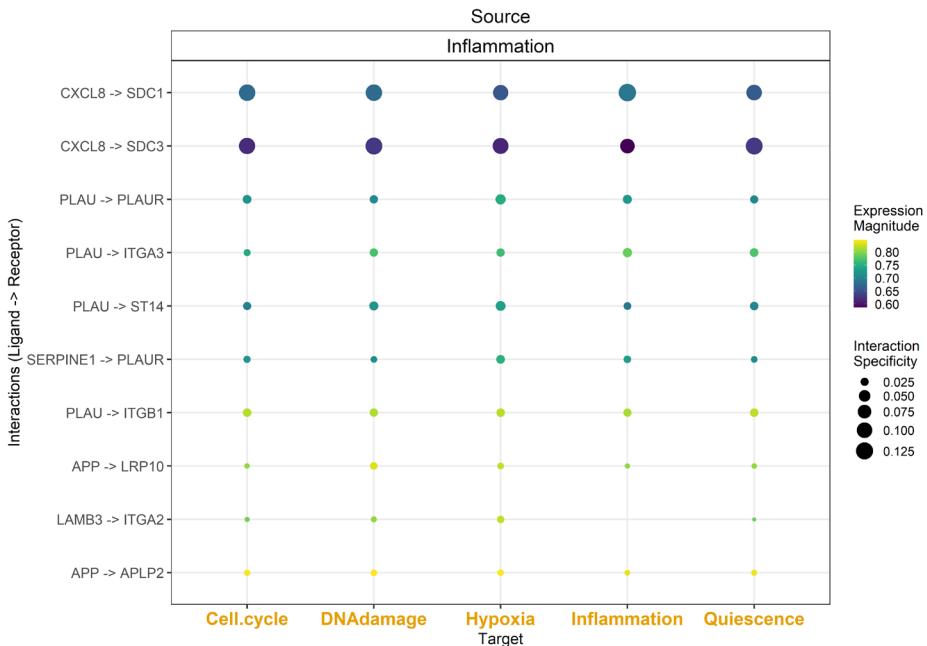


Figure 11. Most relevant cell-cell interactions for Day 2 on CancerSEA signatures.

3.1.3. Day 7 Cell-cell interactions

At day 7, metastasis is where most of cell-cell interaction occurs, as receiver cell interactions, being itself the one as the most frequent source of cell interactions, followed by Stemness and Inflammation clusters. DNA repair cluster is the one with less cell interactions and abundance (Figure 13).

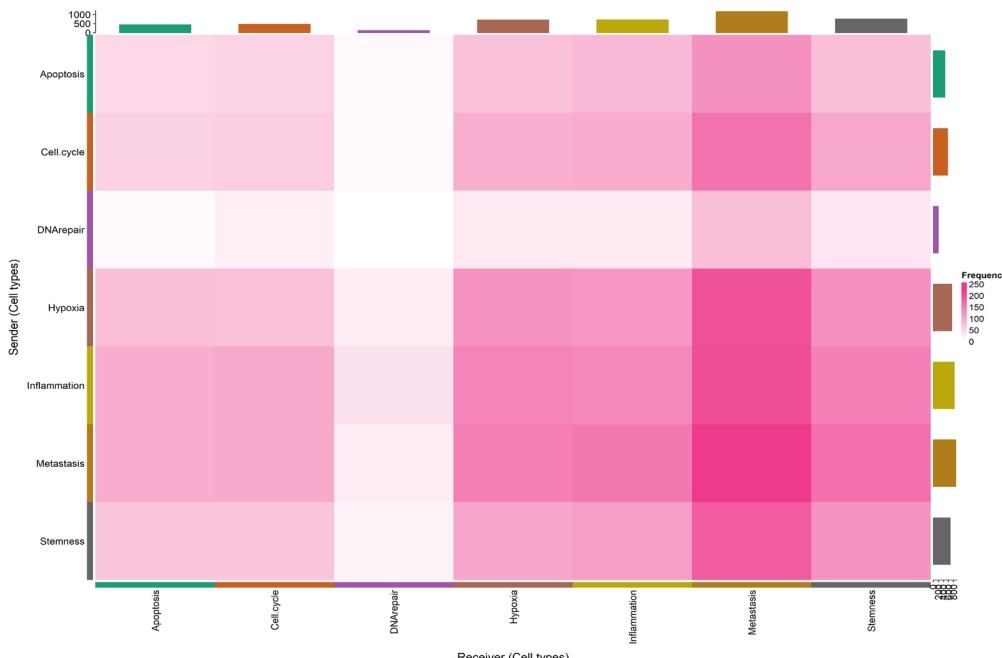


Figure 12. Heatmap of Cell-cell interactions at day 7 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.

At day 7 we identified Hypoxia to inflammation cluster as one of most significant interactions, from the ligand *ANGPTL4* to *SDC4* (Table 4). *TNFSF15* to *TNFRSF6B* from Inflammation cluster to Metastasis refers as a high specific interaction (Figure 14).

Table 4. Top 10 cell-cell interactions for Day 7 CancerSEA signatures.

Ligand	Receptor	Ligand complex	Receptor complex	Aggregate rank
Hypoxia	Inflammation	ANGPTL4	SDC4	3.5367E-08
Inflammation	Cell cycle	CXCL8	SDC1	2.7586E-07
Apoptosis	Stemness	VEGFA	GRIN2B	3.3887E-07
Inflammation	Metastasis	LAMC2	ITGB4	6.3833E-07
	Inflammation	FGF2	SDC4	7.6763E-07
Apoptosis		ANGPTL4		1.1646E-06
Inflammation	Metastasis	DUSP18	ITGB4	2.398E-06
Stemness		LAMB1		3.0332E-06
Inflammation		TNFSF15	TNFRSF6B	3.378E-06
	Inflammation	CD59	TNFSF15	3.378E-06

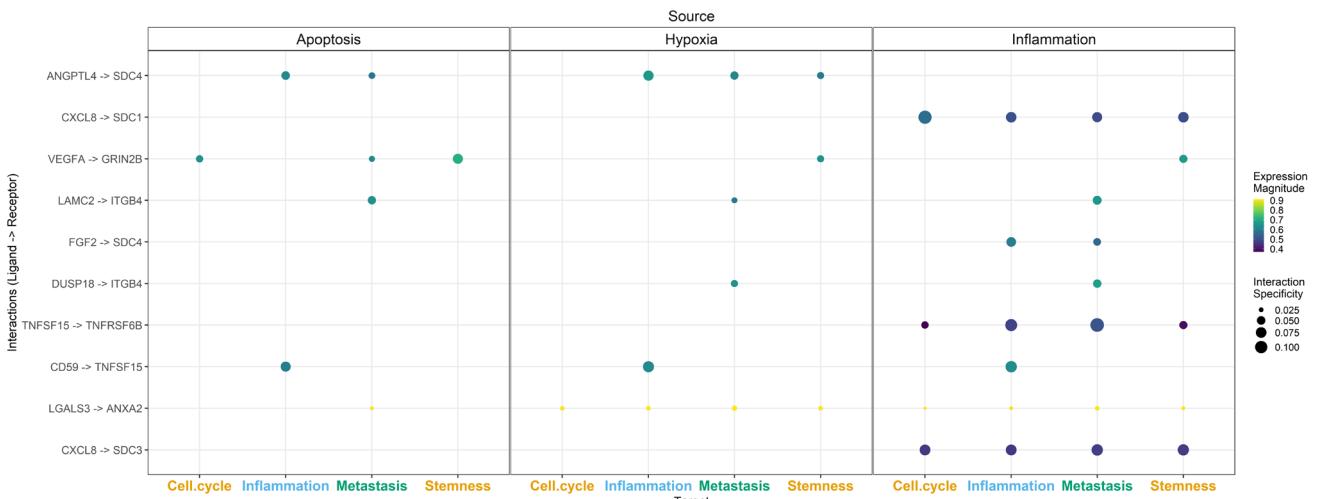


Figure 13. Most relevant cell-cell interactions for Day 7 on CancerSEA signatures.

3.1.4. Day 14 Cell-cell interactions

At day 14, Metastasis cell population becomes the main source and interaction for all the other cell populations. The most frequently interactions are the ones within Metastasis cluster, from metastasis to DNA damage, Cell cycle, and to Stemness. For the other cell clusters, Stemness and DNA damage clusters are frequent main sources for the interaction to Metastasis cell population (Figure 15).

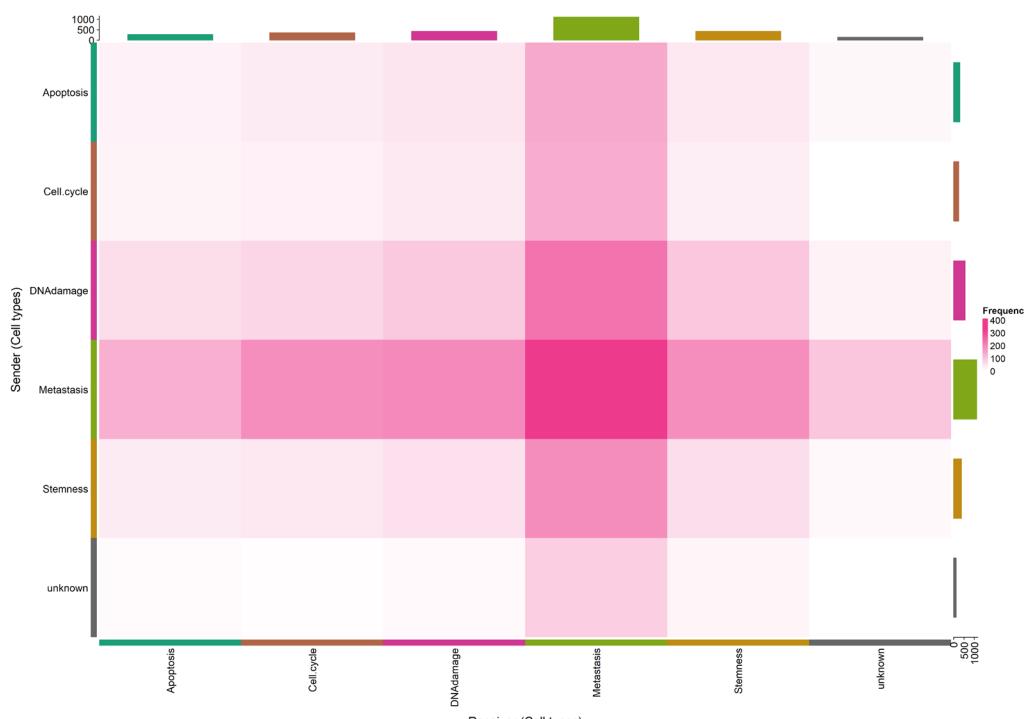


Figure 14. Heatmap of Cell-cell interactions at day 14 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.

At day 14, we identified cluster 5 as the main source the most significant cell-interactions, to clusters 5, 2, 3 or 1, for the ligands *CD59* to *TNFSF15* or *LAMB2/3* and *LAMA3* to *ITGB1/4* (Table 5). Interactions from cluster 5 to 5, from ligands *HSPG2*, *GAST* and *HBEGF* to receptors *LRP1*, *ADGRG1* and *CD82* respectively are high specific interactions (Figure 16).

Table 5. Top 10 cell-cell interactions for Day 14 CancerSEA signatures.

Ligand	Receptor	Ligand complex	Receptor complex	Aggregate rank
Metastasis	Metastasis	LAMC2	ITGA3	2.4274E-07
			ITGB4	5.749E-07
Apoptosis	Metastasis	LGALS9	CD47	7.8774E-07
Metastasis		LAMC2	ITGB1	1.1219E-06
Apoptosis	Cell cycle	LGALS9	SLC1A5	3.1502E-06
		LGALS9	CD44	3.1502E-06
Metastasis	Metastasis	LGALS1	ITGB1	4.7768E-06
		LAMC2	ITGA2	5.6037E-06
		LAMA3	ITGA3	7.0144E-06
Apoptosis	Stemness	LGALS9	CD44	7.0861E-06

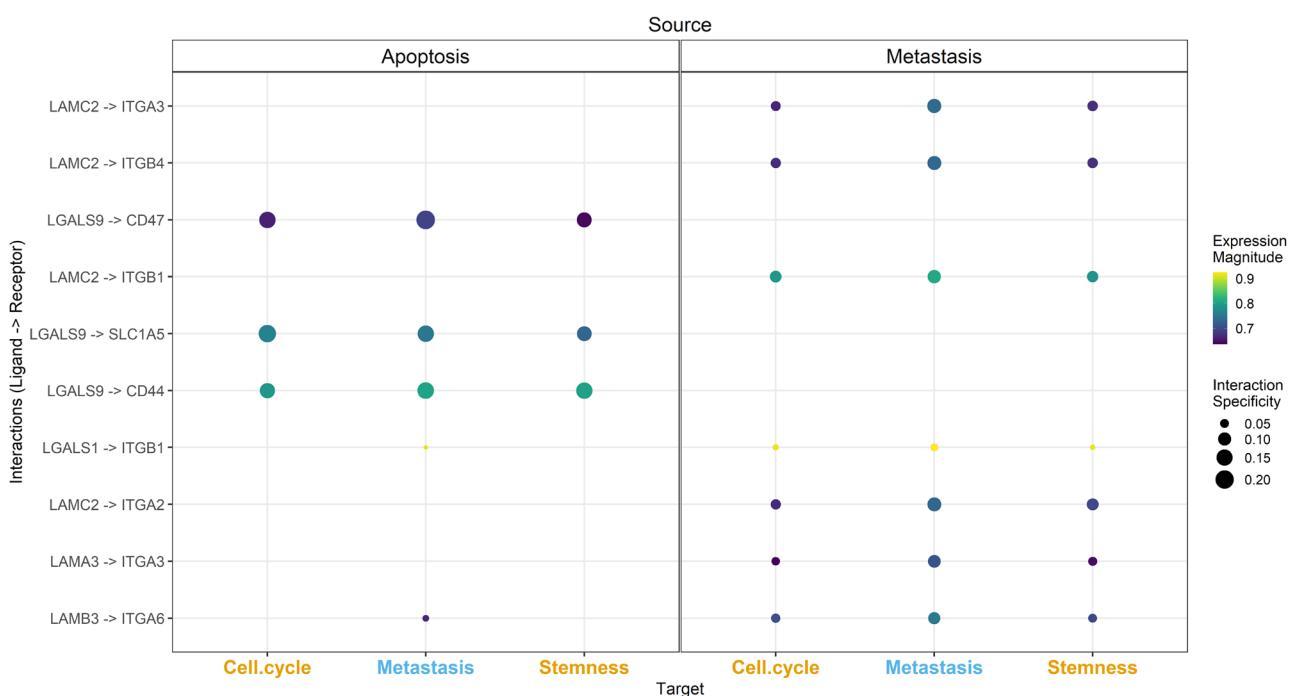


Figure 15. Heatmap of Cell-cell interactions at day 14 between CancerSEA signatures, interaction frequency are represented by the intensity of the color.

3.2. Integrative Analysis

The unintegrated method UMAP is shown in Figure 17, where cells are grouped based on the days. This map clearly illustrates the batch effect of the different days. Clusters were identified without using an integration methodology and are represented in 2D as seen in Figure 18. Additionally, the UMAP is split into different time points, showing clusters from day 0, clusters from day 2, and so on.

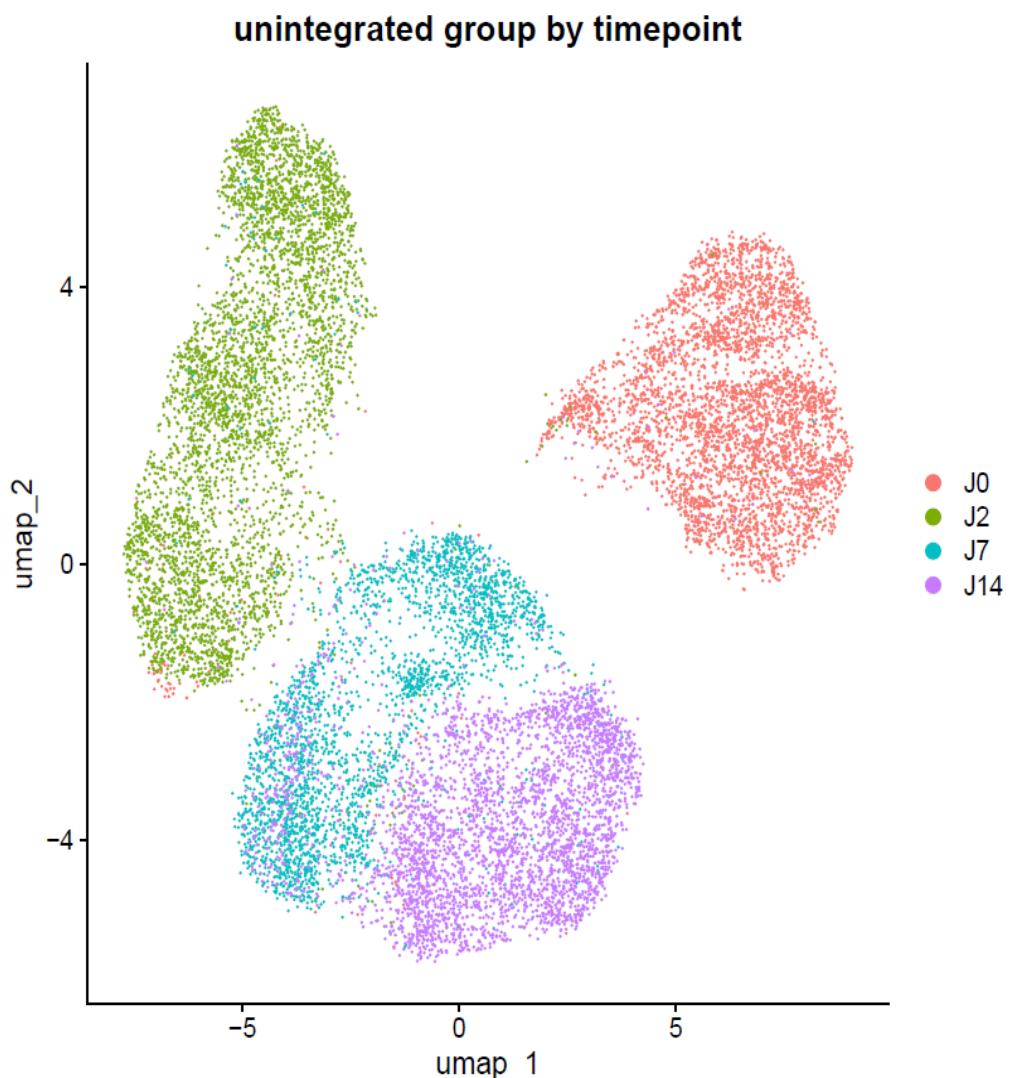


Figure 16. UMAP merging all Time points, representing the unintegrated analysis. Each timepoint is represented in different colors as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).

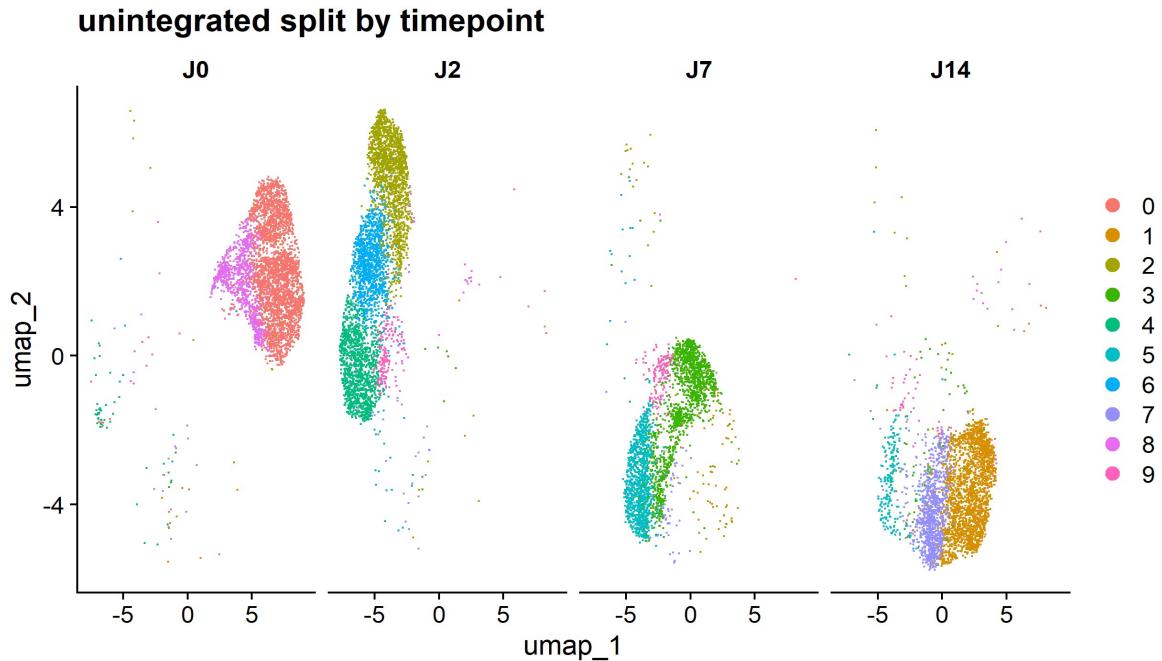


Figure 17. UMAP merging all data times matrices without integrating methodology, represented split by days (J0, J2, J7 and J17), a total of 9 different Clusters are identified.

We then applied the Harmony integration method to mitigate the batch effect across different cell time points. As shown in Figure 19, the batch effect from the Figure 16 was corrected, and cells from different time points were spatially merged. Subsequently, we identified Seurat clusters based on this final correction, recognizing 8 distinct clusters. For easier visualization of the clusters, we created plots split by the different days, from day 0 to day 14 (Figure 20). Finally, based on the CancerSEA table, we distinguished the cancer cell functional states (signatures) separately for each day (Figure 21).

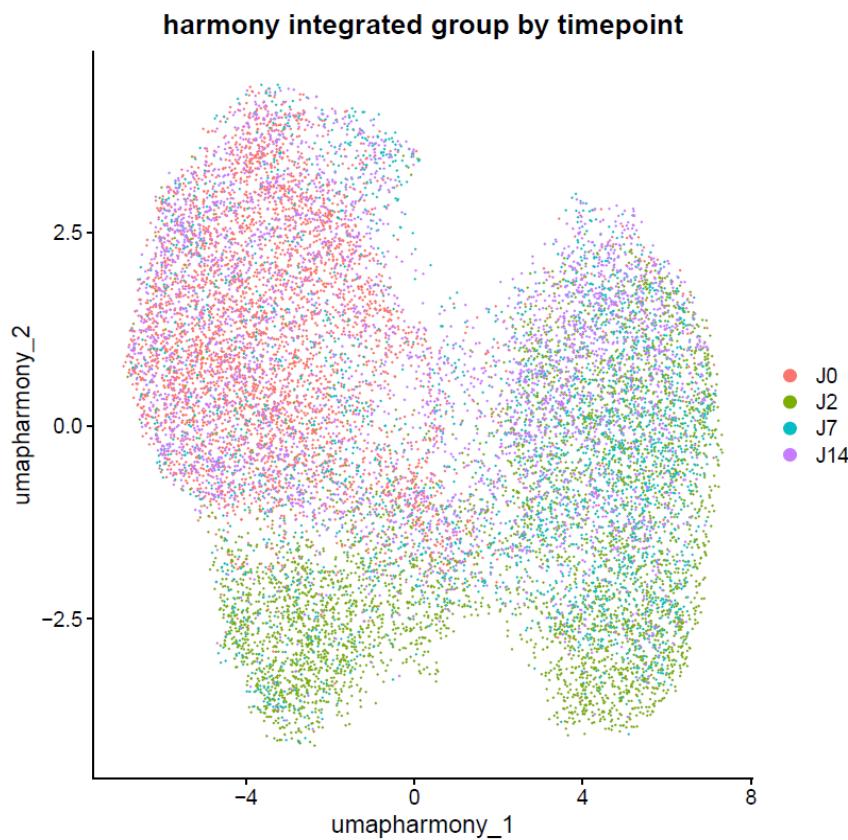


Figure 18. UMAP merging all Time points, representing the integrated analysis using Harmony as Integration methodology. Each timepoint is represented in different colors as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).

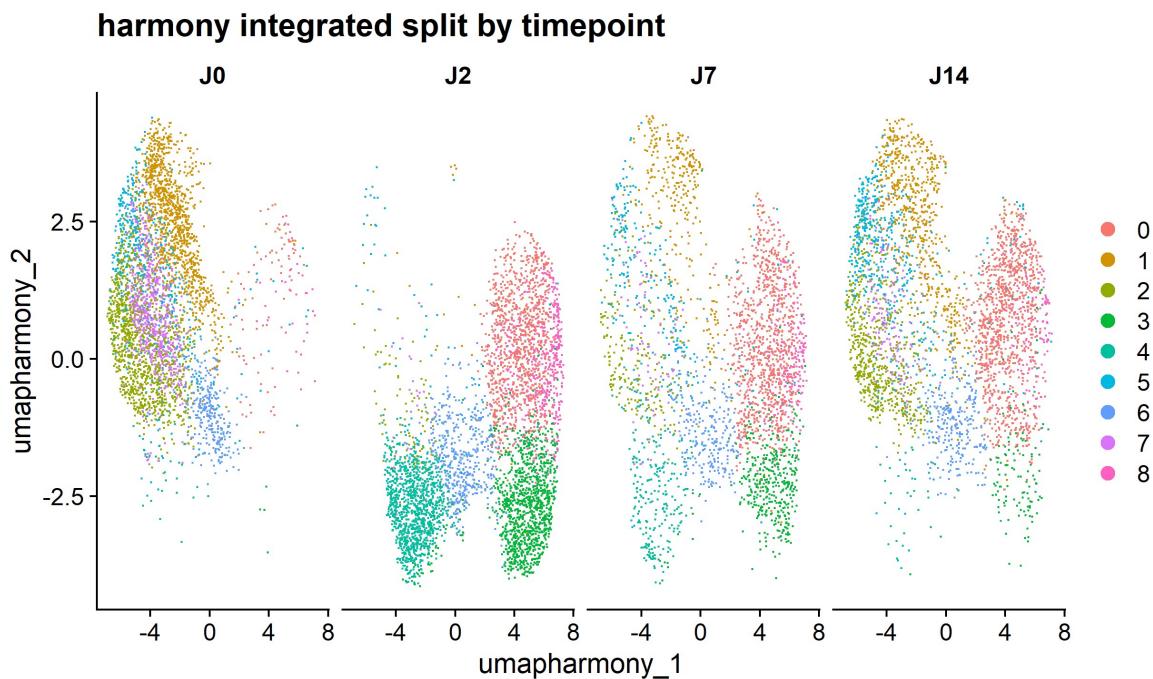


Figure 19. UMAP merging all data times matrices using Harmony as Integration methodology, represented split by days (J0, J2, J7 and J17), a total of 8 different Clusters are identified.

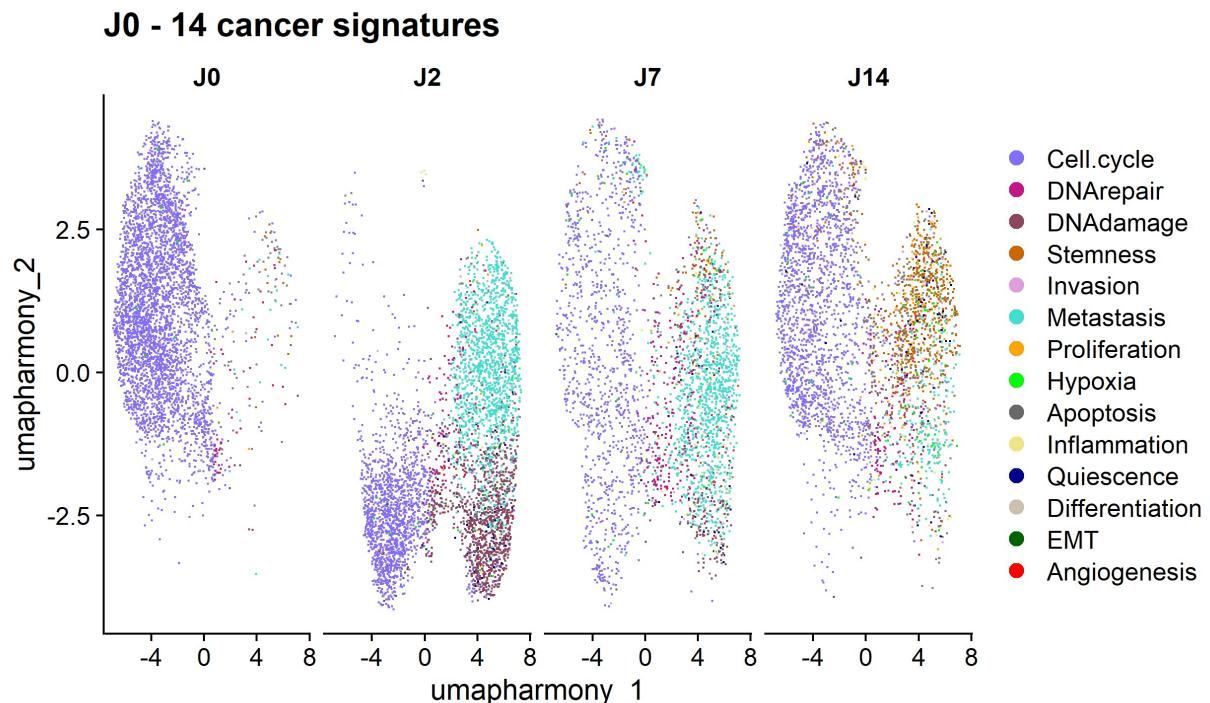


Figure 20. UMAP merging all data times matrices using Harmony as Integration methodology, represented split by days (J0, J2, J7 and J17), the CancerSEA Cancer Cell signatures were represented.

3.2.1. Cell-cell interactions

DNA repair cluster exhibits the less frequency interactions. Interactions from DNA damage to Differentiation is the highest frequency cell-cell interaction between cluster, being DNA damage and differentiation as the most abundant cell clusters (Figure 22). Moreover, CXCL8 refers as the most relevant ligand complex, being the ligand of all 10 most significant interactions which have a high expression, and all from the Inflammation cluster (Table 6 and Figure 23).

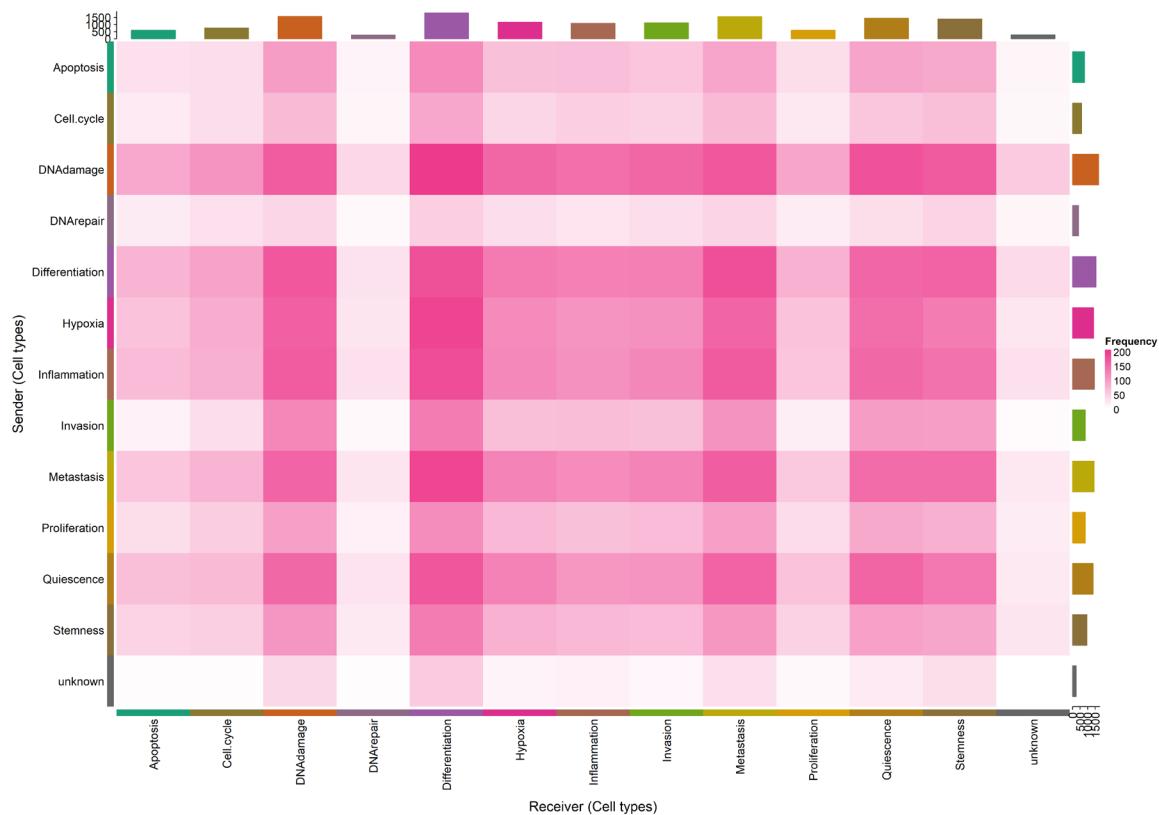


Figure 21. Heatmap of Cell-cell interactions for the integrated model between CancerSEA signatures, interaction frequency are represented by the intensity of the color.

Table 6. Integrative analysis top 10 cell-cell interactions of CancerSEA signatures.

Ligand	Receptor	Ligand complex	Receptor complex	Aggregate rank
Inflammation	Differentiation	CXCL8	SDC3	3.8847E-08
	DNA damage		SDC1	4.0443E-08
	Cell cycle		SDC3	5.8861E-08
	Inflammation		SDC1	6.3113E-08
	Differentiation		SDC3	6.7563E-08
	Hypoxia			7.4624E-08
	Stemness			8.4781E-08
	Metastasis			8.7456E-08
	Quiescence			9.5819E-08
	DNA damage			9.8721E-08

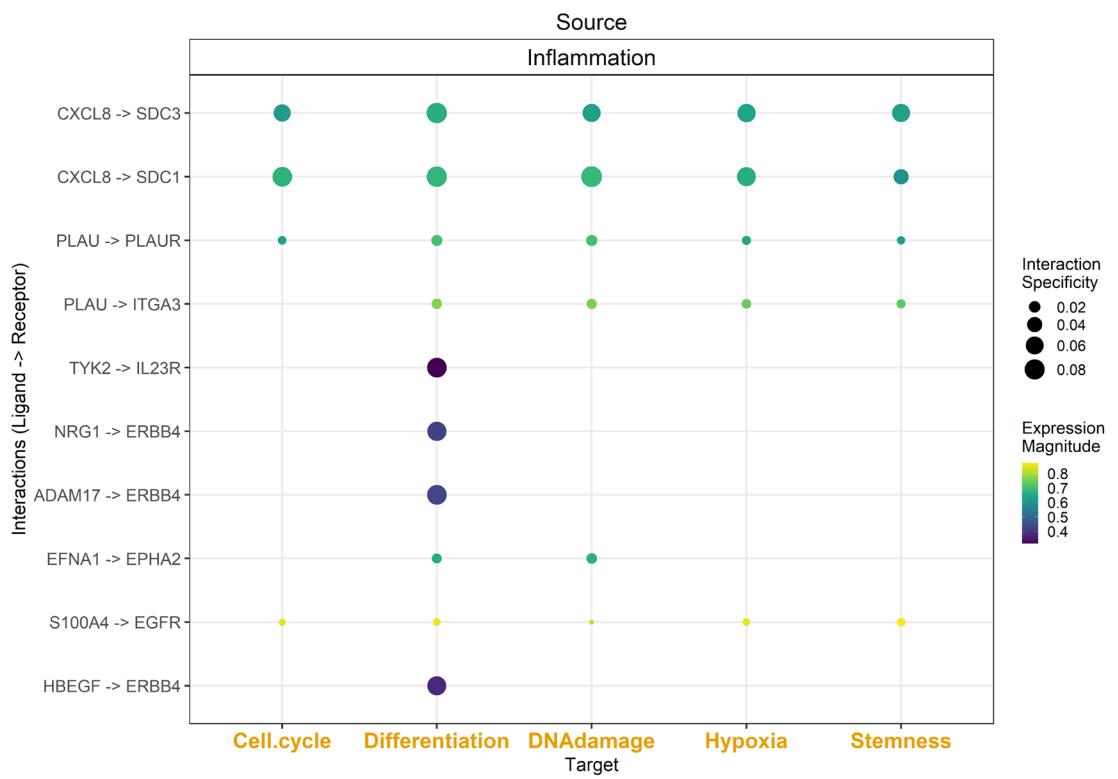


Figure 22. Most relevant cell-cell interactions for the integrative analysis on CancerSEA signatures.

4. Conclusion and Discussion

While the primary aim of this Master's thesis was to investigate and characterize relevant cancer cell subpopulations appearing in response to 5-FU, this objective has been successfully accomplished. Additionally, once these subpopulations were defined, we aimed to establish CCC networks between them, which was archived using Liana's Package and its results analysis. Most of the specific objectives described in the initial section of this work were met, except for the analysis of LR interactions. Although the results for LR interactions were obtained, they were not analyzed.

In terms of planning, the work followed the described plan closely. However, some deadlines that were defined initially were missed towards the final steps, such as the LR analysis, which could have been completed with better planning.

This study maintained transparency by using clear, reproducible research methods, such as open-source programs like R-studio and its packages (like the use of "renv" package), and making scripts and repositories available on platforms like GitHub. By understanding CRC therapy tolerance, the research aims to develop treatments effective for all populations, thus reducing disparities. Encouraging interdisciplinary and diverse team collaborations between the Master's student and the collaborating entities fostered workforce diversity. If the molecular mechanisms studied in this work are well established, effective treatments could reduce overall healthcare costs and decrease hospitalization costs. Early interventions may also provide long-term savings for healthcare system. Improved understanding of the underlying mechanisms of 5-FU treatment resistance in CRC could enhance the quality of life for patients. Overall, the anticipated positive impacts in terms of sustainability and ethical-social considerations have been archived. No significant negative impacts requiring mitigation were identified.

4.1. By Time points Analysis

4.1.1. Clustering and Dynamics

Seurat identified 6 clusters for all days except for day 14, where only 5 clusters were observed. Notably, after 5-FU treatment (day 2), the proportion of cells

from clusters 0 and 5 increased, potentially due to the therapy's effect. Interestingly, in the days following treatment, this dysregulation of cells proportion within clusters returned to day 0 levels. However, on day 14, cells from cluster 6 were no longer recognized, and the proportion of cells from cluster 3 increased.

In terms of cell cycle identification of each cell, we noted that after the treatment (day 2), the number of cells in the G1 phase increased while the percentage of cells in the S phase decreased. In the subsequent days (7 and 14), these percentages returned to day 0 levels. Interestingly, by day 14, there was a resurgence of S phase cells, indicating a potential relapse phase. The initial increase in G1 phase cells suggests that resistant cells may enter a cell cycle arrest stage induced by 5-FU therapy, which is progressively reversed over time. The 5-FU treatment appears to induce cell cycle arrest in the G1 phase and inhibit DNA synthesis, aligning with known mechanisms of action²³. However, the subsequent increase in S (DNA synthesis) and G2M (Mitosis)²⁴ phase cells indicates a return to a replicative and proliferative environment and may indicate that resistant cells have evaded the apoptotic effect of 5-FU. Prolonged G1 and S phases may provide sufficient time for tumor cells to counteract 5-FU induced DNA damage through the activation of DNA repair pathways. All this findings, highlighting the ability of tumor cells to adapt and resume proliferation, a phenomenon supported by recent studies on cellular reprogramming and drug resistance^{25,26}.

CancerSEA Cancer signatures analysis revealed a decrease in DNA repair cells and a concurrent increase in DNA damage cells following 5-FU treatment, likely due to the therapy's mechanism of action. Interestingly, no apoptotic cells were observed immediately after the treatment (day 2), with apoptotic cells appearing on days 7 and 14. While 5-FU therapy initially increases apoptotic cell death, this was not observed in this case, suggesting possible resistance into starting the apoptotic cycle despite the presence of increased DNA damage markers, a possible hallmark of the treatment's effect. At day 7, markers of DNA repair were significantly increased, indicating potential activation of genome repair mechanisms, while DNA damage markers decreased, suggesting partial mitigation of the 5-FU's effect by therapy-resistant cancer cells. Furthermore,

the high proliferative cell population observed at day 7 may have transformed into the high stemness population present at day 14, suggesting the tumor's adaptative response to the therapy and the potential role of stemness in treatment resistance²⁷, and a possible relapse phase after the therapy.

In summary, these findings highlight the remarkable ability of cancer cells to adapt and evade the effects of 5-FU therapy. The observed cell cycle arrest, initial DNA damage, and dysregulation of cell populations followed by a return to proliferation and a potential stemness increase suggest a multi-step resistance process. This underscores the importance of developing therapeutic strategies that can overcome these adaptive mechanisms and improve treatment outcomes for cancer.

Principal interactions:

- **Day 0:**

- DNA Damage to Cell Cycle: Frequent interactions suggest that DNA damage activates cell cycle control mechanism to repair or eliminate damaged cells.
- Hypoxia to Cell Cycle: Hypoxia also influences the cell cycle, possibly due to the colorectal cancer cells' response to the environment stress.

Even though interactions from DNA repair to the cancer cell markers clusters are low, suggesting that the cells may not be focusing on active DNA repair.

- **Day 2 (Treatment Day):**

- Emergence of metastasis cell interactions: Cells start to show high interactions with metastatic cells, interacting with DNA damage, hypoxia, and themselves (metastasis).
- Similarity to Day 0: The interaction profile is similar to Day 0, with the significant addition of metastatic cell interactions.

It is also relevant to note that the absence of apoptotic cells and interactions, indicating that cells are avoiding programmed cell death, a characteristic of metastatic cells acquiring more aggressive properties.

- **Day 7:**

- Interactions from metastatic cells to Hypoxia, Inflammation, Stemness and Cell Cycle: Metastatic cells actively interact with these processes, suggesting an aggressive and adaptable state.
- Limited interactions from DNA Repair: The reduced interactions from DNA Repair cells could reflect the prioritization of the DNA repair focused on promoting proliferation and adaptation.

The presence of apoptotic and stemness cells, suggest that some cells are being removed, while stemness indicates a subpopulation of cells with self-renewal capacity. Inflammation cells as the main communication source of cell communication could facilitate tumor invasion and adaptability.

- **Day 14:**

- Metastatic cells as main source of interactions: Metastatic cells generate numerous interactions with themselves and other processes such as DNA damage, Cell Cycle and Stemness.

By day 14, Stemness and DNA Damage are key targets, indicating mechanisms that maintain the tumor's aggressive profile and adaptability.

As the treatment progresses, an increase in the complexity of interactions is observed, with a focus on metastasis and a relative decrease in DNA repair. Additionally, cells are adapting to the therapy, with an increase in metastatic cells and stemness cells, indicating a more aggressive and resilient tumor behavior. This highlights the high plasticity capacity of tumor cells and the high heterogeneity of the TME created by the 5-FU therapy^{28,29}.

4.1.2. LR interactions

Here, we obtained relevant information from the data, but all the information will be analyzed in a future project due to the complexity of interpreting results related to LR interactions.

4.2. Integrative Analysis

4.2.1. Clustering and Dynamics

Regarding the Integrative Analysis, we identify that the batch effect from the different time points is reduced with the application of Harmony Integrative method. Once the Harmony methodology is applied, we observe that cells from Seurat clusters 1 and 5 practically disappear on the 2nd day, and this cell population increases again after the treatment. Interestingly, cells from clusters 3, 4, and 8 are infrequent on day 0, but appear on the day of the treatment. In the following days, these cell populations disappear. An interesting cluster of cells is cluster 0, which is infrequent on day 0 and day 2 but is maintained during the following days.

When CancerSEA Cancer Cell Markers are identified, we observe that the identified cells from the first day are predominantly related to the cell cycle. These cell cycle cells disappear on day 0 and reappear on days 7 and 14. Moreover, on the day of treatment, cells related to DNA damage and hypoxia appear, forming the new clusters observed on the day of treatment. Interestingly, cells from the DNA damage and metastatic clusters slightly disappear on day 7, while stemness-like cells appear and increase on days 7 and 14. We clearly see a temporal evolution of cellular response and adaptation to treatment, providing insight into the dynamic of tumor cell populations under therapeutic stress. Additionally, the identified cluster highlight the heterogeneous environment created by the 5-FU therapy and help to understand the molecular mechanisms of plasticity cell and its adaptability.

Based on the CancerSEA signature, we identified the CCC between those Cancer Cell Markers. We can identify some interconnections, such as DNA Damage as a significant source of cell interactions, actively interacting with various clusters including Differentiation, Quiescence, and other DNA Damage cells. DNA Repair remains relatively isolated, indicating it is not heavily involved in inter-cluster communications either as a source or a target. On the other hand, the clusters Differentiation, DNA Damage, Metastasis, Quiescence and Stemness are key targets, receiving interactions from multiple sources, indicating their central role in the cell communication network.

As interaction patterns, we identify that the source of interactions from metastasis and Hypoxia, likely driving significant changes in the TME. Differentiation and DNA Damage serve as both important sources and targets, suggesting a dynamic interplay where these processes are both influencing and by influenced by other cellular states.

As well as the by time points analysis, the integrative analysis highlights the complex and dynamic nature of cell-to-cell communications within the tumor environment, emphasizing the critical roles of metastasis, hypoxia, differentiation, and DNA Damage in orchestrating these interactions, in the context of 5-FU therapy on Colorectal Cancer Cells.

4.2.2. LR Interactions

Here, as well as the by time points results, we obtained relevant information from the data, but all the obtained information is going to be analyzed in a future project.

4.3. Biological significance and Final conclusions

In this study, we aimed to understand the complex dynamics of CRC cell populations in response to 5-FU therapy by performing integrative and temporal analysis of scRNA-seq data. Our findings reveal several key insights into the biological processes and cellular interactions that underpin tumor plasticity and resistance to the treatment.

4.3.1. Dynamics from Seurat Clusters

Our temporal analysis identified six distinct cluster of cells, with notable changes in their proportions and interactions over time, especially after the administration of 5-FU. By day 14, the absence of cluster 6 and the increase in cluster 3 indicate a substantial shift in cell population dynamics, reflecting the tumor's adaptative response to treatment.

4.3.2. Cell Cycle Regulation

The increase in G1 phase cells post-treatment indicates a cell cycle arrest, likely as a surviving mechanism to prevent the propagation of damaged DNA. The resurgence of cells on the phases S and G2M by day 14 suggests a

recovery of proliferative capacity, pointing to the cells' ability to overcome the initial cell cycle arrest and resume replication. This is indicative of resistance to the cytotoxic effects of 5-FU, allowing the tumor to maintain its growth potential.

4.3.3. 5-FU therapy adaptability

The initial response to the treatment is identified by the visualization of DNA damage, which activates DNA damage response pathways. By day 7, there is an increase in DNA repair cell markers, suggesting the activation of genome repair mechanisms as a resilience strategy against 5-FU-induced damage, and a phenotypic evolution of the cancer cells. Moreover, the rise in stemness markers by days 7 and 14, especially on day 14, indicates a shift towards a more stem-like, resilient cell population. These cells possess self-renewal capabilities such as CSCs. This indicates that cancer cells can adaptively upregulate DNA repair pathways to counteract therapeutic stress and evolve phenotypically depending on environmental conditions.

Additionally, the presence of hypoxia markers on the day of treatment could reflect the tumor's adaptation to a stressed microenvironment, which can further drive aggressive behavior and metastasis. The increase in metastatic cell interactions and the persistence of metastatic clusters over time underscore the therapy-induced selection for cells with high metastatic potential, contributing to disease progression and recurrence.

4.3.4. Biological Implications and Future Directions

The study underscores the importance of understanding both genetic and nongenetic mechanisms driving tumor progression, resistance, and tolerance. The high plasticity and heterogeneity of tumor cells in response to 5-FU therapy reveal potential targets for therapeutic intervention. The findings suggest that targeting stemness and metastatic pathways could be critical in overcoming resistance and preventing relapse.

Future research should focus on elucidating the molecular mechanisms behind this adaptive response and exploring combination therapies that target both proliferative and stem-like tumor cell populations to improve treatment outcomes for CRC patients. Also, a more detailed study must be done based on

the LR interactions not discussed in this study. Once transcriptomic and cellular mechanisms are well understood, further functional studies, both *in vitro* and then *in vivo*, should be conducted to validate the identified resistance and adaptation mechanisms. Additionally, investigating the combination of 5-FU with other therapeutic agents that can inhibit DNA repair pathways or stem cell signaling could improve treatment efficacy and reduce the likelihood of tumor recurrence.

5. Glossary

Colorectal Cancer (CRC)

single-cell RNA sequencing (scRNA-seq): A powerful technique for measuring the expression levels of all genes across thousands to millions of individual cells, enabling the identification of rare cell subpopulations, inference of complex gene-gene regulatory factors, and tracing of developmental lineages.

cell-cell communication (CCC): The process by which cells interact and communicate with each other through the expression of genes encoding corresponding ligands, receptors, intermediate signaling proteins, and intracellular targets.

tumor microenvironment (TME): The environment surrounding a tumor, including various cell types such as cancer cells, stromal fibroblasts, and immune cells, as well as the extracellular matrix and signaling molecules.

Colonic stem cells (CSCs): Defined by the ability to self-renew while maintaining their ability to generate a progeny of both tumorigenic and nontumorigenic cancer cells through asymmetric division³¹.

Proliferative CSCs (proCSCs): A subtype of CSCs that are transcriptionally similar to healthy CSCs but are trapped in a mitotic state and fail to differentiate into the secretory and absorptive cells of the healthy colon.

revival stem cells (revCSCs): A type of stem cells that help resist 5-FU therapy. During CRC progression, these cells polarize between proCSC and revCSC phenotypes to establish primary tumors, but they cannot expand to form primary tumors due to their slow cell-cycle activity.

Institut de Génomique Fonctionnelle (IGF): A research institute located in Montpellier, France, focusing on functional genomics.

Lyon Cancer Research Center (CRCL): A cancer research center located in Lyon, France, dedicated to studying cancer biology and developing new therapeutic strategies.

5-fluorouracil (5-FU): Represents an anti-metabolite with substitution of fluorine for hydrogen at the C-5 position of uracil. The thymine-uracil/5-FU exchange caused by thymine replacement in DNA consequently leads to the formation of adenine-uracil/5-FU base pairs³².

ligand-receptor interactions (LR): The simplest class of CCC inference methods are those that infer protein-protein interactions from dissociated single-cell transcriptomics data, commonly referred to as ligand-receptor interaction inference methods³³.

shared nearest neighbor (SNN): An algorithm used in clustering analysis to identify groups of similar data points based on their nearest neighbors in a shared space.

Differential Gene Expression (DGE): The process of comparing gene expression levels between different conditions or groups to identify genes that are significantly upregulated or downregulated.

6. Bibliografia

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
2. Cho, Y.-H. *et al.* 5-FU promotes stemness of colorectal cancer via p53-mediated WNT/β-catenin pathway activation. *Nat. Commun.* **11**, 5321 (2020).
3. Jeught, K. V. D., Xu, H.-C., Li, Y.-J., Lu, X.-B. & Ji, G. Drug resistance and new therapies in colorectal cancer. *World J. Gastroenterol.* **24**, 3834–3848 (2018).
4. Qin, X. & Tape, C. J. Functional analysis of cell plasticity using single-cell technologies. *Trends Cell Biol.* S0962892424000060 (2024) doi:10.1016/j.tcb.2024.01.006.
5. Househam, J. *et al.* Phenotypic plasticity and genetic control in colorectal cancer evolution. *Nature* **611**, 744–753 (2022).
6. Elbadawy, M., Usui, T., Yamawaki, H. & Sasaki, K. Development of an Experimental Model for Analyzing Drug Resistance in Colorectal Cancer. *Cancers* **10**, 164 (2018).
7. Mohr, S. E. *et al.* Methods and tools for spatial mapping of single-cell RNAseq clusters in *Drosophila*. *Genetics* **217**, iyab019 (2021).
8. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).

9. Detection of differentially abundant cell subpopulations in scRNA-seq data.
<https://www.pnas.org/doi/epdf/10.1073/pnas.2100293118>
doi:10.1073/pnas.2100293118.
10. Bridges, K. & Miller-Jensen, K. Mapping and Validation of scRNA-Seq-Derived Cell-Cell Communication Networks in the Tumor Microenvironment. *Front. Immunol.* **13**, (2022).
11. Tape, C. J. Plastic persisters: revival stem cells in colorectal cancer. *Trends Cancer* **10**, 185–195 (2024).
12. Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* **37**, 471–484 (2020).
13. Competències transversals - Compromís global - UOC.
<https://www.uoc.edu/portal/ca/compromis-social/ciutadania-global/competencies-transversals/index.html>.
14. SCINA: Semi-supervised Category Identification and Assignment | single cell r package | cell clustering tools | deconvolution software.
<https://dbai.biohpc.swmed.edu/scina/>.
15. Yuan, H. *et al.* CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res.* **47**, D900–D908 (2019).
16. Türei, D. *et al.* Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* **17**, e9923 (2021).
17. Dimitrov, D. *et al.* Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat. Commun.* **13**, 3224 (2022).
18. Waltman, L. & van Eck, N. J. A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471 (2013).

19. Chari, T. & Pachter, L. The specious art of single-cell genomics. *PLOS Comput. Biol.* **19**, e1011288 (2023).
20. LIANA Tutorial. https://saezlab.github.io/liana/articles/liana_tutorial.html.
21. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
22. Chen, B., Banton, M. C., Singh, L., Parkinson, D. B. & Dun, X. Single Cell Transcriptome Data Analysis Defines the Heterogeneity of Peripheral Nerve Cells in Homeostasis and Regeneration. *Front. Cell. Neurosci.* **15**, (2021).
23. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **3**, 330–338 (2003).
24. Cell Cycle | Cell Cycle: DNA Damage Checkpoints. 2–9 (2021) doi:10.1016/B978-0-12-819460-7.00178-X.
25. Azwar, S., Seow, H. F., Abdullah, M., Faisal Jabar, M. & Mohtarrudin, N. Recent Updates on Mechanisms of Resistance to 5-Fluorouracil and Reversal Strategies in Colon Cancer Treatment. *Biology* **10**, 854 (2021).
26. Watson, A. J. M. Apoptosis and colorectal cancer. *Gut* **53**, 1701–1709 (2004).
27. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).
28. Lee, S. Y. *et al.* Regulation of Tumor Progression by Programmed Necrosis. *Oxid. Med. Cell. Longev.* **2018**, 3537471 (2018).
29. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

30. Ghafouri-Fard, S. *et al.* 5-Fluorouracil: A Narrative Review on the Role of Regulatory Mechanisms in Driving Resistance to This Chemotherapeutic Agent. *Front. Oncol.* **11**, (2021).
31. Ricci-Vitiani, L., Fabrizi, E., Palio, E. & De Maria, R. Colon cancer stem cells. *J. Mol. Med.* **87**, 1097–1104 (2009).
32. 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. *Pharmacol. Ther.* **206**, 107447 (2020).
33. Dimitrov, D. *et al.* LIANA+: an all-in-one cell-cell communication framework. Preprint at <https://doi.org/10.1101/2023.08.19.553863> (2023).

7. Annex

All the working scripts have been uploaded on the following GitHub repository:

https://github.com/miquelg151/happy_penguin

QC metrics are defined in the 01_scripts/output/plots/QC

The top 50 most relevant cell markers of each Seurat Cluster saved in the
01_scripts/output/tables/sig.markers.by.timepoint.xlsx