

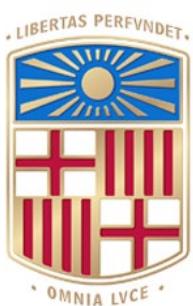
# **Impact of 5-FU treatment on tumor cell plasticity in colorectal cancer**

**UOC**

A scRNAseq analysis

Universitat  
Oberta  
de Catalunya

**Miquel Gratacós i Aurich**



UNIVERSITAT DE  
BARCELONA

MU Bioinf. i Bioest.

Àrea de treball final

**Tutor/a de TFM**

Laia Bassaganyas

**Professor/a responsable de  
l'assignatura**

Nom i Cognoms

Data Lliurament





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 © ANY EL-TEU-NOM.

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**

© (l'autor/a)

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

<b>Títol del treball:</b>	<i>Impact of 5-FU treatment on tumor cell plasticity in colorectal cancer</i>
<b>Nom de l'autor:</b>	<i>Miquel Gratacós i Aurich</i>
<b>Nom del director/a:</b>	<i>Laia Bassaganyas</i>
<b>Nom del PRA:</b>	<i>Nom i dos cognoms</i>
<b>Data de lliurament (mm/aaaa):</b>	<i>MM/AAAA</i>
<b>Titulació o programa:</b>	<i>Pla d'estudis</i>
<b>Àrea del Treball Final:</b>	<i>Area de TF</i>
<b>Idioma del treball:</b>	<i>anglès</i>
<b>Paraules clau</b>	<i>Màxim 3 paraules clau, validades pel tutor/a del TF</i>
<b>Resum del Treball</b>	
<i>Màxim 250 paraules, amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball</i>	
<b>Abstract</b>	
<i>A maximum of 250 words, detailing the purpose, context of application, methodology, results and conclusions of the work</i>	

## **Index**

1.	Introduction .....	1
1.1.	Context and Justification of the Work.....	1
1.2.	Objectives .....	2
2.1.1.	General objectives .....	2
2.1.2.	Specific objectives .....	2
1.3.	Impact on sustainability, ethical-social and diversity .....	3
1.4.	Approach and method followed.....	4
1.5.	Planning .....	6
1.5.1	Tasks.....	6
1.5.2.	Callendar .....	7
1.6.	Brief summary of products obtained.....	8
1.7.	Brief description of the other chapters of the report .....	8
2.	Material and Methods.....	10
2.1.	By Time points analysis .....	10
2.1.1.	Quality Control and normalization of Raw matrix. ....	10
2.1.2.	Dimensionality reduction.....	11
2.1.3.	Cluster the cells and UMAP. ....	11
2.1.4.	Cell-cell interaction analysis.....	12
2.2.	Integrative Analysis .....	12
2.2.1.	Perform integration and cluster identification .....	12
2.2.2.	Identify conserved cell type markers.....	13
2.2.3.	Cell-cell interaction analysis.....	13
3.	Results .....	14
3.1.	Analysis by time points.....	14

3.2. Integrative Analysis .....	17
4. Conclusion and Discussion .....	21
5. Glossari .....	22
6. Bibliografia .....	23
7. Annex .....	25

## ***List of Figures***

Figure 1 Gantt Diagram of the followed timeline during the project.	<b>¡Error!</b>
<b>Marcador no definido.</b>	
Figure 2. 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.	14
Figure 3. 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.	15
Figure 4. UMAP from scRNAseq analysis by time points. Clustering is defined by the CancerSEA cancer cell marker dataset. 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.	16
Figure 5. Proportion Plot of the cell populations at each time point based on the Cancer Cell Markers dataset, CancerSEA.	<b>¡Error! Marcador no definido.</b>
Figure 6. UMAP from the Integrative analysis, each timepoint is represented in different colors and as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).	17
Figure 7	<b>¡Error! Marcador no definido.</b>
Figure 8	<b>¡Error! Marcador no definido.</b>
Figure 9	<b>¡Error! Marcador no definido.</b>
Figure 10	<b>¡Error! Marcador no definido.</b>

## ***List of Tables***

Table 1. Deadlines and duration of the project planification.	7
---	---

v

# 1. Introduction

## 1.1. Context and Justification of the Work

Colorectal cancer (CRC) is the third most diagnosed cancer, affecting both men and women worldwide. Unfortunately, drug resistance is one of the main reasons for the low survival rates of CRC patients [1].

5-Fluorouracil (5-FU)-based chemotherapy, the mainstay therapy for patients with CRC, is often associated with resistance and consequently tumor recurrence. Despite the advancements in various modulation strategies, nearly half of CRC patients experience recurrence within 5 years of their treatment. In response to anticancer treatments, including 5-FU, tumor cells develop tolerance, promoting tumor recurrence and metastasis development [2], [3].

Cancer cells are heterogeneous in morphology, inheritance, and functions. Throughout cancer cells, the clinical outcomes of 5-FU treatments are often associated with cancer stem cells (CSCs), which are a small population of cancer cells that own the ability to self-renew and generate the tumor. It has been considered that CSCs might regulate the mechanisms of intrinsic or acquired drug resistance [4], enhancing the relapse of tumors treated with chemotherapy [2].

However, mechanisms of drug tolerance in cancer are still poorly understood, for this reason, revealing the underlying mechanism, is necessary for improving the outcome of CRC patients. Additionally, transcriptomic, and epigenetic studies describe cellular plasticity as one of the hallmarks of treatment tolerance. Specially, epigenetic modifications promote drug efflux, and the regulation of drug targets and associated signaling pathways. Nevertheless, in the context of tumor drug resistance, the translation of messenger RNA into proteins remains unexplained [5].

In recent years, with the development of single-cell RNA sequencing (scRNA-seq) technology have emerged powerful new tools and techniques that have provided new insights into the definition of molecular signatures of cell types, subtypes, data analysis, visualization, and mining of scRNA-seq datasets, such as Seurat analysis platform [6], [7]. For these reasons, these new technologies

are crucial for the understanding and study of transcriptomic and cell population development.

## 1.2. Objectives

### 2.1.1. General objectives

Originally, this study is part of a project from Institute de Génomique Fonctionnelle (IGF) of Montpellier, the Lyon Cancer Research Center (CRCL) which has several objectives (5):

1. Demonstrate that translational control is one of the molecular mechanisms responsible for cellular reprogramming in response to 5-FU.
2. **Isolate relevant cellular subpopulations for study. \***
3. **Phenotypically characterize the cellular subpopulations. \***
4. Identify the role of these subpopulations in treatment tolerance and tumor recurrence.

\* Specifically, on this study, the objectives that are going to be archived, are the 2-3.

### 2.1.2. Specific objectives

Based on the principal objectives aimed for this study, we've established a more detailed scheme, to create the workflow needed for this master thesis project:

2. Employ and optimize single-cell RNA-seq workflow using Seurat Package from R to obtain relevant cellular subpopulations.
  3. Perform detailed analysis of these cell subpopulations using appropriate methodologies, thanks to Seurat Package workflow and optimization.
- 2.1 Comparison between integrative analysis against the non-integrative analysis.
- 2.2 Identification of cell subpopulations and cluster markers.
- 3 Phenotypically characterize the cellular subpopulations using cell-cell communication understanding.

- 3.1  Ligand-receptor interaction through the analysis of the different clusters obtained.
- 4 Evaluate the contribution of the identified cellular subpopulations to 5-FU treatment tolerance and tumor recurrence.

### **1.3. Impact on sustainability, ethical-social and diversity**

Here, we describe the points related to the guide of sustainability, ethical-social and diversity of UOC[8], that this project considers:

#### **(1) Sustainability:**

- a. Environmental:
  - i. Reduce waste: More effective treatment can decrease medical waste from ineffective therapies.
  - ii. Resource Efficiency: Better protocols lead to more efficient use of medical resources.
- b. 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.
- c. Social:
  - i. Improved Quality of Life: Effective management enhances patients' quality of life and social contributions.
  - ii. Increased Productivity: Reduced disease burden allows patients to remain active in the workforce.

#### **(2) Ethical-Social Considerations:**

- a. Ethical Research:
  - i. Transparency: Maintaining clear, reproducible research methods.
- b. Social Responsibility:

- i. Awareness: Promoting CRC awareness and early detection.
- ii. Community Engagement: Addressing community needs and ensuring research benefits diverse populations.

(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.
  - ii. Support: Providing opportunities for underrepresented groups in science and medicine.

To conclude, this project aims to improve CRC treatment outcomes and contribute to sustainability, ethical practice, and diversity healthcare, promoting a more equitable and responsible medical and care environment.

#### **1.4. Approach and method followed**

As mentioned above, this project has been done on the collaboration between the IGF and CRCL, to understand the CRC cell plasticity. Based on this context, an experiment was performed on a CRC cancer cell line (HCT116), which were treated with 5-FU, and the evolution of the cells was observed until day 20. Furthermore, RNA from the colonies was extracted at different time points and sequenced on single-cell, specifically at day 0, 2, 7, and 14. So, the data is a longitudinal dataset of scRNAseq, from before the treatment until the last day of the culturing, to obtain a map of the populations of the different cell lines during the 5-FU treatment on CRC cells.

Preliminary results from the CRCL have demonstrated that 5-FU therapy induces increased translational activity during the tolerance phase, which cells no longer divide. Surprisingly, they've also shown that this protein synthesis

correlates with the emergence of several tumor subpopulations, including cells exhibiting characteristics of pluripotent cells.

Based on this preliminary results and dataset, we performed a scRNAseq analysis throughout the Seurat v5 package [7] on R-studio [9], where two different studies were performed; (1) one script was designed for analyzing scRNAseq data by time points, and (2) another script was planned for making and integrative analysis of all datasets. The analysis divided by different time points must be considered on account of the remarkable preliminary results mentioned, since it could be useful to distinguish between the different phases that cell subpopulations experience and understand how cell subpopulations and their proportions evolve during the therapy.

On the other hand, the integrative analysis could be key for the understanding of the interactions between the cluster and cell subpopulations and cell-cell interactions on a CRC treated with 5-FU environment, what could help to understand the mechanisms and key cell-cell and cell subpopulations interactions during on a possible 5-FU tolerance, which remains misunderstood.

Both approaches could complement each other to comprehend the under covered communicative web between the tumoral cells and the phenotypic change that they develop during the therapy, all making a big step on a possible explanation for the cell plasticity of the tumoral cells, specifically of CRC, exposed to a 5-FU treatment.

Moreover, for each analysis we perform two different clusters methodologies; (1) clustering based on Seurat algorithms, and (2) throughout the identification of Cell Cycle of each cell, using Scina [10], Cancer Cell Markers are clustered and identified based on CancerSEA dataset [11].

After the classification of the different clusters, we performed an analysis of ligand-receptors using the database for intra- and intercellular signaling knowledge, OmniPath [12], by means of the use of LIANA package[13].

Finally, once the significant markers and cell-cell interactions are addressed, a biological analysis based on the results of the study is 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 Figure1, 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.

- By timepoints analysis
- Integrative analysis

#### **Translatome Analysis:**

Duration: 4 weeks

Task: Perform detailed analyses of the translatome of identified cellular subpopulations to understand treatment tolerance mechanisms, using appropriate methodologies.

#### **Cluster markers analysis and Phenotypic Characterization**

Duration: 5 weeks

Task: Identification of cell subpopulations and cluster markers.

Task: Final identification and characterization of Seurat Clusters. Phenotypically characterize cellular subpopulations through Cancer-Cell Markers.

#### **Biological description:**

Duration: 2 weeks

Task: Evaluate the contribution of the identified cellular subpopulations to 5-FU treatment tolerance and tumor recurrence. Studying ligand-receptor interaction of the different cell types or clusters.

### Preparation of the Manuscript:

Duration: 4 weeks.

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

#### 1.5.2. Callendar

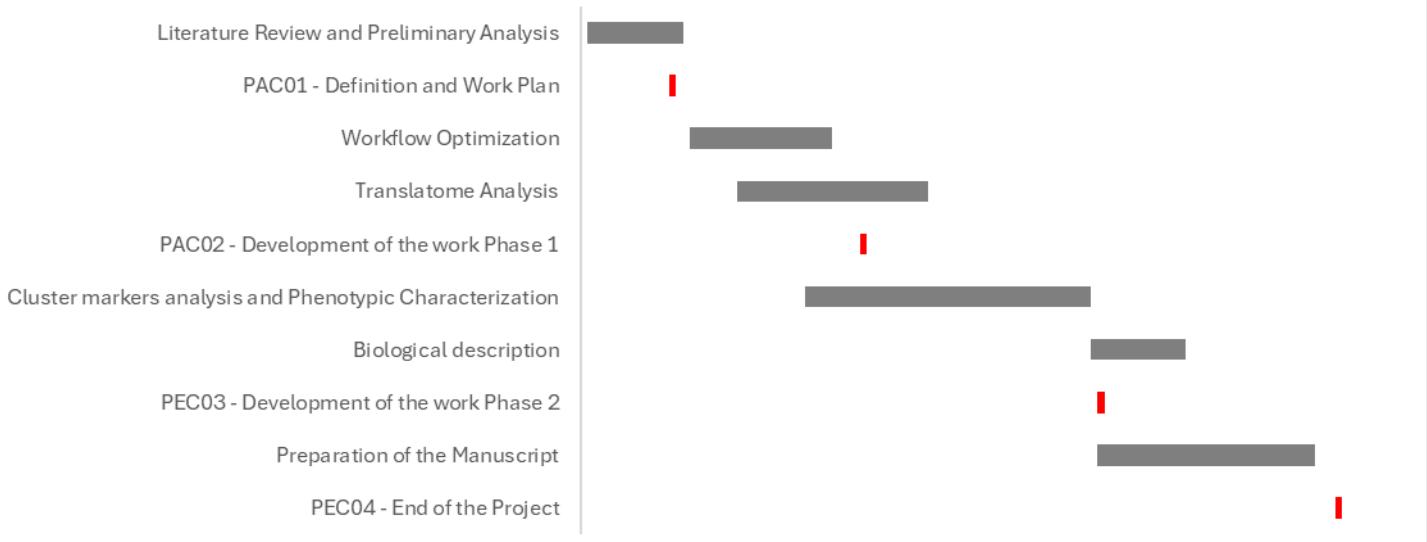
The time schedule for this project was defined to achieve all the objectives of the study.

**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
<b>PAC01 - Definition and Work Plan</b>	12/03/2024	1	12/03/2024
Workflow Optimization	15/03/2024	21	05/04/2024
Translatome Analysis	22/03/2024	28	19/04/2024
<b>PAC02 - Development of the work Phase 1</b>	09/04/2024	1	09/04/2024
Cluster markers analysis and Phenotypic Characterization	01/04/2024	42	13/05/2024
Biological description	13/05/2024	14	27/05/2024
<b>PEC03 - Development of the work Phase 2</b>	14/05/2024	1	14/05/2024
Preparation of the Manuscript	14/05/2024	32	15/06/2024
<b>PEC04 - End of the Project</b>	18/06/2024	1	18/06/2024

Gantt Diagram of the Workflow

28/02/2024 19/03/2024 08/04/2024 28/04/2024 18/05/2024 07/06/2024



**Figure 1.** 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.
  - 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.
- Ligand-Receptor Analysis for the by time points analysis and Integrative:
  - o Table of the top 50 most concordant interactions for each Time point and Integrative analysis.

## **- MILO RESULTS**

## **1.7. Brief description of the other chapters of the report**

2. Material and methods: In this chapter, the explanation in deep of the analysis performed on R-studio and all the methodology followed. Inside this chapter, it is subdivided into different subcategories:
  - 2.1. By time points analysis: In this part, the analysis of the data from the different time points is described.
    - 2.1.1. Quality Control and normalization of Raw matrix: In this part, the initial analysis of the scRNAseq is described, where the quality control of the data is done.

2.1.2. Dimensionality reduction: In this second part of the analysis of the scRNAseq is explained. To analyze scRNAseq data, it is needed to reduce the dimensionality of the data.

2.1.3. Cluster the cells and UMAP: After dimensionality reduction, the clusters of each matrix are identified and all their potentially cell markers are described.

2.1.4. Cell-cell interaction analysis:

2.2. Integrative analysis: After by time points analysis, we decided to perform an analysis that integrates all data by time points, in a one whole single data matrix, and the integrative methodology is described.

2.2.1. Perform integration:

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 are exposed. It is divided into different subsections:

3.1. Results by time points:

3.2. Results integrative analysis:

4. Conclusion and discussion: Here, the biological description of the results is explained, and the final conclusions of this project are clarified.

## 2. Material and Methods

As mentioned on the previous sections, the experiment consists of single-cell RNA-seq from cancer cell lines, from different time points, 0, 2, 7 and day 14. ON the time point 0, no 5-FU CDC cancer treatment was applied, but on day 2, the 5-FU treatment was started.

To study the cell population of this experiment, we performed an analysis based on Seurat v5 package on R. Other packages were also relevant during the study, like ggplot2 or SCINA.

Moreover, we performed two different approaches, one based on the study by timepoints, which consist in the evolution of the cell population during the therapy and seeing the different cell populations proportion.

On the other hand, an integrative analysis was performed, which consists of the integration of the whole time points data on one single matrix.

### 2.1. By Time points analysis

#### 2.1.1. Quality Control and normalization of Raw matrix.

For each matrix we work on an object called Seurat object, which storage all the raw information about the experiment, such as meta data, assay, counts matrix, or project name, with a set of designed parameters.

To erase all “junk” data from our raw matrix, with Seurat we explored QC metrics to filter cells based on our criteria. The amount of mitochondrial and ribosomes genes was calculated. We selected a threshold for selecting the amount of mitochondrial, and ribosomes genes, and the nCount and nFeatures from each matrix, with the help of “violin plot”. Also, the “FeatureScatter” from all the values mentioned above was plotted, to review the possible thresholds to apply, referred as Quality Control plots (Supplementary data).

Once the undesired cells from the datasets is removed, the normalization and scaling from each matrix is performed. The normalization method employed is a global-scaling normalization method “LogNormalize” which normalizes the feature expression measurements from each cell by the total expression, multiplies this by a scale factor, and log-transforms the result. Then, for future

downstream analysis, like PCA, a dataset of high cell-to-cell variation is calculated with 5000 features per dataset.

Next, we performed a scale based on a linear transformation in order to dimensional reduction the data, and then data was re-scaled by cell cycle difference (S/G2M vs G1), and mitochondrial gene expression.

### 2.1.2. Dimensionality reduction.

Once matrix of each time point is processed, normalized, and scaled, the Principal Component Analysis (PCA) on the scaled data. To determine the dimensionality of the dataset, an Elbow plot, and a Dimensional reduction heatmap (attached on **Supplementary data**) of each Time point is done. Identifying the true dimensionality of a dataset can be challenging, so we used multiple approaches.

### 2.1.3. Cluster the cells and UMAP.

The clusters are obtained based on the selected dimensions. The identified clusters of cells are measured by a shared nearest neighbor (SNN) modularity optimization based clustering algorithm [14]. For the purposes of visualizing the different clusters, thanks to an algorithm that takes a high-dimensional dataset and plots it in a figure called **UMAP**, which is a non-linear dimensional reduction technique, used to explore the datasets. The main objective of these algorithms is to understand the underlying structure in the dataset, to place similar cells together in low-dimensional space. **Even** though, some authors find 2D visualization techniques like UMAP as a valuable tool for exploring datasets, it has some limitations, so we decided to avoid drawing biological conclusions solely based on visualization techniques. For each cluster, the most differentially expressed genes are found and saved on a top 50 table, which are useful for further identification of each Cell Type for their potentially cell marker function.

Also, the Cell Cycle UMAP with cluster by each cell cycle is done where S, G1 and G2M phases are identified. Then, SCINA package is used for the identification of different Cancer Cell Markers signatures as different clusters.

#### 2.1.4. Cell-cell interaction analysis

## 2.2. Integrative Analysis

Integration of single-cell datasets, like across experimental batches, donors, or conditions, is a really important step in scRNA-seq workflows. Integrative analysis helps to match shared cell types and states across datasets, which can boost statistical power, and facilitate accurate comparative analysis across datasets. In our project, the experimental batch is going to be the different time points datasets (day 0, 2, 7 and 14).

### 2.2.1. Perform integration and cluster identification

Before starting the Integrative analysis by Seurat, re-scaled seurat objects by time points from the previous analysis are merged in a one single object, so the Quality Control and low-quality cells are already removed, so is not needed.

One important feature of Seurat v5 is that allows to store data in layers. These layers can store raw, un-normalized counts, normalized data, or z-scored/variance-stabilized data. In our data, we have 12 layers (a counts layer and a scale data layer for each time point).

So, once data is uploaded, the normalization and scaling data from the count's matrix is performed. We can visualize the results of a standard analysis without integration. The resulting clusters are defined by cell types and stimulation conditions.

Then, we aim to integrate data from the conditions, so that cells from the same cell type/subpopulations will cluster together. Our main goal in the integration is not to remove biological differences across conditions, but to learn sheared cell types/states. The integration method used for our project is the  Harmony method, a well established method.

After scRNA-seq data integration using Harmony method, we identified the clusters of the integrative dataset and UMAPs and DimPlots are plotted.

### 2.2.2. Identify conserved cell type markers

To identify canonical cell type markers genes that are conserved across time points, we perform the FindAllMarkers() in order to find the genes that are different between the identified clusters.

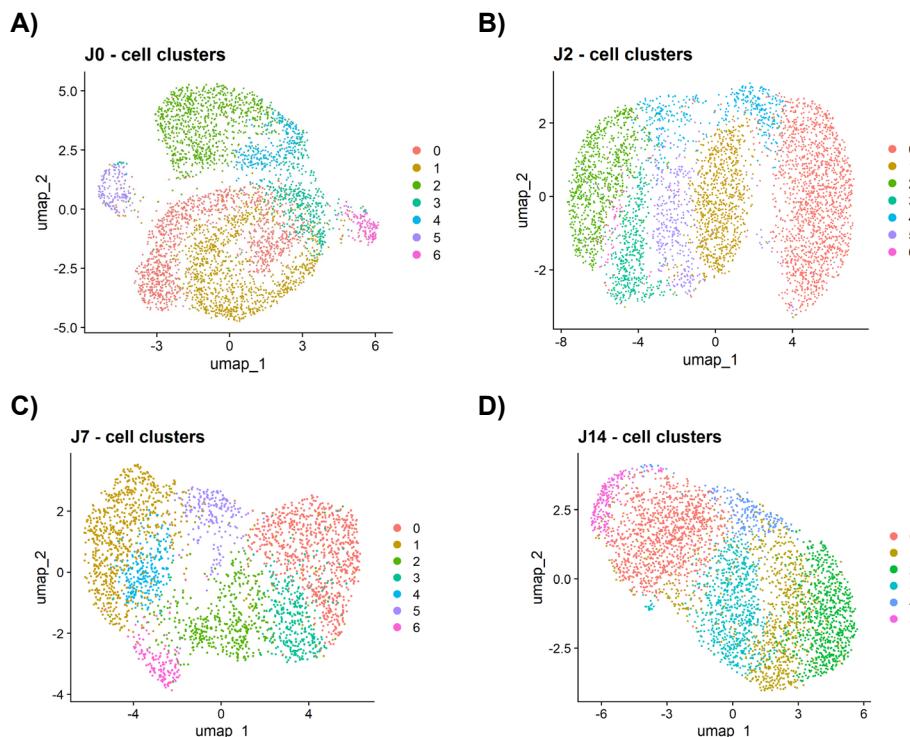
Also, CancerSEA markers are identified for the integrative analysis, and UMAPs.

### 2.2.3. Cell-cell interaction analysis

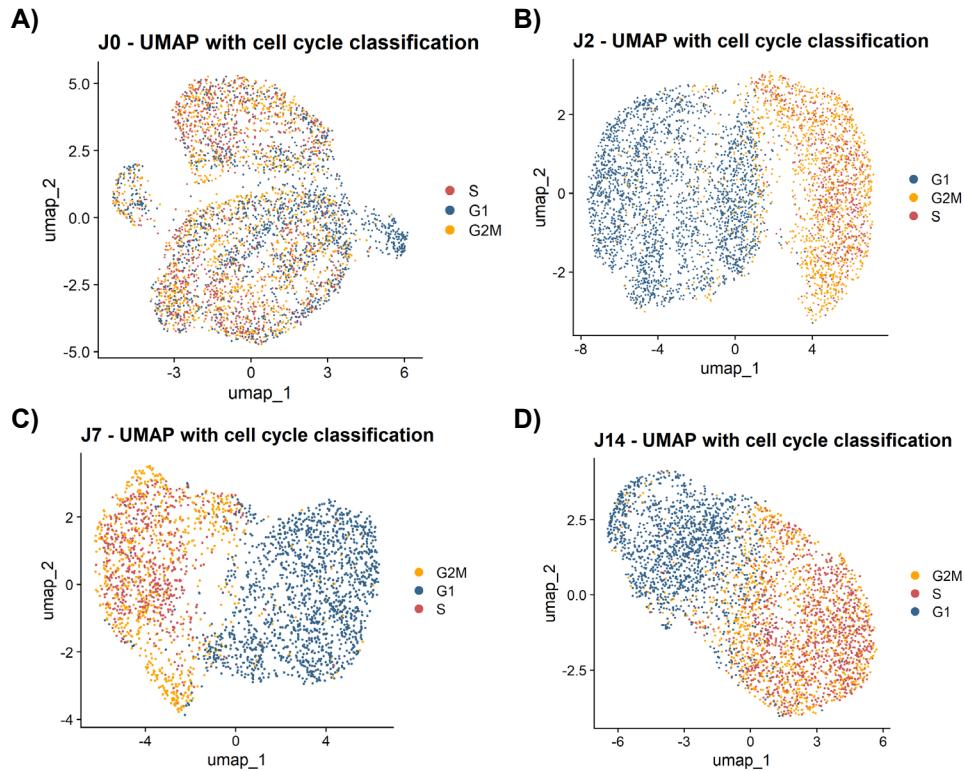
### 3. Results

#### 3.1. Analysis by time points

Based on the different time points (day 0, 2, 7 and 14), we obtained different number of clusters on each time point. For the day 0, 2 and 7 we obtained 6 clusters, even as it is seen in the Figure 2 the clusters are spatially different. On the Figure 3, there are less cells on G2M phase than the other days, and at day 14 is when there are more cells on S phase.



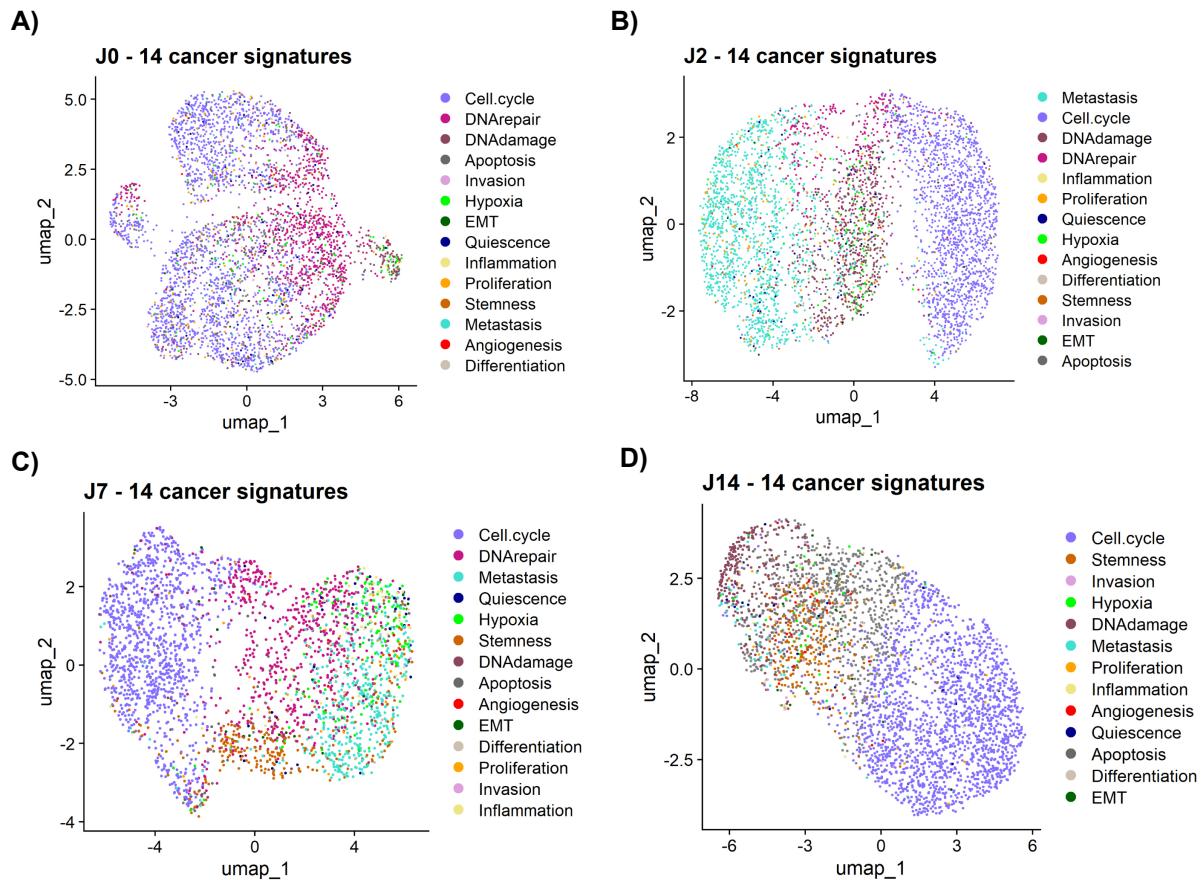
**Figure 2.** UMAP from scRNASeq analysis by time points. Clustering, defined by Seurat v5 Algorithm **A)** Represents the initial time point of the experiment; before the 5-FU treatment, 6 different clusters are identified. **B)** Time point 2 of the experiment: HCT116 cells after the 5-FU treatment,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



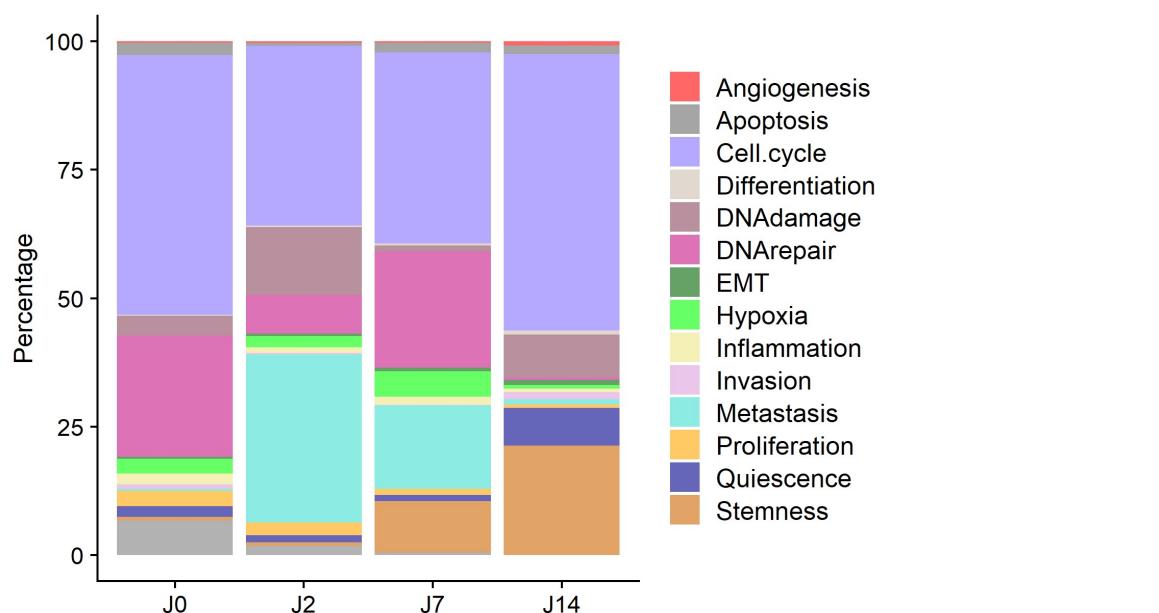
**Figure 3.** UMAP from scRNAseq analysis by time points. Clustering defined by Cell Cycle Ranked, G2M, G1 and S phase are labeled. **A)** Initial time point of the experiment, before the 5-FU treatment. **B)** Time point 2 of the experiment, HCT116 cells after the 5-FU treatment. **C)** 7 days of experiment. **D)** Day 14 of the experiment.

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

For the CancerSEA Cell markers identified, we can see that metastasis cells are less abundant on the day 0 and 14 not as seen the days 2 and 7. Cells on stemness state are not present at days 0, 2, and at days 7 and 14 the number of stemness cells increases. Moreover, we identified that cells on DNA repair weren't present for the day 14 (Figure 4 & 5).



**Figure 4.** UMAP from scRNAseq analysis by time points. Clustering is defined by the CancerSEA cancer cell marker dataset. **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.

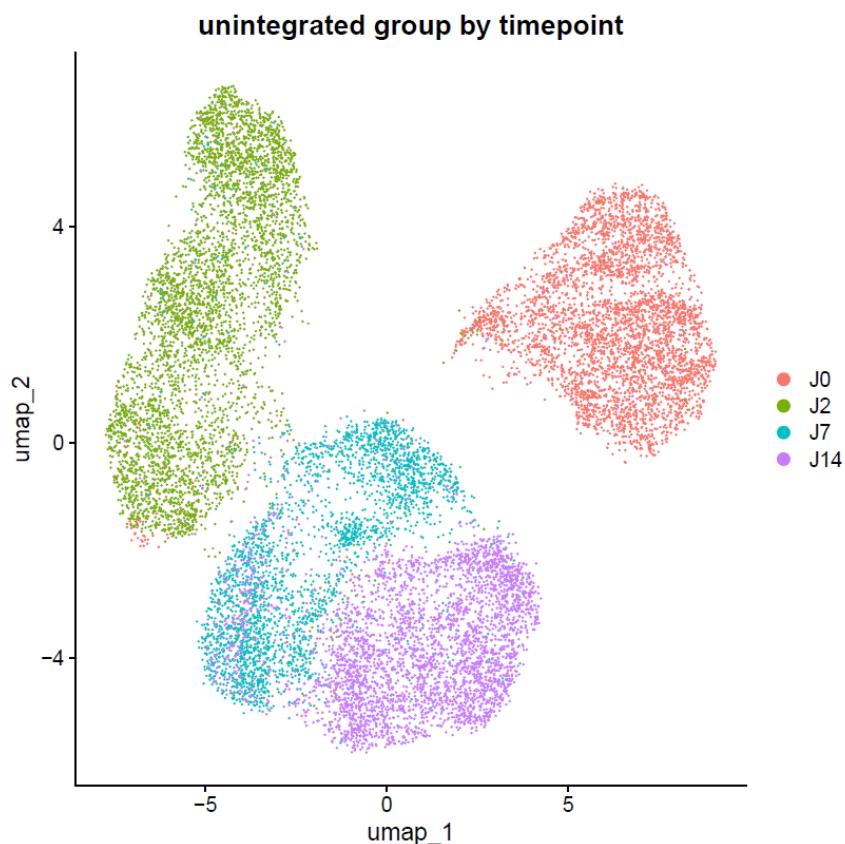


**Figure 5.** Proportion Plot of the cell populations at each time point based on the Cancer Cell Markers dataset, CancerSEA.

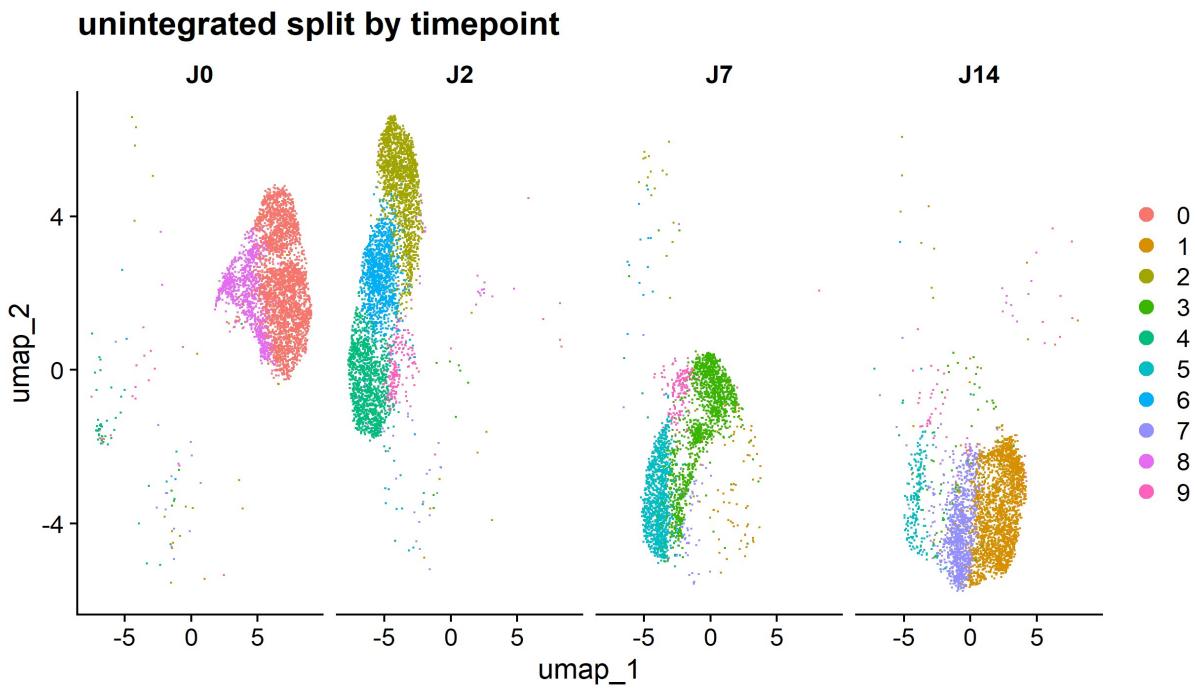
## LIANA RESULTS

### 3.2. Integrative Analysis

The unintegrated UMAP is seen in the Figure 6, and all cells are grouped based on the days. On this map, it can clearly be seen the batch effect of the different days. Then, clusters are identified without the use of integrating methodology, and 2d represented as in the Figure 7, also the UMAP is split into different time points, in the manner that the clusters from the day 0 are represented, clusters from day 2, and so on.



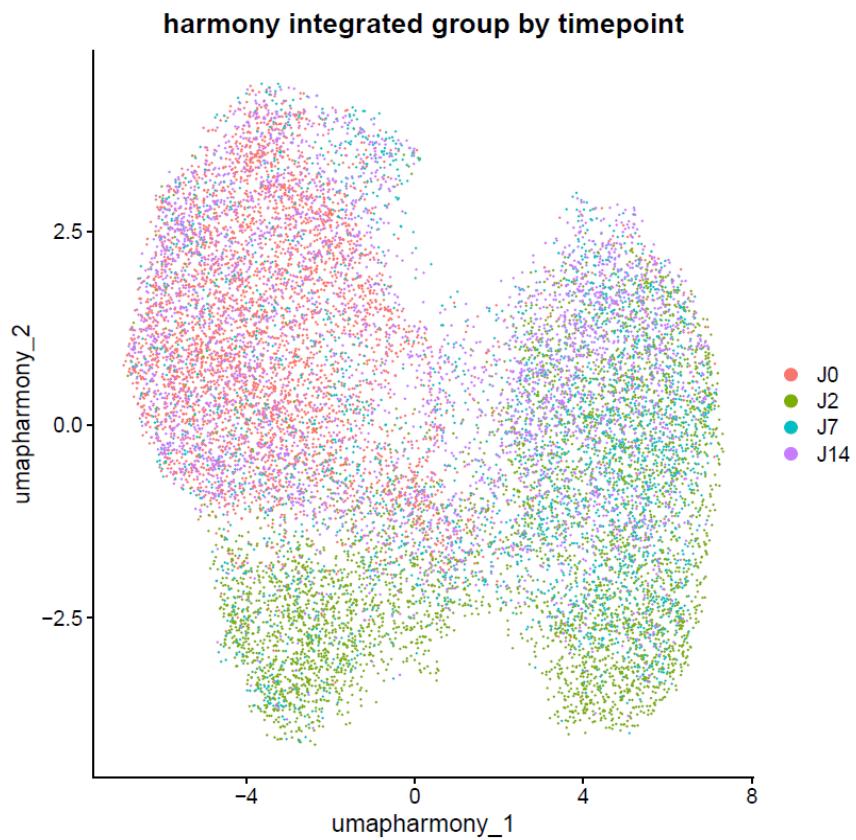
**Figure 6.** UMAP merging all Time points, representing the unintegrated analysis. Each timepoint is represented in different colors as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).



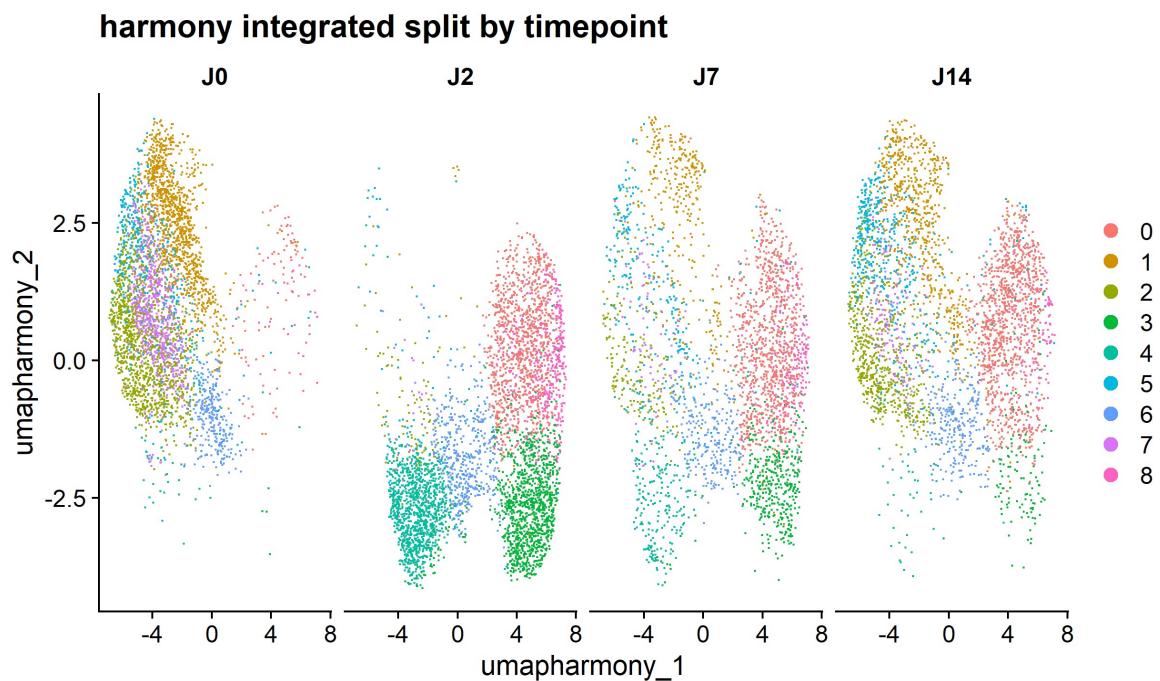
**Figure 7.** UMAP merging all data times matrices without integrating methodology, represented split by days (J0, J2, J7 and J17), a total of 9 different Clusters are identified.

Then, we performed the Harmony integrated method to avoid the batch effect of the different cell time points. As seen In the Figure 8, we can see that batch effect has been corrected, and cells from different time points merge spatially. Then, we identified the clusters based on this final correction, and 8 different clusters are recognized (Figure 9). For more accessible seeing of the clusters, we made the plots split by the different days, from day 0 to 14. Finally, based on the CancerSEA table, we distinguished the Cell Markers separately by the different days.

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



**Figure 8.** UMAP merging all Time points, representing the integrated analysis. Each timepoint is represented in different colors as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).



**Figure 9.**

### J0 - 14 cancer signatures

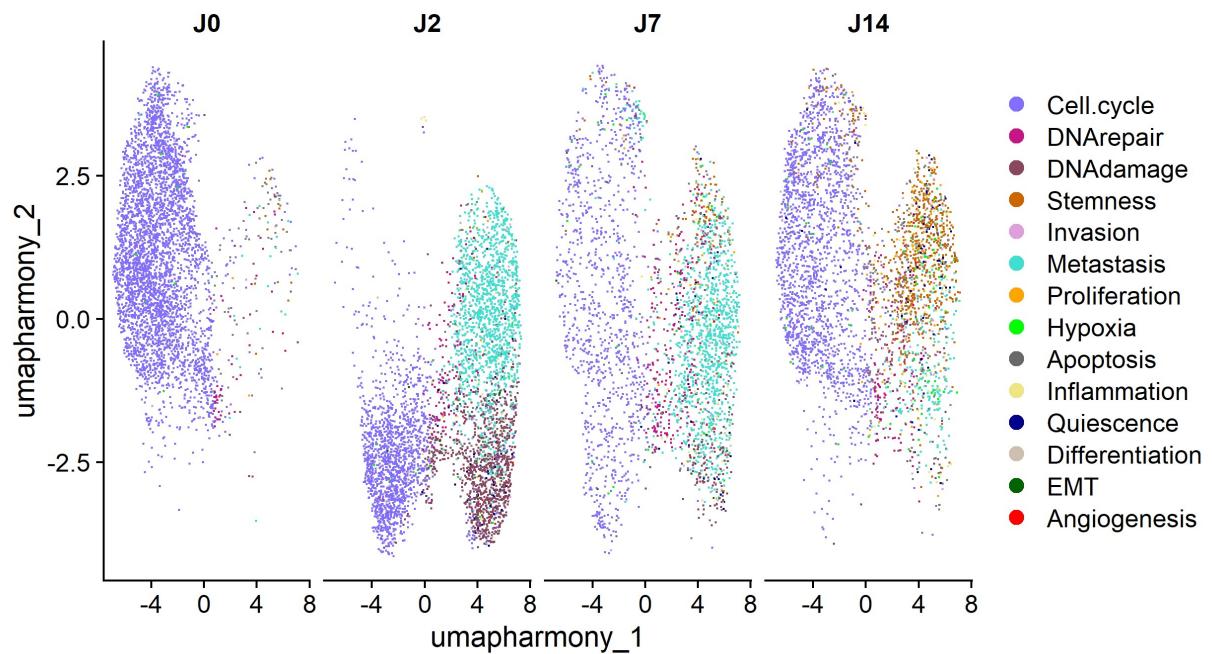


Figure 10.

Finally, the LIANA package is used to identify cell-cell interactions.

## 4. Conclusion and Discussion

Aquest capítol ha d'incloure:

- Una descripció de les conclusions del treball:
  - Un cop s'han obtingut els resultats quines conclusions s'estreu?
  - Aquests resultats són els esperats? O han estat sorprenents? Per què?
- Una reflexió crítica sobre l'assoliment dels objectius plantejats inicialment:
  - Hem assolit tots els objectius? Si la resposta és negativa, per quin motiu?
- Una anàlisi crítica del seguiment de la planificació i metodologia al llarg del producte:
  - S'ha seguit la planificació?
  - La metodologia prevista ha estat prou adequada?
  - Ha calgut introduir canvis per garantir l'èxit del treball? Per què?
- Dels impactes previstos a 1.3 (ètic-socials, de sostenibilitat i de diversitat), avaluar/esmentar si s'han mitigat (si eren negatius) o si s'han aconseguit (si eren positius).
- Si han aparegut impactes no previstos a 1.3, avaluar/esmentar com s'han mitigat (si eren negatius) o què han aportat (si eren positius).
- Les línies de treball futur que no s'han pogut explorar en aquest treball i han quedat pendents.

## 5. Glossari

Definició dels termes i acrònims més rellevants utilitzats dins la Memòria.

## 6. Bibliografia

- [1] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.
- [2] Y.-H. Cho *et al.*, "5-FU promotes stemness of colorectal cancer via p53-mediated WNT/β-catenin pathway activation," *Nat. Commun.*, vol. 11, no. 1, p. 5321, Oct. 2020, doi: 10.1038/s41467-020-19173-2.
- [3] K. V. D. Jeught, H.-C. Xu, Y.-J. Li, X.-B. Lu, and G. Ji, "Drug resistance and new therapies in colorectal cancer," *World J. Gastroenterol.*, vol. 24, no. 34, pp. 3834–3848, Sep. 2018, doi: 10.3748/wjg.v24.i34.3834.
- [4] M. Elbadawy, T. Usui, H. Yamawaki, and K. Sasaki, "Development of an Experimental Model for Analyzing Drug Resistance in Colorectal Cancer," *Cancers*, vol. 10, no. 6, Art. no. 6, Jun. 2018, doi: 10.3390/cancers10060164.
- [5] M. Luo, X. Yang, H.-N. Chen, E. C. Nice, and C. Huang, "Drug resistance in colorectal cancer: An epigenetic overview," *Biochim. Biophys. Acta BBA - Rev. Cancer*, vol. 1876, no. 2, p. 188623, Dec. 2021, doi: 10.1016/j.bbcan.2021.188623.
- [6] S. E. Mohr, S. G. Tattikota, J. Xu, J. Zirin, Y. Hu, and N. Perrimon, "Methods and tools for spatial mapping of single-cell RNAseq clusters in *Drosophila*," *Genetics*, vol. 217, no. 4, p. iyab019, Apr. 2021, doi: 10.1093/genetics/iyab019.
- [7] Y. Hao *et al.*, "Dictionary learning for integrative, multimodal and scalable single-cell analysis," *Nat. Biotechnol.*, vol. 42, no. 2, pp. 293–304, Feb. 2024, doi: 10.1038/s41587-023-01767-y.
- [8] "Competències transversals - Compromís global - UOC." Accessed: Jun. 02, 2024. [Online]. Available: <https://www.uoc.edu/portal/ca/compromis-social/ciutadania-global/competencies-transversals/index.html>
- [9] "R: The R Project for Statistical Computing." Accessed: May 13, 2024. [Online]. Available: <https://www.r-project.org/>
- [10] "SCINA: Semi-supervised Category Identification and Assignment | single cell r package | cell clustering tools | deconvolution software." Accessed: May 13, 2024. [Online]. Available: <https://dbai.biohpc.swmed.edu/scina/>
- [11] H. Yuan *et al.*, "CancerSEA: a cancer single-cell state atlas," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D900–D908, Jan. 2019, doi: 10.1093/nar/gky939.
- [12] D. Türei *et al.*, "Integrated intra- and intercellular signaling knowledge for multicellular omics analysis," *Mol. Syst. Biol.*, vol. 17, no. 3, p. e9923, Mar. 2021, doi: 10.15252/msb.20209923.
- [13] D. Dimitrov *et al.*, "Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data," *Nat. Commun.*, vol. 13, no. 1, p. 3224, Jun. 2022, doi: 10.1038/s41467-022-30755-0.
- [14] L. Waltman and N. J. van Eck, "A smart local moving algorithm for large-scale modularity-based community detection," *Eur. Phys. J. B*, vol. 86, no. 11, p. 471, Nov. 2013, doi: 10.1140/epjb/e2013-40829-0.



## 7. Annex

Llistat d'apartats que són massa extensos per incloure dins la memòria i tenen un caràcter autocontingut (per exemple, manuals d'usuari, manuals d'instal·lació, etc.)

Depenent del tipus de treball, és possible que no calgui afegir cap annex.