

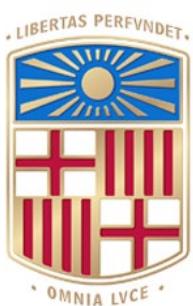
Impact of 5-FU treatment on tumor cell plasticity in colorectal cancer

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

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</i>
Nom del PRA:	Nom i dos cognoms
Data de lliurament (mm/aaaa):	06/2024
Titulació o programa:	Pla d'estudis
Àrea del Treball Final:	Area de TF
Idioma del treball:	<i>anglès</i>
Paraules clau	Màxim 3 paraules clau, validades pel tutor/a del TF
Resum del Treball	
Màxim 250 paraules, amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball	
Abstract	
A maximum of 250 words, detailing the purpose, context of application, methodology, results and conclusions of the work	

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	9
2.	Material and Methods.....	11
2.1.	By Time points analysis	11
2.1.1.	Quality Control and normalization of Raw matrix.	11
2.1.2.	Dimensionality reduction.....	12
2.1.3.	Cluster the cells and UMAP.	12
2.1.4.	Cell-cell interaction analysis.....	13
2.2.	Integrative Analysis	13
2.2.1.	Perform integration and cluster identification	13
2.2.2.	Identify conserved cell type markers.....	14
2.2.3.	Cell-cell interaction analysis.....	14
3.	Results	15
3.1.	Analysis by time points.....	15

3.1.1.	Day 0 Cell-cell interactions	19
3.1.2.	Day 2 Cell-cell interactions	22
3.1.3.	Day 7 Cell-cell interactions	24
3.1.4.	Day 14 Cell-cell interactions.....	27
3.2.	Integrative Analysis	30
3.2.1.	Cell-cell interactions	34
4.	Conclusion and Discussion	37
5.	Glossari	38
6.	Bibliografia	39
7.	Annex	41

List of Figures

Figure 1. Gantt Diagram of the followed timeline during the project.....	8
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.....	15
Figure 3 Proportion Plot of the cells proportion at each time point based on the Seurat clustering.....	16
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.....	17
Figure 5. Proportion Plot of the cell populations at each time point based Cell cycle markers	17
Figure 6. 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.....	18
Figure 7. Proportion Plot of the cell populations at each time point based on the Cancer Cell Markers dataset, CancerSEA.	19
Figure 8	19
Figure 9. Most relevant cell-cell interactions for Day 0 on Seurat clustering	20
Figure 10	21
Figure 11. Most relevant cell-cell interactions for Day 0 on CancerSEA Cancer Cell Markers.	22
Figure 12	22

Figure 13. Most relevant cell-cell interactions for Day 2 on Seurat clustering ..	23
Figure 14	23
Figure 15. Most relevant cell-cell interactions for Day 2 on CancerSEA Cancer Cell Markers.....	24
Figure 16	25
Figure 17. Most relevant cell-cell interactions for Day 7 on Seurat clustering ..	26
Figure 18	26
Figure 19. Most relevant cell-cell interactions for Day 7 on CancerSEA Cancer Cell Markers.....	27
Figure 20	28
Figure 21. Most relevant cell-cell interactions for Day 14 on Seurat clustering	29
Figure 22	29
Figure 23. Most relevant cell-cell interactions for Day 14 on CancerSEA Cancer Cell Markers.....	30
Figure 24. 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).	31
Figure 25. 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.....	32
Figure 26. 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).....	32
Figure 27. 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.....	33
Figure 28. UMAP merging all data times matrices using Harmony as Integration methodology, represented split by days (J0, J2, J7 and J17), the CancerSEA Cancer Cell Markers were represented.....	33

Figure 29	34
Figure 30. Most relevant cell-cell interactions for the integrative analysis on Seurat clustering.	35
Figure 31	35
Figure 32. Most relevant cell-cell interactions for the integrative analysis on CancerSEA Cancer Cell Markers.	36

List of Tables

Table 1. Deadlines and duration of the project planification.	7
Table 2 Top 10 cell-cell interactions for Day 0 on Seurat Clusters (0-6).	20
Table 3 Top 10 cell-cell interactions for Day 0 CancerSEA Cancer Cell Markers.	21
Table 4 Top 10 cell-cell interactions for Day 2 Seurat Clusters (0-6).	22
Table 5 Top 10 cell-cell interactions for Day 2 CancerSEA Cancer Cell Markers.	24
Table 6 Top 10 cell-cell interactions for Day 7 Seurat Clusters (0-6).	25
Table 7 Top 10 cell-cell interactions for Day 7 CancerSEA Cancer Cell Markers.	27
Table 8 Top 10 cell-cell interactions for Day 14 Seurat Clusters (0-5).	28
Table 9 Top 10 cell-cell interactions for Day 14 CancerSEA Cancer Cell Markers.	30
Table 10 Integrative analysis top 10 cell-cell interactions Seurat Clusters	34
Table 11 Integrative analysis top 10 cell-cell interactions of CancerSEA Cancer Cell Markers.	36

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 significantly contributes to the low survival rates observed in CRC patients [1]. The primary treatment for CRC, 5-Fluorouracil (5-FU)-based chemotherapy, is often associated with resistance, leading to tumor recurrence. Despite the advancements in various modulation strategies, approximately half of CRC patients experience recurrence within 5 years of their treatment. Tumor cells develop tolerance to anticancer treatments, including 5-FU, which promotes recurrence and metastasis [2], [3].

Cancer cells are heterogeneous in morphology, inheritance, and functions. Within this heterogeneity, cancer stem cells (CSCs) are a small population of cells with the ability to self-renew and regenerate tumors. CSCs are believed to play a crucial role in regulating mechanisms of intrinsic or acquired drug resistance [4]; thus enhancing tumor relapse post-chemotherapy. Despite advancements in understanding the mechanism underlying drug tolerance in cancer remain poorly understood. Therefore, elucidating these mechanisms is essential for improving CRC patient outcomes [2].

Additionally, transcriptomic, and epigenetic studies have highlighted cellular plasticity as a hallmark of treatment tolerance. Epigenetic modifications facilitate drug efflux and regulate drug targets and associated signaling pathways. However, the translation of messenger RNA into proteins within the context of tumor drug resistance remains largely unexplored [5].

In recent years, single-cell RNA sequencing (scRNA-seq) technology has revolutionized the field by providing powerful tools and techniques for defining molecular signatures of cell types and subtypes, as well as for data analysis, visualization, and mining of scRNA-seq datasets, exemplified by platforms like Seurat [6], [7]. These technologies are crucial for understanding transcriptomic and cell population dynamics. The scRNA-seq technology is invaluable, as it

enabling researchers to measure the expression levels of all genes across thousands to millions of individual cells [8].

This study is part of a project from the Institut de Génomique Fonctionnelle (IGF) of Montpellier and the Lyon Cancer Research Center (CRCL). The overarching objectives include demonstrating that translational control is a key molecular mechanism in cellular reprogramming in response to 5-FU, and specifically for this study, isolating and phenotypically characterizing relevant cellular subpopulations. By employing and optimizing the single-cell RNA-seq workflow using the Seurat package in R, this research aims to perform a detailed analysis of these subpopulations and cluster markers and characterizing these cells phenotypically through cell-cell communication studies. Ultimately, the goal is to evaluate the contribution of these subpopulations to 5-FU treatment tolerance and tumor recurrence.

1.2. Objectives

2.1.1. General objectives

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

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, this study will focus in achieving objectives 2 and 3.

2.1.2. 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 single-cell RNA-seq workflow using Seurat Package in R to isolate relevant cellular subpopulations.

2. Perform detailed analysis of these cell subpopulations using appropriate methodologies, facilitated by Seurat Package workflow and optimization.
 - 2.1 Compare integrative analysis with non-integrative analysis.
 - 2.2 Identify cell subpopulations and cluster markers.
- 3 Phenotypically characterize the cellular subpopulations using cell-cell communication.
- 3.1 Analyze ligand-receptor interactions within 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[9], 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 isolate relevant cellular subpopulations and phenotypically characterize the cellular subpopulations, in a longitudinal experiment of 5-FU therapy. Based on this context, an experiment was performed on a CRC cancer cell line (HCT116), which was treated with 5-FU, the standard treatment for CRC, and the evolution of cell populations was observed until day 20. RNA from the colonies was extracted at different time

points and sequenced on single-cell, specifically, the obtained data is from days 0, 2, 7, and 14. So, the data is a longitudinal dataset of scRNA-seq, from before the treatment until the last day of the culturing. This approach allows us to obtain a map of the evolution of different cell populations and proportions during the 5-FU treatment on CRC-like cells during a determined time.

Preliminary results from the CRCL have demonstrated that 5-FU therapy induces increased translational activity during the tolerance phase, in 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.

So, based on these preliminary results and dataset, we performed a scRNA-seq analysis throughout the Seurat v5 package [7] on R-studio [10], where two different studies were performed; (1) one script was designed for analyzing scRNA-seq data by time points separately, (2) another script was planned for making and integrative analysis of all dataset. 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, which could help to understand how cell plasticity works in terms of how cell populations environment work.

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-like cell which have been treated with 5-FU, which 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 toward a possible explanation for the cell plasticity of the tumoral cells, specifically of CRC, exposed to 5-FU treatment.

Moreover, for each analysis we perform two different cluster methodologies; (1) clustering based on Seurat algorithms, and (2) throughout the identification of the Cell Cycle of each cell, using Scina [11], Cancer Cell Markers are clustered and identified based on CancerSEA dataset [12].

After the classification of the different clusters, for both approaches, we performed an analysis of ligand-receptors using the database for intra- and intercellular signaling knowledge, OmniPath [13], employing the LIANA package[14].

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
PAC01 - Definition and Work Plan	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
PAC02 - Development of the work Phase 1	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
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
PEC04 - End of the Project	18/06/2024	1	18/06/2024

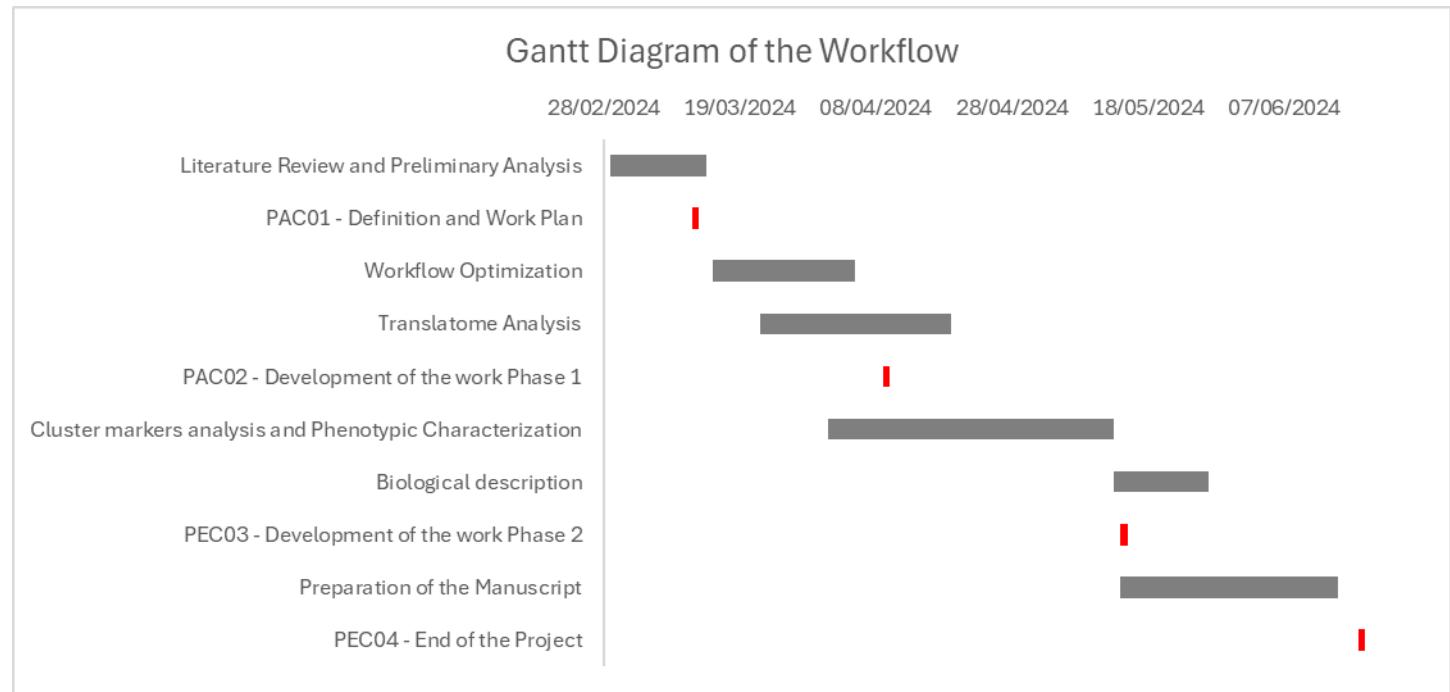


Figure 1. Gantt Diagram of the followed timeline during the project.

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.

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 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:

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.

2. Material and Methods

As mentioned in the previous sections, the experiment consists of single-cell RNA-seq cancer cell lines (HCT116), from different time points: 0, 2, 7, and day 14. At day 0, no 5-FU treatment was applied; treatment began on day 2. At each time point, RNA was extracted and sequenced for single-cell RNA analysis.

To study the single-cell data and cell population evolution from this experiment, we performed an analysis using the Seurat v5 package in R. Additional packages, such as `ggplot2` or SCINA, were also employed.

Two different approaches were taken: one involved analyzing the datasets separately by time points to observe cell population evolution and proportions during the longitudinal study. The other approach was an integrative analysis, combining all datasets into a single dataset with corrected batch effects.

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 removed “junk” data from our raw matrix. Using Seurat, we explored QC metrics to filter cells based on our criteria. The amounts of mitochondrial and ribosomal genes were also calculated. Based on QC plots, such as violin plots and FeaturesScatter plots, we selected thresholds for mitochondrial and ribosomal gene content, and the desired nCount and nFeatures from each matrix (Supplementary data).

Once the undesired cells 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, such as Principal Component Analysis (PCA), a dataset with high cell-to-cell variation was created with 5000 features per dataset.

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) and mitochondrial gene expression.

2.1.2. Dimensionality reduction.

Once each time point matrix was processed, normalized, and scaled. 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 (Supplementary data)

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 clustering [15]. For visualization, we used 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 techniques. 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.

Using the SCINA package, we identified Cancer Cell Marker signatures from the CancerSEA database and correlated them to Seurat clusters. This allowed us to add Differential Gene Expression (DGE) markers for each Seurat cluster and predict their cancer signatures. Based on these markers, we clustered cells from different time points and created a proportion plot for the longitudinal study.

2.1.4. Cell-cell interaction analysis

To study cell-cell communication (CCC), we used the LIANA framework, which allows for the use of various ligand-receptor methods. For the by-time point analysis, we performed cell-cell interaction analysis on each dataset for the two clustering methods: (1) Seurat Clusters and (2) CancerSEA Cancer Cell Markers. We identified the top 50 most concordant interactions using a consensus of all statistically significant methods.

2.2. Integrative Analysis

Integration of single-cell datasets across experimental batches, donors, or conditions is crucial in 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 re-scaled Seurat objects from each time points into a single object, ensuring Quality Control and removal of low-quality cells were already performed. Seurat v5 allows storing data in layers, which can hold raw, un-normalized 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 and stimulation conditions.

Our goal was to integrate data from the different conditions, so cells of the same type/supopulations would cluster together. The integration method used was Harmony, a well-established technique.

After integrating the scRNA-seq data using the Harmony method, we identified clusters in the integrative dataset and generated UMAPs and DimPlots.

2.2.2. Identify conserved cell type markers

To identify canonical cell type markers conserved genes across time points, we used the FindAllMarkers() function to find genes that differ between identified clusters.

Additionally, we identified CancerSEA markers 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 between cells for the two clusters: (1) Seurat Clusters and (2) CancerSEA Cancer Cell Markers, based on a consensus that integrates the predictions of individual methods. This is done by ranking and aggregating (RRA) the ligand-receptor interaction predictions from all methods.

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 we obtained 6 clusters were identified, although their localization differed on the 2D maps (Figure 2). 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 3).

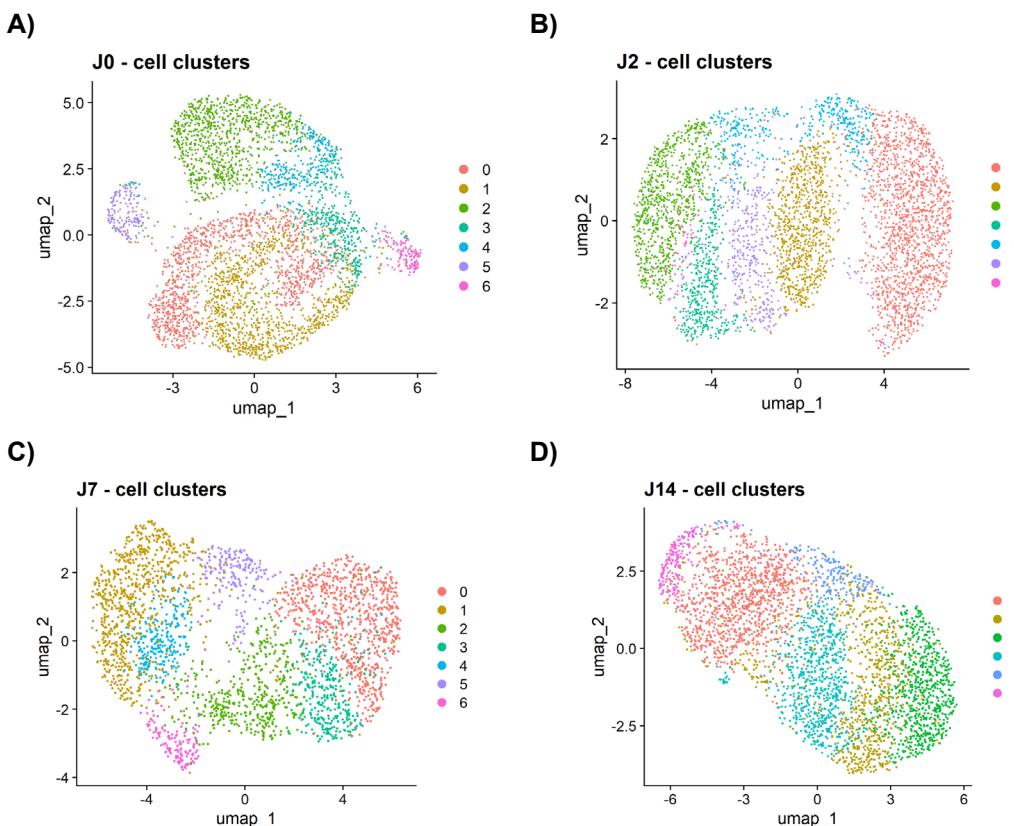


Figure 2. UMAP from scRNAseq analysis by time points. Clustering, defined by Seurat v5 Algorithm **A)** Represents the initial time point of the experiment; before the 5-FU treatment, 6 different clusters are identified. **B)** Time point 2 of the experiment: HCT116 cells after the 5-FU treatment, 6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

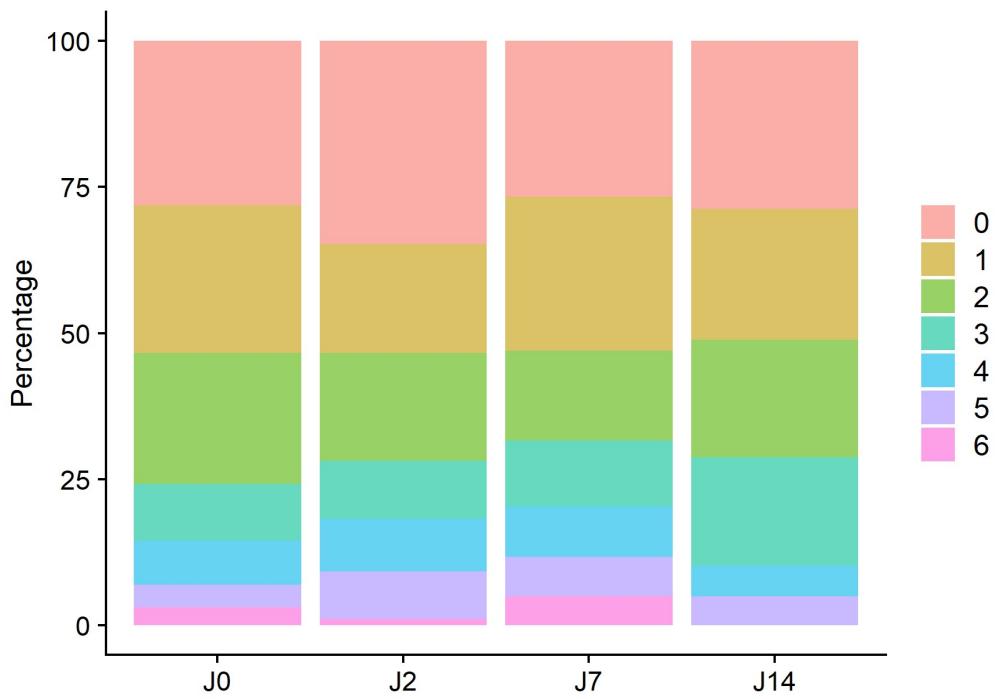


Figure 3 Proportion Plot of the cells proportion at each time point based on the Seurat clustering.

In Figure 4, we can see that on day 0, most cells, although in different cell cycling phases, are spatially represented in a similar 2D. In contrast, for the other time points, especially day 2, the spatial difference between cell cycles is clearly represented. 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 5).



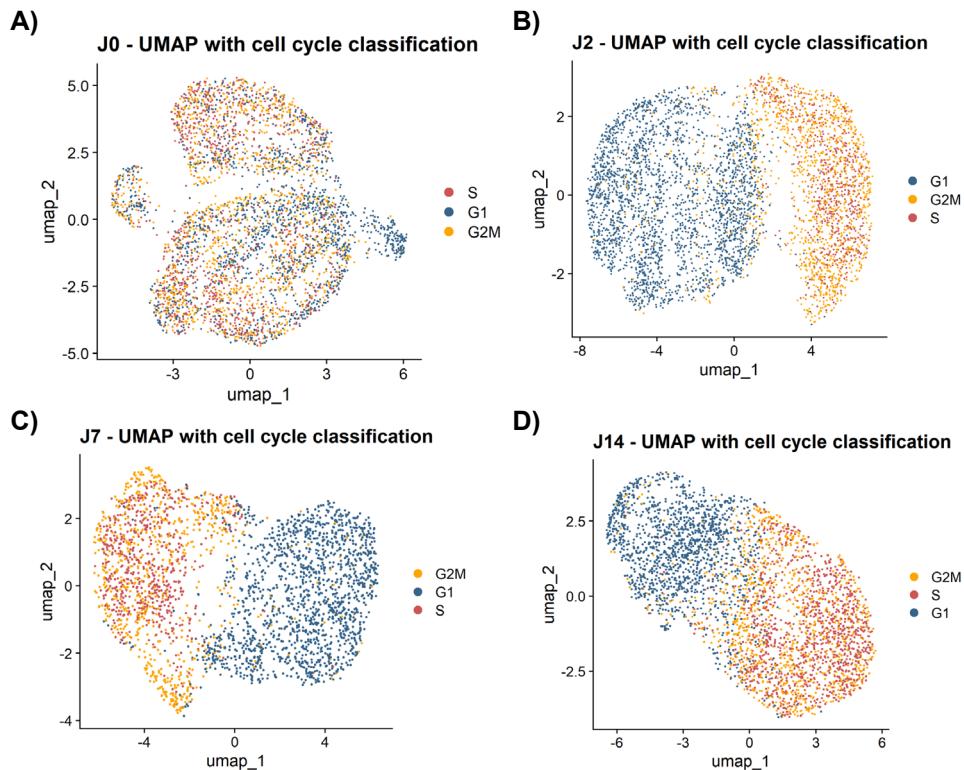


Figure 4. UMAP from scRNAseq analysis by time points. Clustering defined by Cell Cycle Ranked, G2M, G1 and S phase are labeled. **A)** Initial time point of the experiment, before the 5-FU treatment. **B)** Time point 2 of the experiment, HCT116 cells after the 5-FU treatment. **C)** 7 days of experiment.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

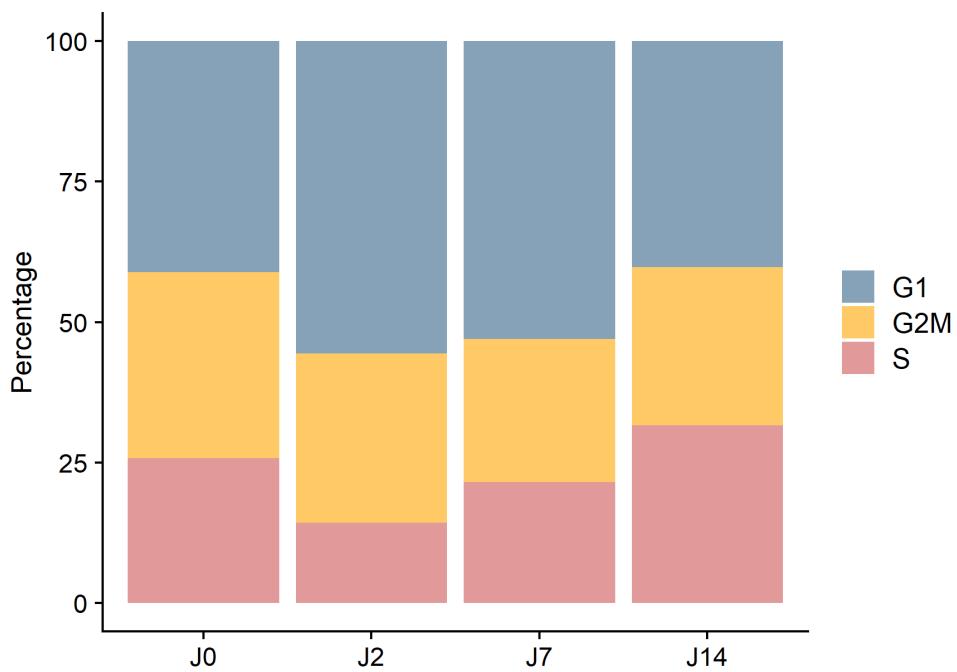


Figure 5. Proportion Plot of the cell populations at each time point based Cell cycle markers

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 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 6 and 7).

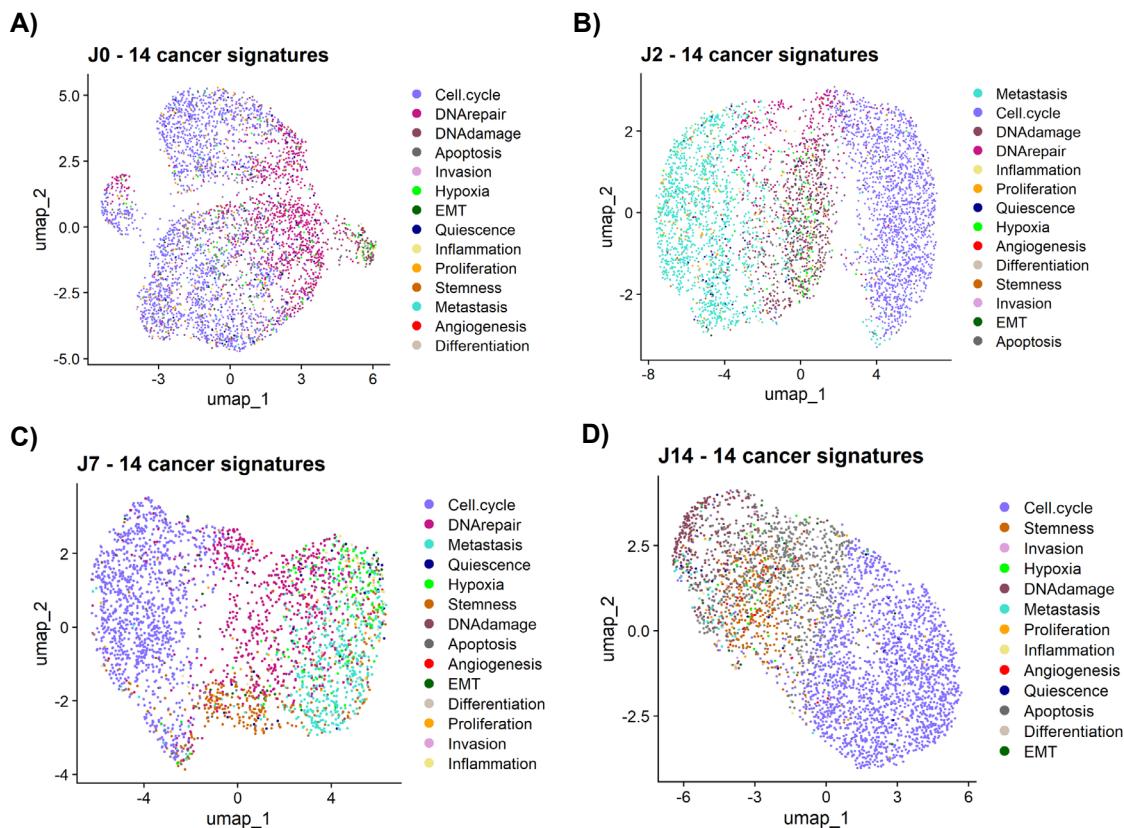


Figure 6. UMAP from scRNAseq analysis by time points. Clustering is defined by the CancerSEA cancer cell marker dataset. **A)** Initial time point of the experiment, before the 5-FU treatment. **B)** Time point 2 of the experiment: HCT116 cells after the 5-FU treatment. **C)** 7 days of experiment. **D)** Day 14 of the experiment.

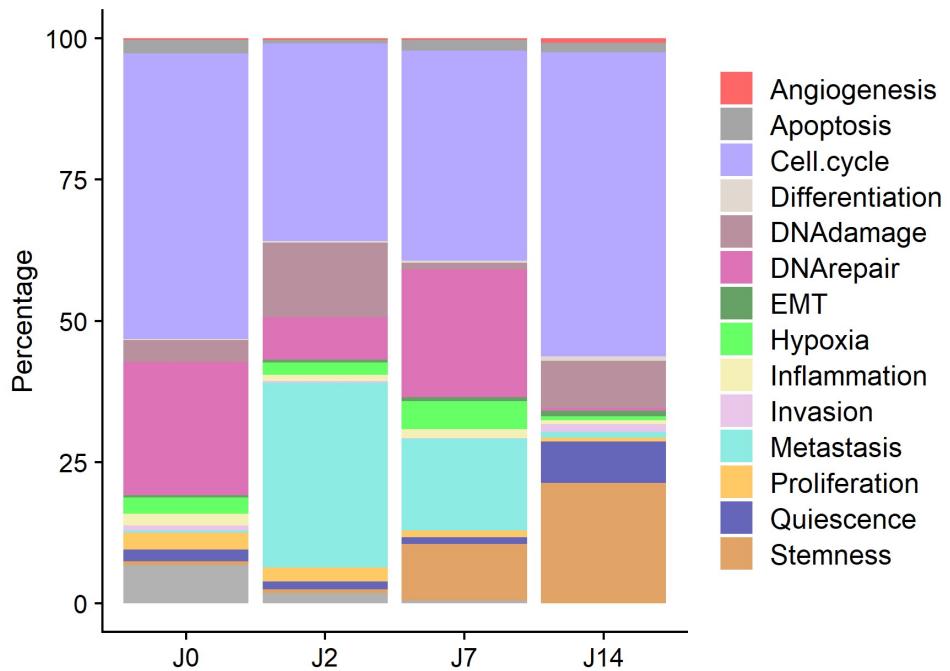


Figure 7. Proportion Plot of the cell populations at each time point based on the Cancer Cell Markers dataset, CancerSEA.

3.1.1. Day 0 Cell-cell interactions

Based on cell-cell interactions analysis, the most concurrent interactions between Seurat Clusters

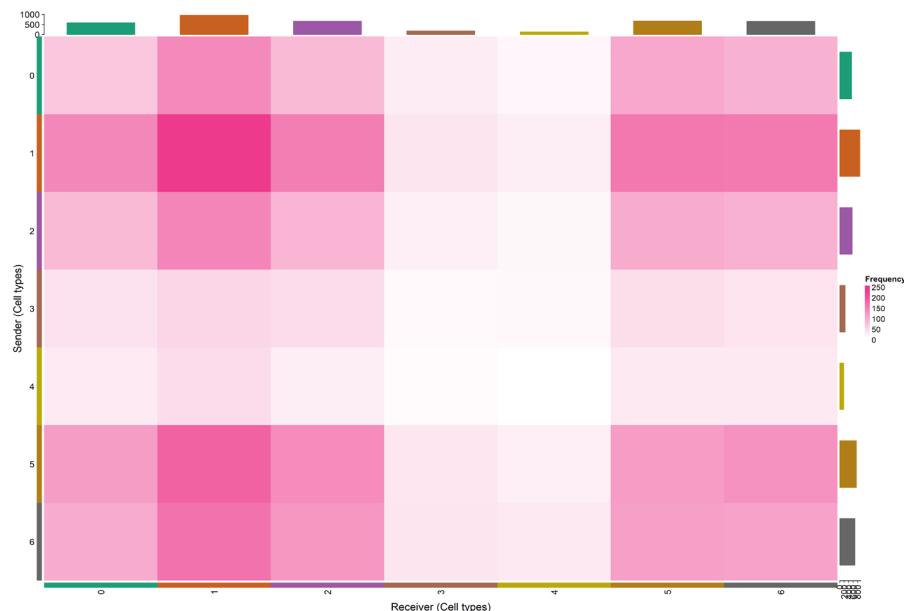


Figure 8

Table 2 Top 10 cell-cell interactions for Day 0 on Seurat Clusters (0-6).

source	target	ligand.complex	receptor.complex
3	2	LGALS1	ITGB1
5	2	SEMA3A	PLXNA1
5	1	SEMA3A	PLXNA3
3	5	LGALS1	ITGB1
3	1	MIF	CD44
5	6	SEMA3A	PLXNA3
5	5	SEMA3A	PLXNA1
2	1	VCAN	CD44
6	6	LAMA3	ITGB4
3	6	MIF	CD44

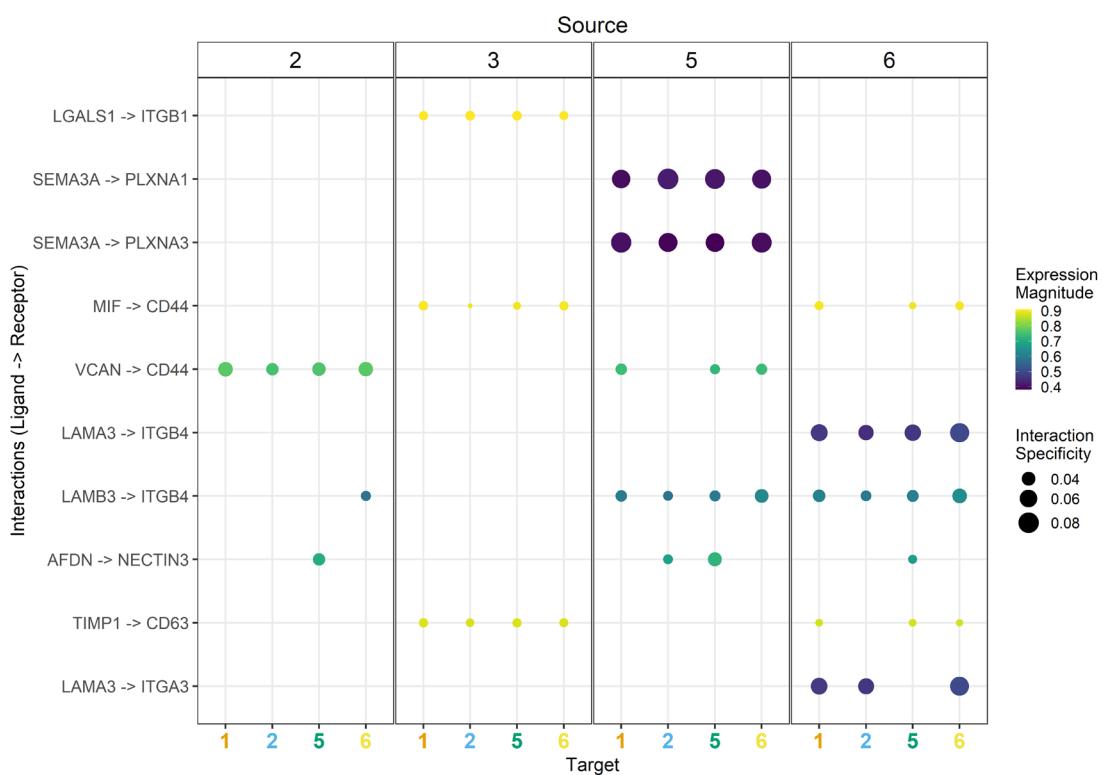


Figure 9. Most relevant cell-cell interactions for Day 0 on Seurat clustering

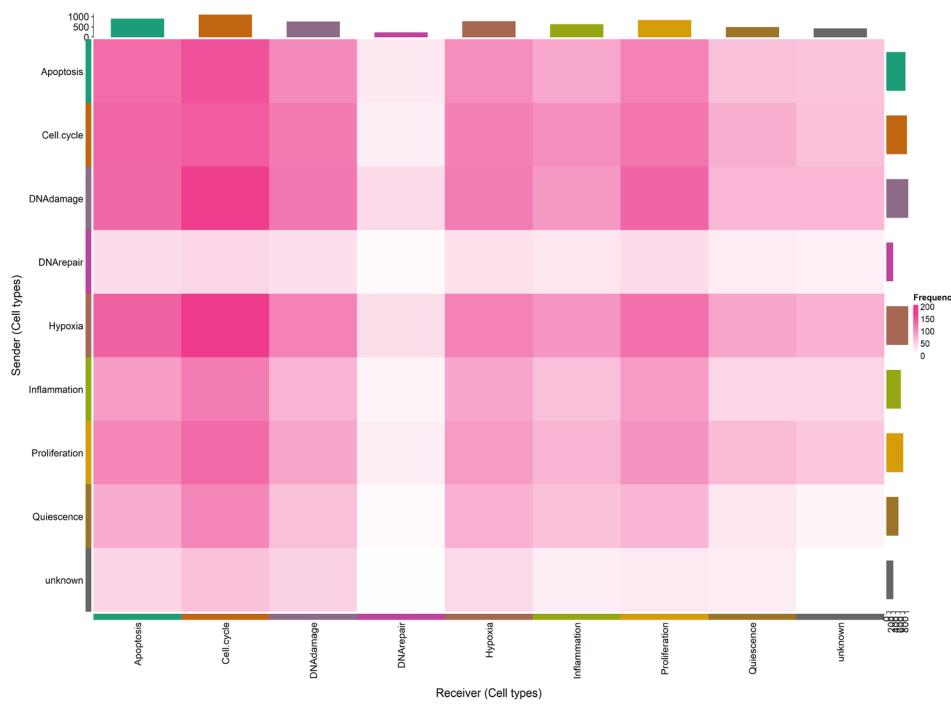


Figure 10

Table 3 Top 10 cell-cell interactions for Day 0 CancerSEA Cancer Cell Markers.

source	target	ligand.complex	receptor.complex
Inflammation	DNAdamage	CXCL8	SDC3
Inflammation	Cell.cycle	CXCL8	SDC3
Proliferation	Cell.cycle	FGF7	FGFR1
Hypoxia	Inflammation	LGALS1	ITGB1
DNAdamage	Quiescence	ALDH1A3	CRABP2_RXRA
DNAdamage	Quiescence	ALDH1A3	CRABP2_RARA
DNAdamage	Quiescence	ALDH1A3	CRABP2_RARG
Inflammation	unknown	CXCL8	SDC3
Hypoxia	Cell.cycle	LGALS1	ITGB1
DNArepair	Hypoxia	MIF	CD44

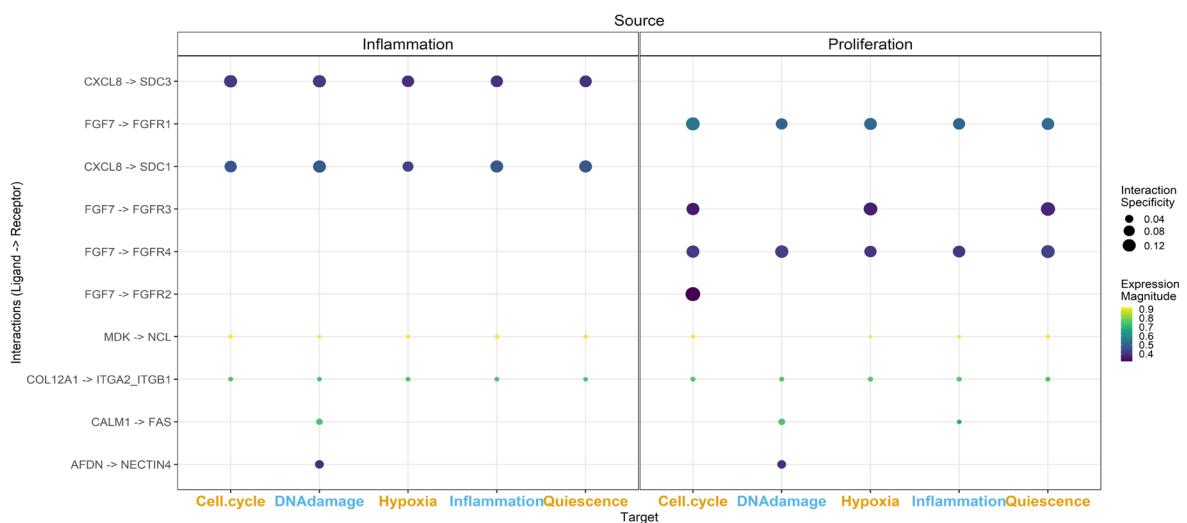


Figure 11. Most relevant cell-cell interactions for Day 0 on CancerSEA Cancer Cell Markers.

3.1.2. Day 2 Cell-cell interactions

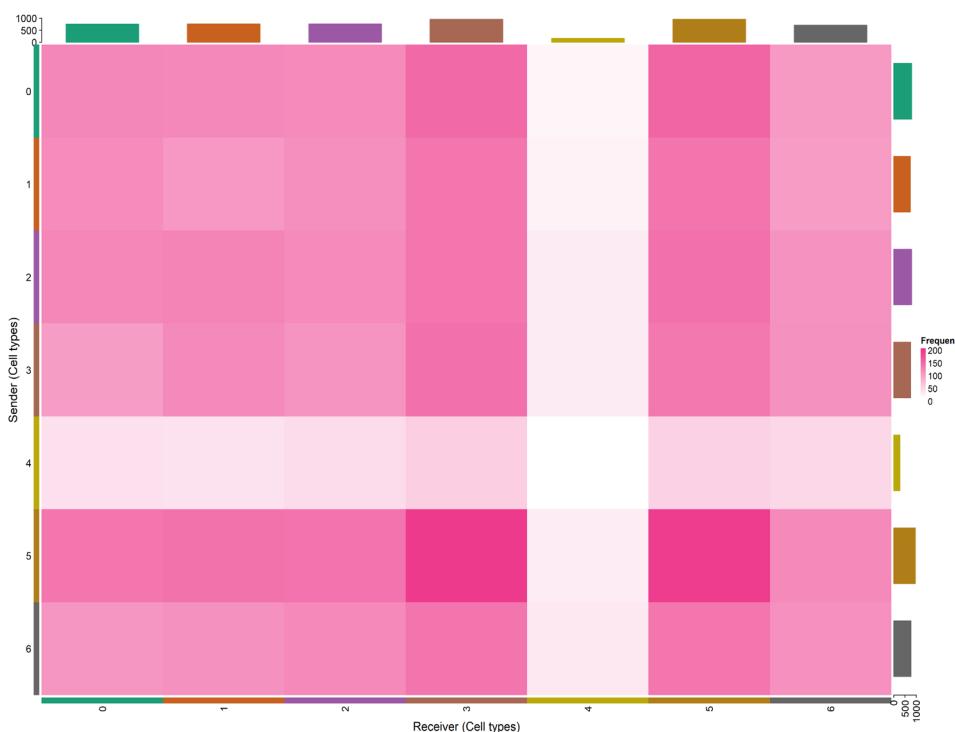


Figure 12

Table 4 Top 10 cell-cell interactions for Day 2 Seurat Clusters (0-6).

source	target	ligand.complex	receptor.complex
--------	--------	----------------	------------------

3	3	TFPI	F3
6	0	SEMA3A	NRP2
4	6	LGALS3	ANXA2
6	5	SEMA3A	NRP2
6	1	SEMA3A	NRP2
3	6	PVR	NECTIN3
6	3	NECTIN3	PVR
1	6	NECTIN2	NECTIN3
6	1	NECTIN3	NECTIN2
2	4	APP	RPSA

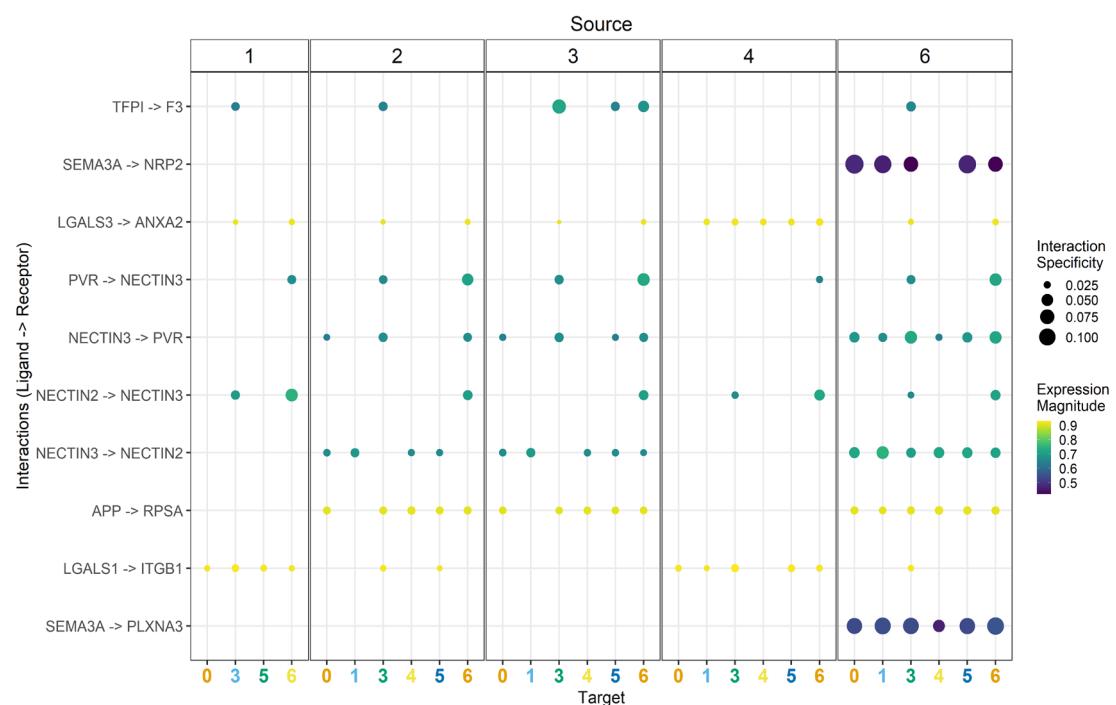


Figure 13. Most relevant cell-cell interactions for Day 2 on Seurat clustering

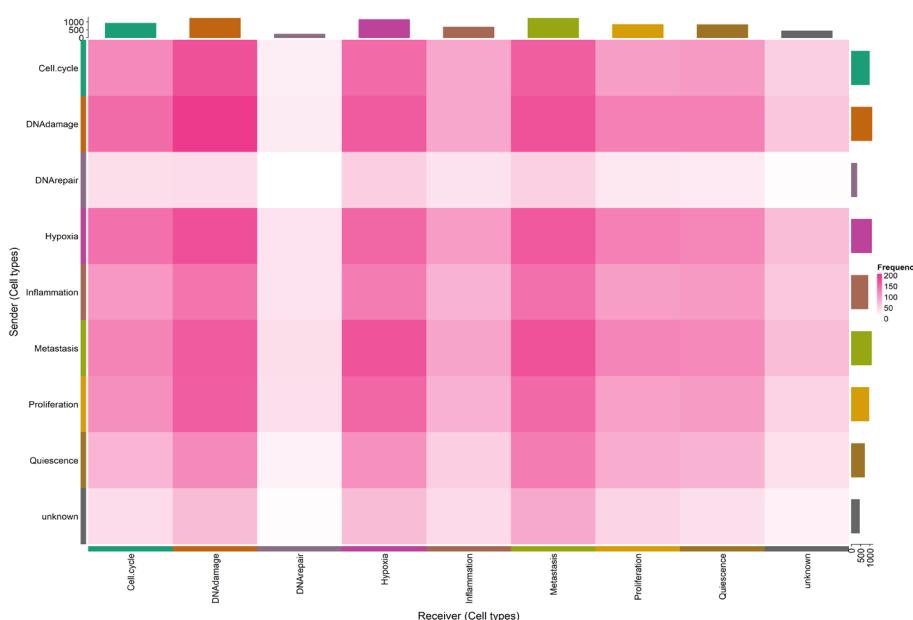


Figure 14

Table 5 Top 10 cell-cell interactions for Day 2 CancerSEA Cancer Cell Markers.

source	target	ligand.complex	receptor.complex
Inflammation	Inflammation	CXCL8	SDC1
	Cell.cycle	CXCL8	SDC1
	DNAdamage	CXCL8	SDC1
	DNAdamage	CXCL8	SDC3
	Quiescence	CXCL8	SDC3
	Cell.cycle	CXCL8	SDC3
	Quiescence	CXCL8	SDC1
	Hypoxia	CXCL8	SDC3
	unknown	CXCL8	SDC3
	unknown	CXCL8	SDC1

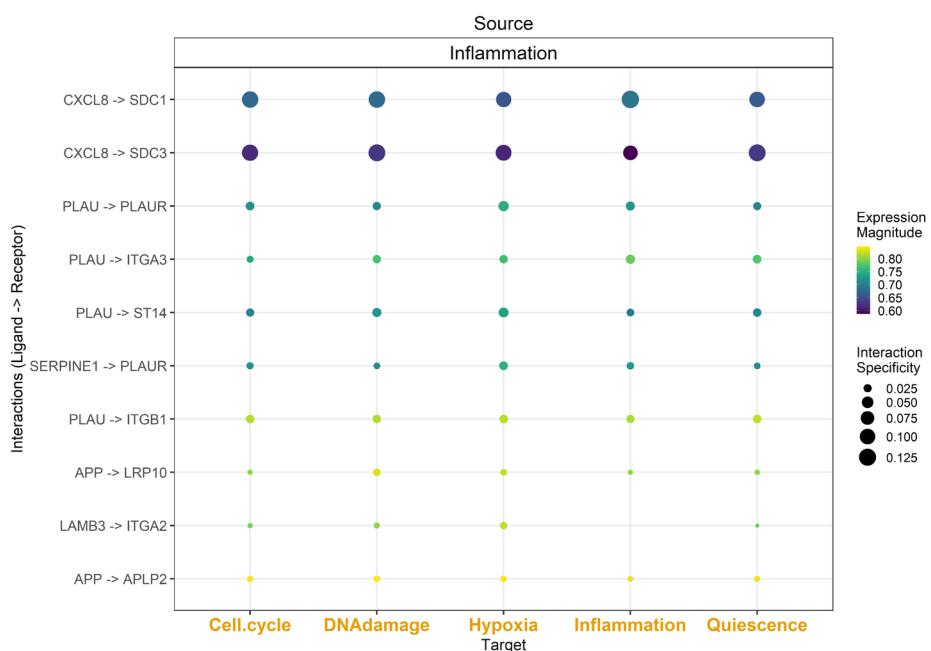


Figure 15. Most relevant cell-cell interactions for Day 2 on CancerSEA Cancer Cell Markers.

3.1.3. Day 7 Cell-cell interactions

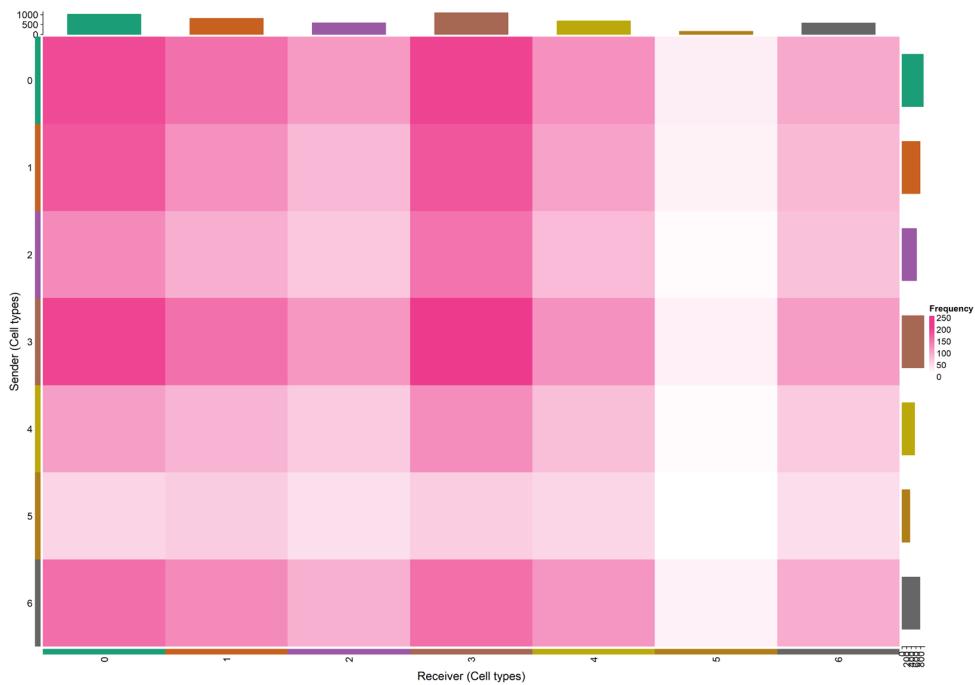


Figure 16

Table 6 Top 10 cell-cell interactions for Day 7 Seurat Clusters (0-6).

source	target	ligand.complex	receptor.complex
5	6	MDK	NCL
5	6	MIF	CD44
4	6	CALM1	HMMR
1	6	CALM1	HMMR
6	6	CALM1	HMMR
4	1	HMGB1	SDC1
5	2	MIF	CD44
5	0	TIMP1	CD63
2	6	TFPI	F3
0	3	LAMB3	ITGB4

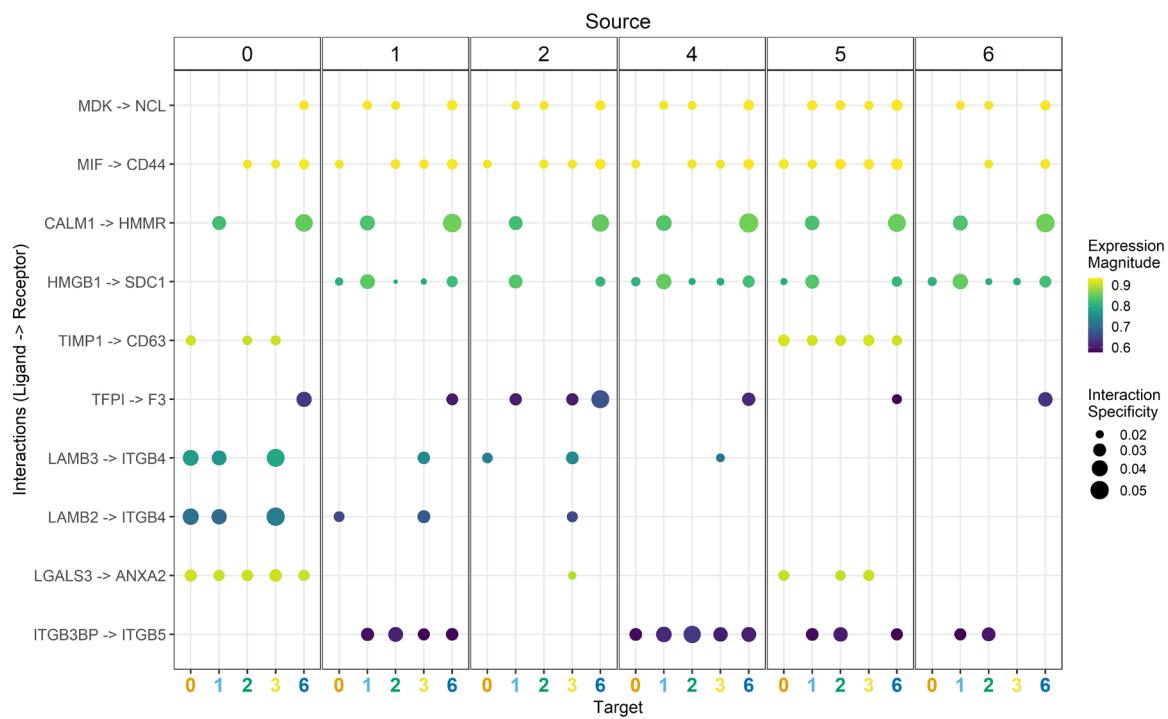


Figure 17. Most relevant cell-cell interactions for Day 7 on Seurat clustering

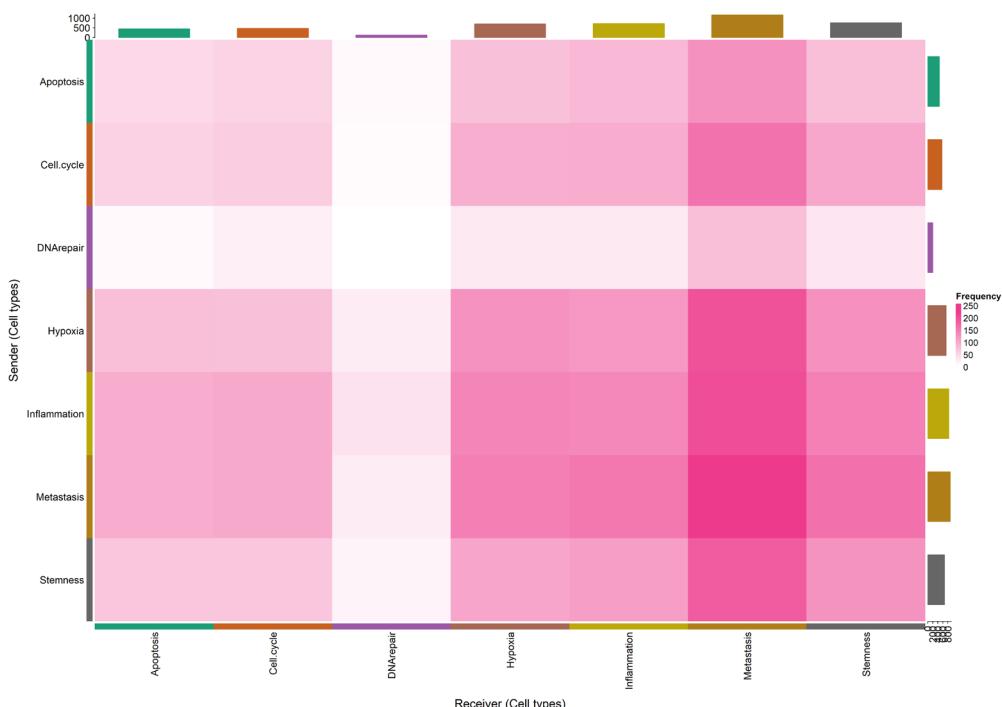


Figure 18

Table 7 Top 10 cell-cell interactions for Day 7 CancerSEA Cancer Cell Markers.

source	target	ligand.complex	receptor.complex
Hypoxia	Inflammation	ANGPTL4	SDC4
	Cell.cycle	CXCL8	SDC1
	Apoptosis	VEGFA	GRIN2B
	Inflammation	LAMC2	ITGB4
	Inflammation	FGF2	SDC4
	Apoptosis	ANGPTL4	SDC4
	Inflammation	DUSP18	ITGB4
	Stemness	LAMB1	ITGB4
	Inflammation	TNFSF15	TNFRSF6B
	Inflammation	CD59	TNFSF15

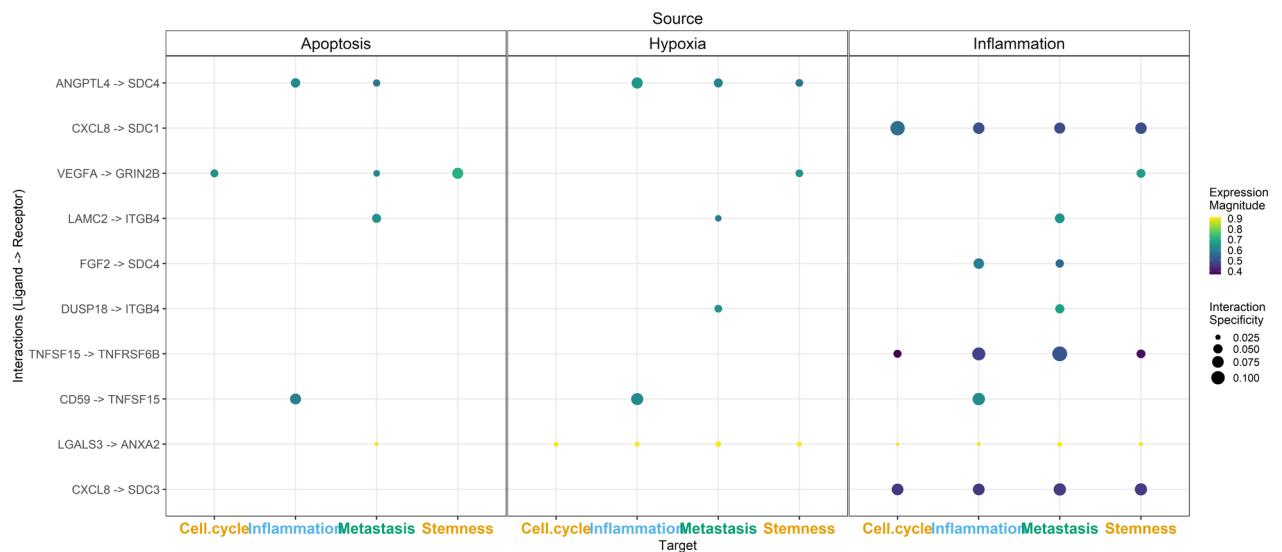


Figure 19. Most relevant cell-cell interactions for Day 7 on CancerSEA Cancer Cell Markers.

3.1.4. Day 14 Cell-cell interactions

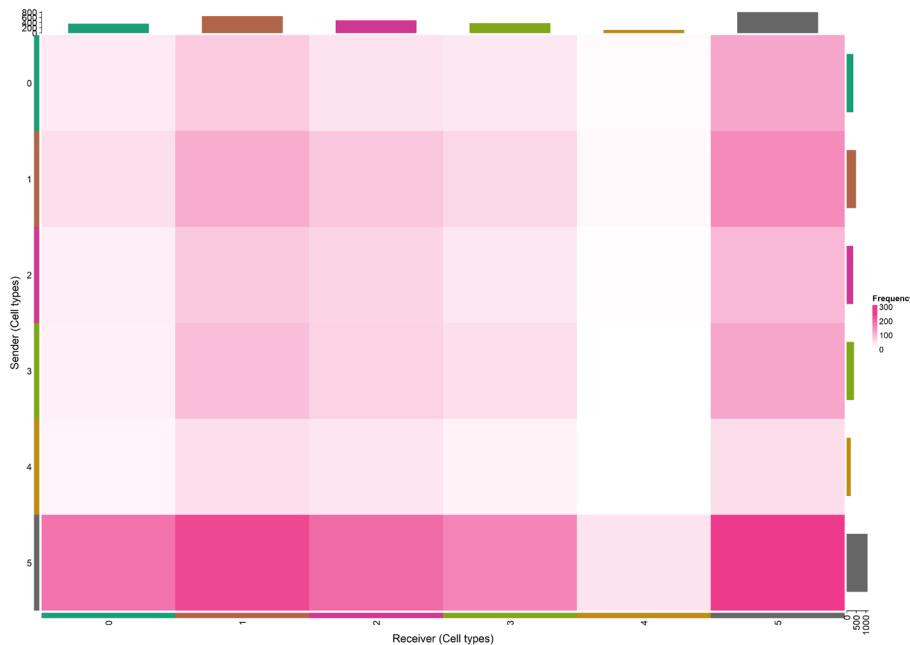


Figure 20

Table 8 Top 10 cell-cell interactions for Day 14 Seurat Clusters (0-5).

source	target	ligand.complex	receptor.complex
5	5	CD59	TNFSF15
5	5	PSAP	LRP1
5	5	LAMB3	ITGB4
5	5	HSPG2	LRP1
5	2	LGALS3BP	ITGB1
5	5	LAMB2	ITGB4
5	5	LAMB3	COL17A1
5	5	LAMA3	ITGB4
5	3	LGALS3BP	ITGB1
5	1	LGALS3BP	ITGB1

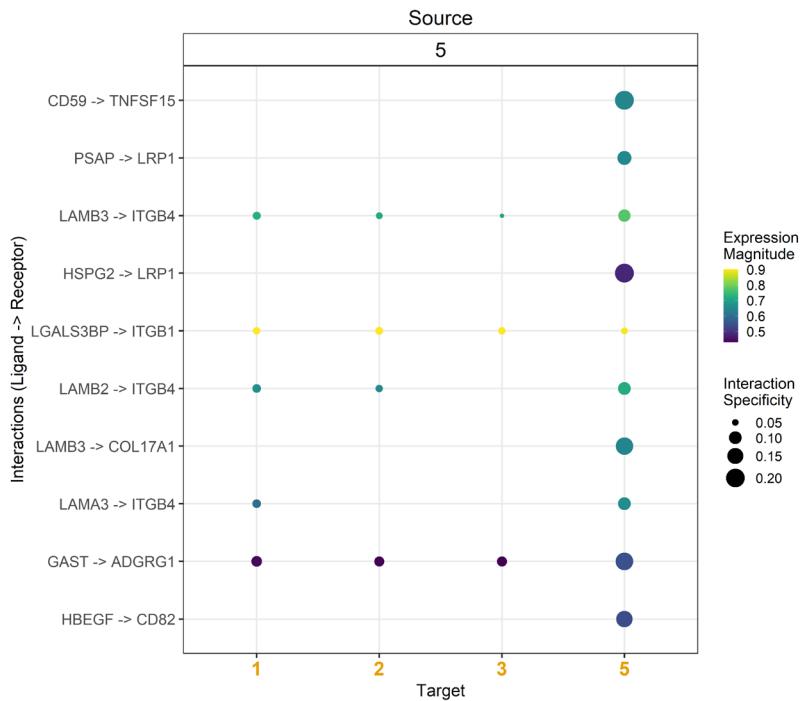


Figure 21. Most relevant cell-cell interactions for Day 14 on Seurat clustering

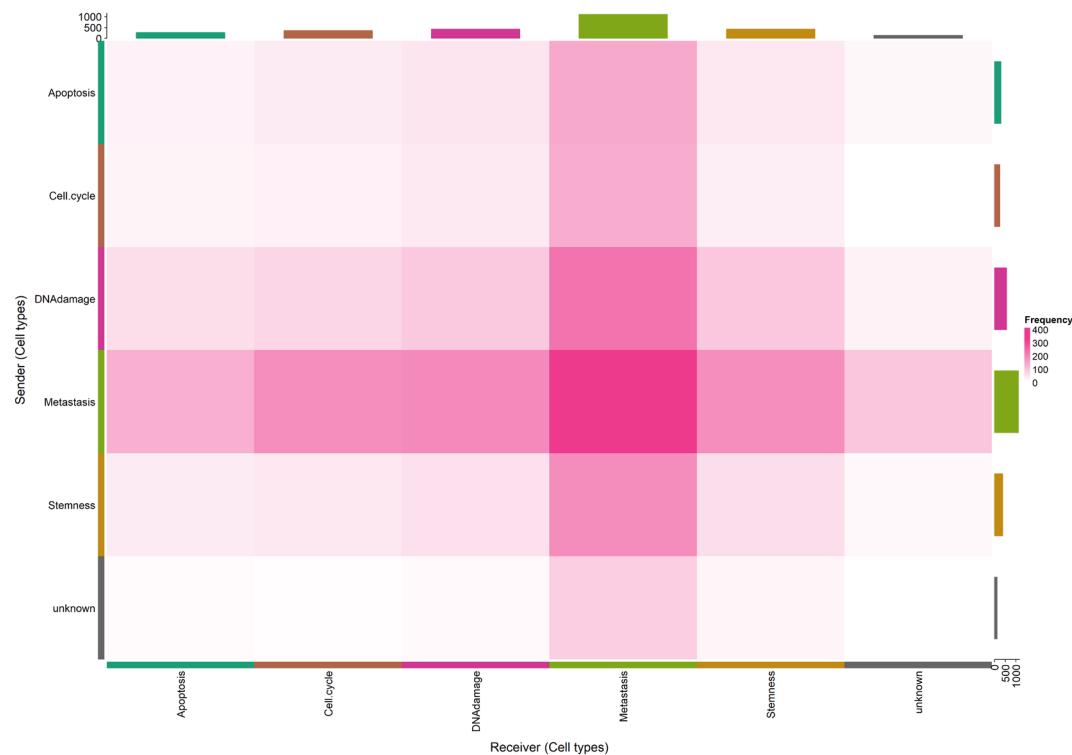


Figure 22

Table 9 Top 10 cell-cell interactions for Day 14 CancerSEA Cancer Cell Markers.

source	target	ligand.complex	receptor.complex
Metastasis	Metastasis	LAMC2	ITGA3
Metastasis	Metastasis	LAMC2	ITGB4
Apoptosis	Metastasis	LGALS9	CD47
Metastasis	Metastasis	LAMC2	ITGB1
Apoptosis	Cell.cycle	LGALS9	SLC1A5
Apoptosis	Metastasis	LGALS9	CD44
Metastasis	Metastasis	LGALS1	ITGB1
Metastasis	Metastasis	LAMC2	ITGA2
Metastasis	Metastasis	LAMA3	ITGA3
Apoptosis	Stemness	LGALS9	CD44

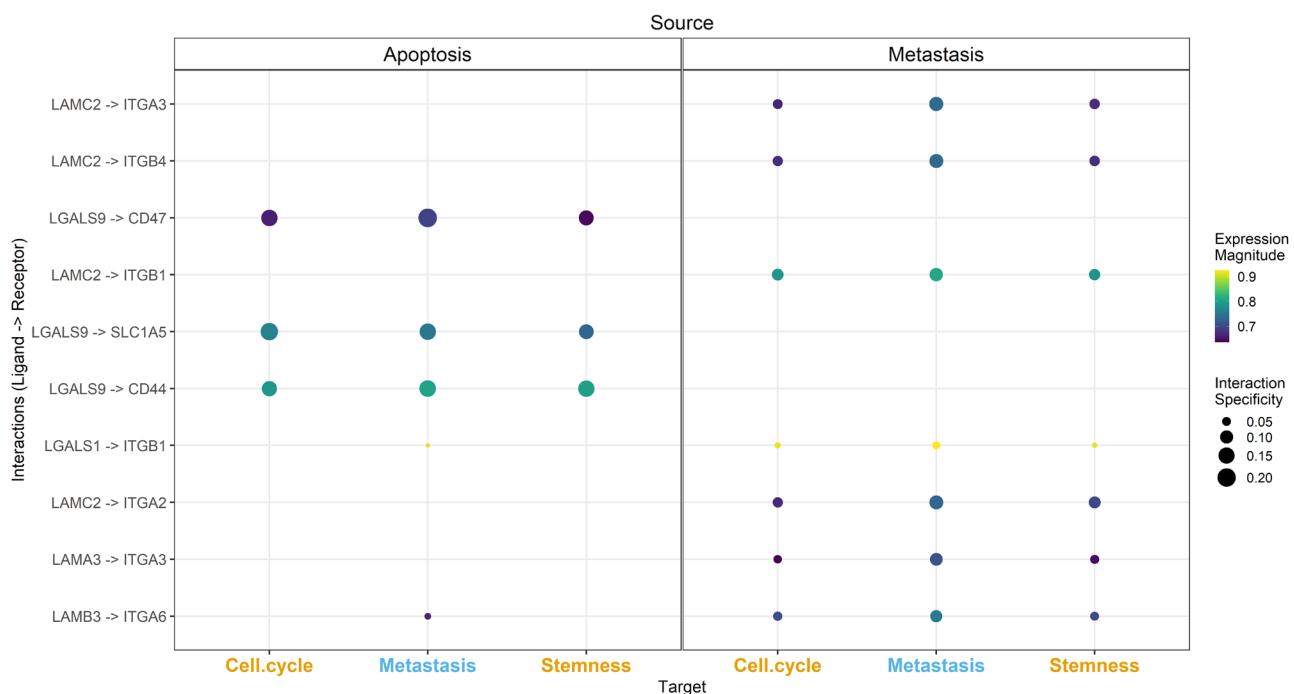


Figure 23. Most relevant cell-cell interactions for Day 14 on CancerSEA Cancer Cell Markers.

3.2. Integrative Analysis

The unintegrated method UMAP is shown in Figure 6, 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 represent in 2D as seen in Figure 8. Additionally, the UMAP is split into different time points, showing clusters from day 0, clusters from day 2, and so on.

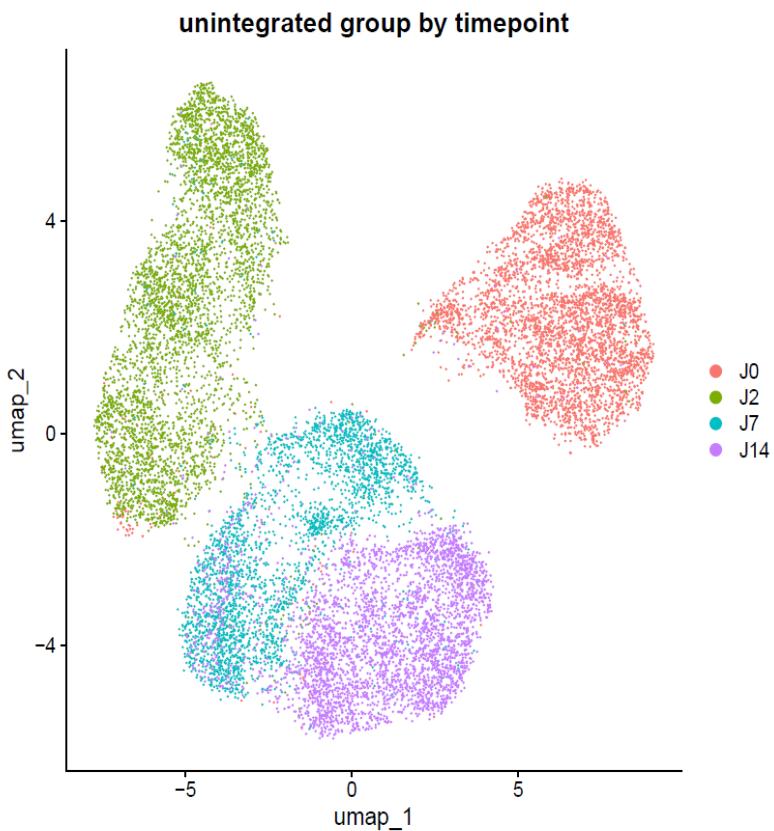


Figure 24. UMAP merging all Time points, representing the unintegrated analysis. Each timepoint is represented in different colors as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).

We then applied the Harmony integration method to mitigate the batch effect across different cell time points. As seen in Figure 9, the batch effect from the Figure 6 has been corrected, and cells from different time points are spatially merged. Subsequently, we identified Seurat clusters based on this final correction, recognizing 8 distinct clusters (Figure 9). For easier visualization of the clusters, we created plots split by the different days, from day 0 to day 14. Finally, based on the CancerSEA table, we distinguished the cell markers separately for each day (Figure 10).

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

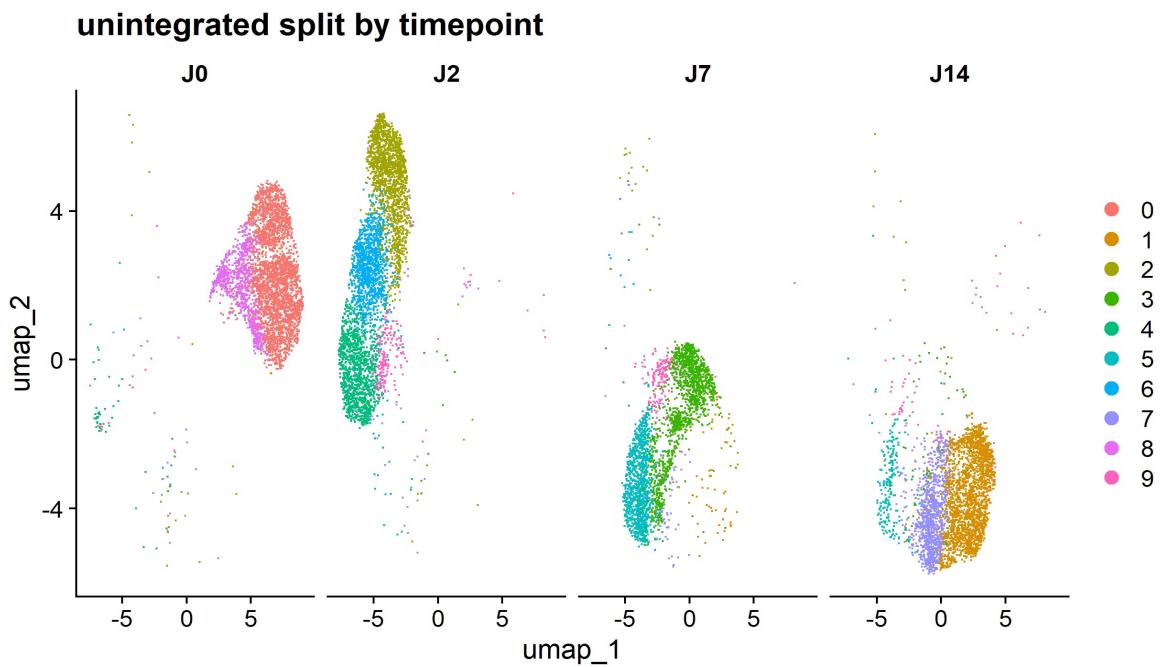


Figure 25. UMAP merging all data times matrices without integrating methodology, represented split by days (J0, J2, J7 and J17), a total of 9 different Clusters are identified.

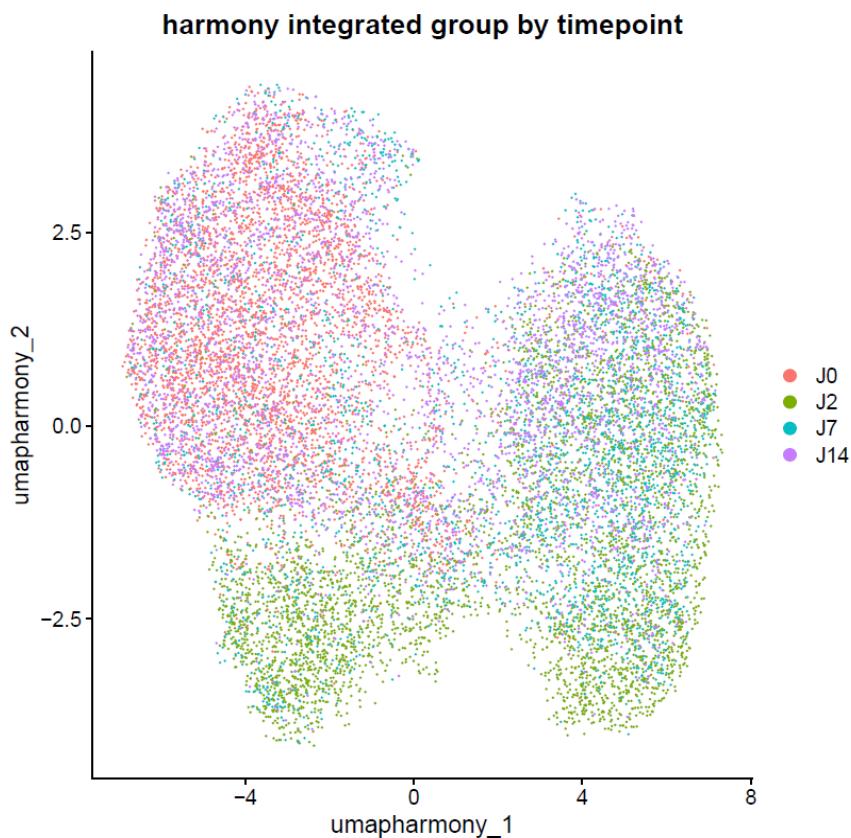


Figure 26. UMAP merging all Time points, representing the integrated analysis using Harmony as Integration methodology. Each timepoint is represented in different colors as J0 (day 0), J2 (day 2), J7 (day 7) and J14 (day 14).

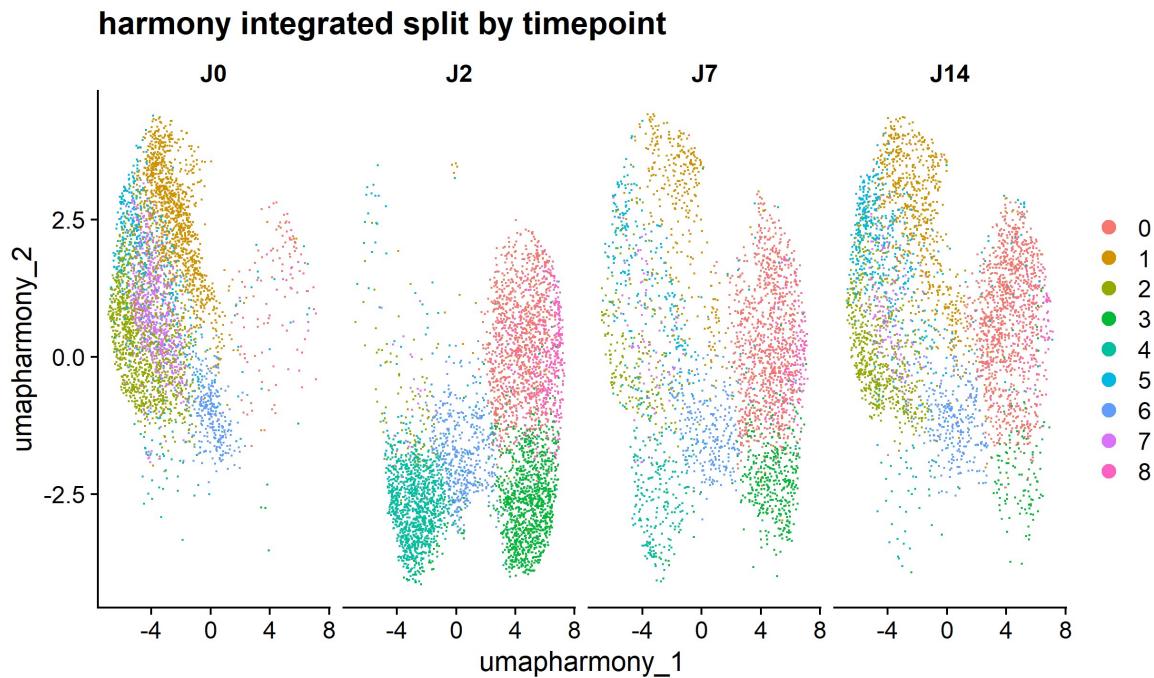


Figure 27. UMAP merging all data times matrices using Harmony as Integration methodology, represented split by days (J0, J2, J7 and J17), a total of 8 different Clusters are identified.

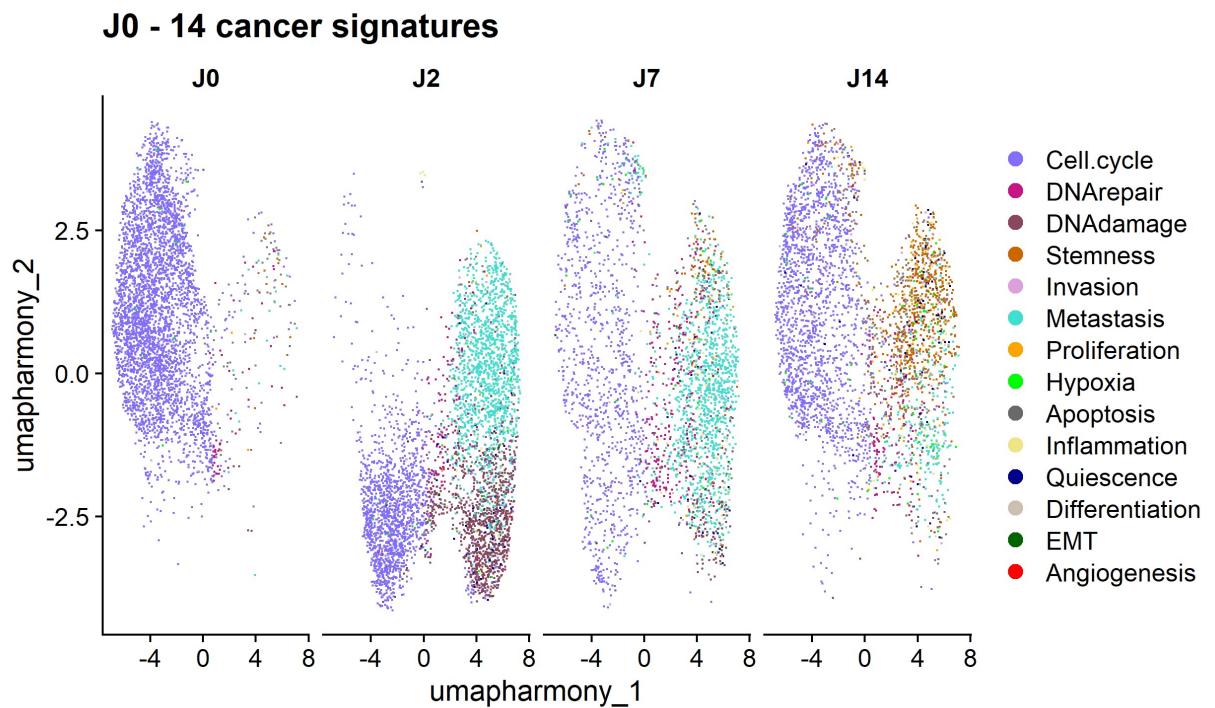


Figure 28. UMAP merging all data times matrices using Harmony as Integration methodology, represented split by days (J0, J2, J7 and J17), the CancerSEA Cancer Cell Markers were represented.

3.2.1. Cell-cell interactions

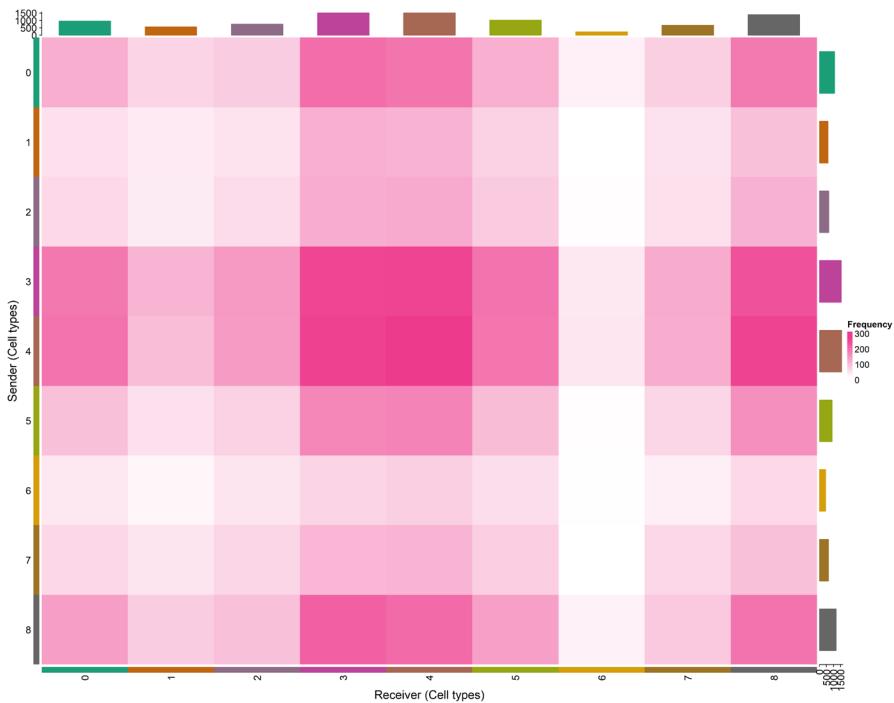


Figure 29

Table 10 Integrative analysis top 10 cell-cell interactions Seurat Clusters

source	target	ligand.complex	receptor.complex
6	1	MDK	NCL
8	8	LAMA4	ITGA3
4	3	HBEGF	CD82
6	7	MDK	NCL
3	3	GAST	ADGRG1
4	8	LGALS1	ITGB1
8	3	LAMA4	ITGA3
3	8	LGALS3	ANXA2
4	8	LAMA3	ITGA3
6	3	TIMP1	CD63

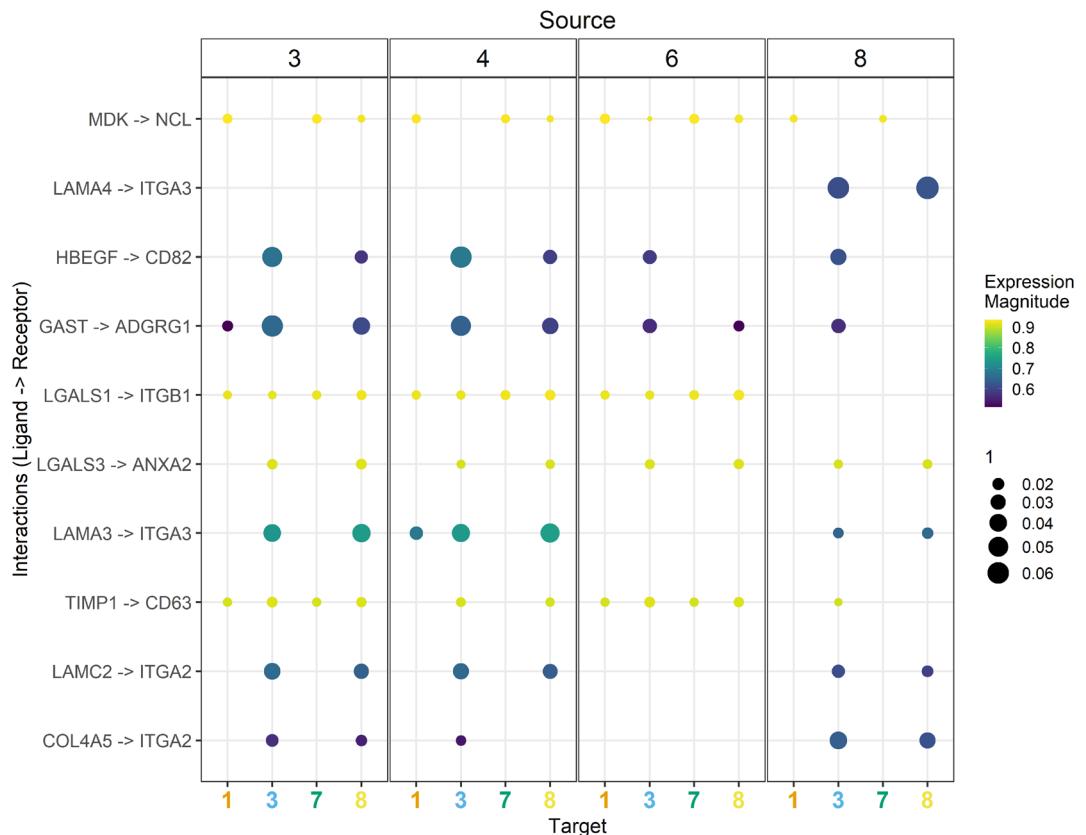


Figure 30. Most relevant cell-cell interactions for the integrative analysis on Seurat clustering.

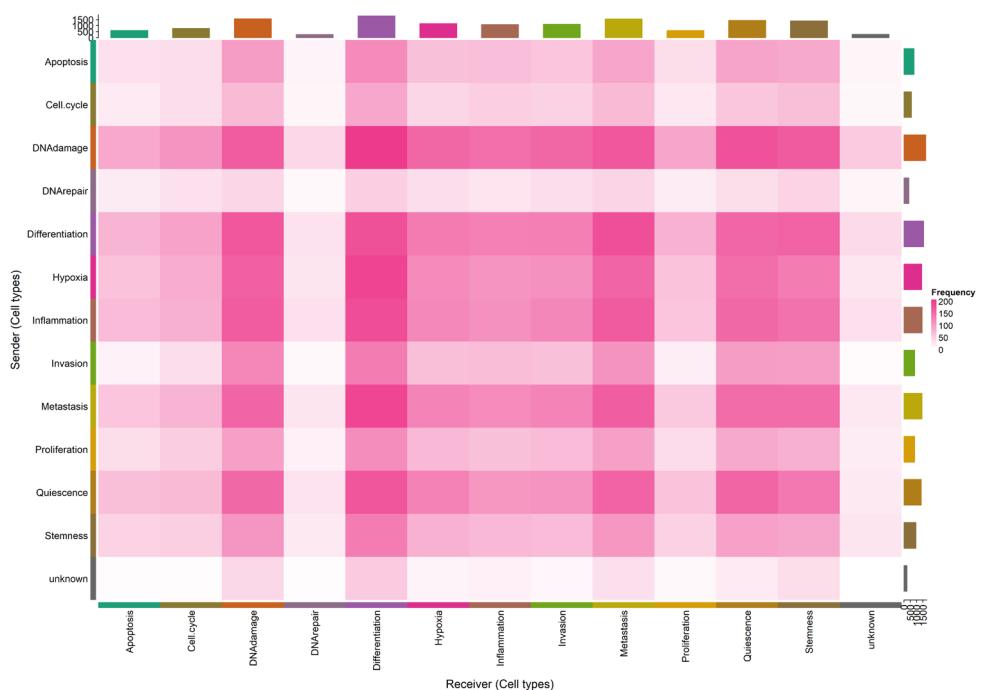


Figure 31

Table 11 Integrative analysis top 10 cell-cell interactions of CancerSEA Cancer Cell Markers.

source	target	ligand.complex	receptor.complex
Inflammation	Differentiation	CXCL8	SDC3
Inflammation	DNAdamage	CXCL8	SDC1
Inflammation	Cell.cycle	CXCL8	SDC1
Inflammation	Inflammation	CXCL8	SDC3
Inflammation	Differentiation	CXCL8	SDC1
Inflammation	Hypoxia	CXCL8	SDC3
Inflammation	Stemness	CXCL8	SDC3
Inflammation	Metastasis	CXCL8	SDC3
Inflammation	Quiescence	CXCL8	SDC3
Inflammation	DNAdamage	CXCL8	SDC3

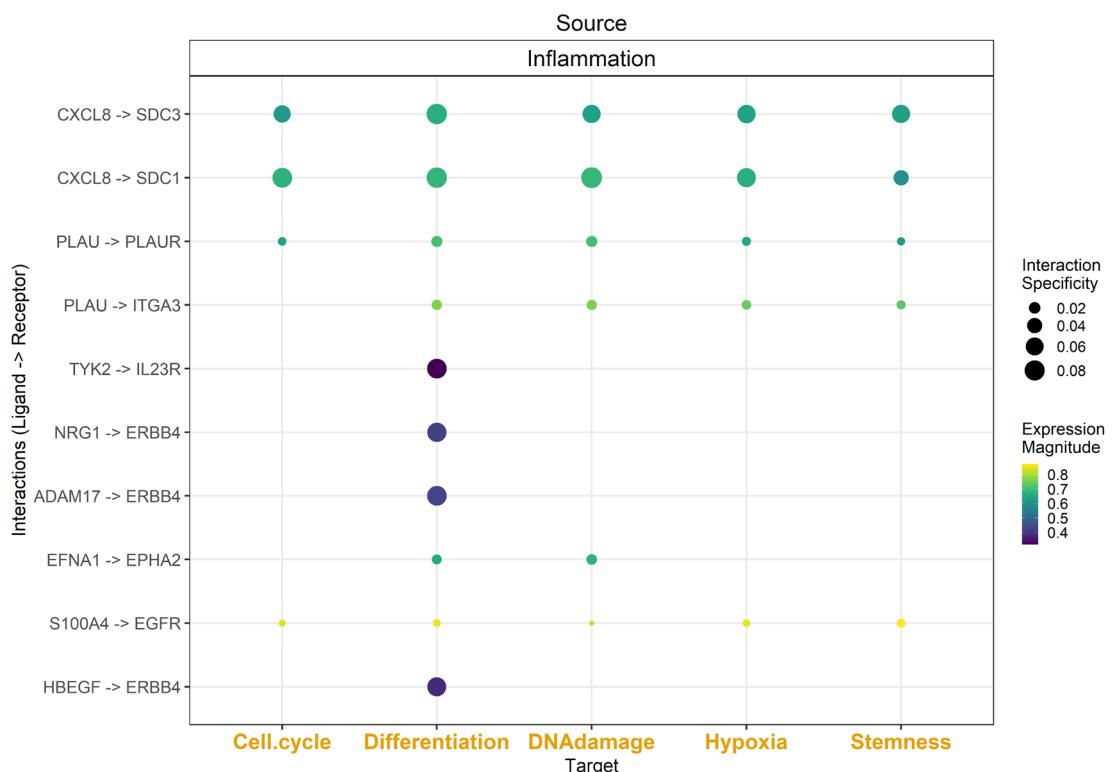


Figure 32. Most relevant cell-cell interactions for the integrative analysis on CancerSEA Cancer Cell Markers.

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] "Detection of differentially abundant cell subpopulations in scRNA-seq data." Accessed: Jun. 04, 2024. [Online]. Available: <https://www.pnas.org/doi/epdf/10.1073/pnas.2100293118>
- [9] "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>
- [10] "R: The R Project for Statistical Computing." Accessed: May 13, 2024. [Online]. Available: <https://www.r-project.org/>
- [11] "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/>
- [12] 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.
- [13] 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.
- [14] 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.

- [15] 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.