I. **Tittle**: Optimization of the scRNA-seq workflow using the *Seurat* package from R, in the context of colorectal cancer data.

II. **Key words**: scRNA-seq, Seurat package, R, Colorectal cancer, Epithelial cells, Optimization, Bioinformatics, Cancer, Adaptive response, Adaptive evolution, Integration.

2. **Context and Justification of the Work:**

5-Fluorouracil (5-FU), an agent used in all chemotherapy combinations for colorectal cancer (CRC) as well as in many other cancers, is often associated with resistance and consequently tumor recurrence. Indeed, 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.

Mechanisms of drug tolerance remain poorly understood and may be explained by non-genomic processes. Transcriptomic and epigenetic studies have described cellular plasticity as a process contributing to treatment tolerance. However, the translation of messenger RNA into proteins remains unexplored in the context of tumor tolerance to treatments.

The tolerance phase is followed by a significant resumption of cell division. Furthermore, single-cell RNA sequencing (scRNAseq) analyses were performed on CRC cells before and after 5-FU treatment. Preliminary scRNAseq analyses suggest the presence of 11 clusters before, during, and after treatment, indicating that 5-FU promotes cellular reprogramming/enrichment so that tumor cells become tolerant to 5-FU. This highlights a yet poorly evaluated role of tumor cell plasticity in the response to chemotherapy in CRC.

The objective of my thesis project will be to demonstrate how translational control is responsible for cellular reprogramming in response to 5-FU to obtain characterization of the different subpopulations involved in tumor recurrence.

3. **General Description**:

The Master's thesis project aims to investigate the role of translational control in cellular reprogramming in response to 5-Fluorouracil (5-FU) treatment in colorectal cancer (CRC). The project will utilize single-cell RNA sequencing (scRNAseq) techniques, particularly focusing on optimizing the workflow using the Seurat package from R to isolate relevant cellular subpopulations. The obtained subpopulations will then be subjected to detailed analyses of their translatome to understand the mechanisms underlying treatment tolerance. Phenotypic characterization of these cellular subpopulations will be performed using cell and molecular biology techniques. Finally, the project will evaluate the contribution of the identified subpopulations to 5-FU treatment tolerance and tumor recurrence. Overall, the project aims to shed light on the role of translational control in CRC treatment response and provide insights into potential therapeutic strategies to combat tumor recurrence.

4. **Sustainability, social-ethical and diversity impact**:

In the context of the thesis project investigating translational control in CRC treatment response, the project aims to:

- **Sustainability:** Minimize environmental impact through eco-friendly lab practices, such as the reduction of paper use or taking care of the waste during lab practices.
- **Social-Ethical Impact:** Adhere to ethical standards, considering societal implications of findings.
- **Diversity Impact:** Foster diversity within the team, ensure equitable access to resources, and promote inclusivity in research outcomes.

Overall, the project emphasizes responsible, ethical, and inclusive research practices while striving for environmental sustainability.

5. **Objectives**:

    I. **General objectives**: The objectives of the thesis project can be enumerated as follows:

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

    II. **Specific objectives**: Here aret he specific objectives of these work:

        1. Investigate the role of translational control in 5-FU-induced cellular reprogramming.
        2. Employ and optimize single-cell RNA-seq workflow using Seurat Package from R in order to obtain relevant cellular subpopulations.
        3. Perform detailed analyses of the translatome of these subpopulations using appropriate methodologies, thanks to Seurat Package workflow and optimization.
            1. Comparison between Integrative analysis against the non-integrative analysis.
        4. Phenotypically characterize the cellular subpopulations using cell and molecular biology techniques.
        5. Evaluate the contribution of the identified cellular subpopulations to 5-FU treatment tolerance and tumor recurrence.

6. **Approach and method to be followed**:

The project will employ several strategies to achieve its objectives, with a focus on optimizing the workflow for single-cell RNA sequencing (scRNAseq) using the Seurat package from R, characterizing cellular subpopulations, and evaluating their role in tumor recurrence post-5-FU treatment.

- **Possible Strategies:**
    1) Literature Review and Preliminary Analysis: Initially, a comprehensive review of existing literature will be conducted to understand previous research findings and identify gaps in knowledge. Preliminary analyses of scRNAseq data may also be performed to gain insights into potential cellular subpopulations.
    2) Experimental Design and Data Collection: Based on the literature review and preliminary analysis, an experimental design will be formulated to isolate relevant cellular subpopulations and collect scRNAseq data before and after 5-FU treatment.
    3) Workflow Optimization: The Seurat package from R will be utilized for scRNAseq data analysis. Various strategies for workflow optimization, such as parameter tuning and integration of multiple datasets, will be explored to enhance the accuracy and efficiency of data processing.
        - Integrative analysis: Implementation of the Integrative analysis and comparison against the non-Integrative analysis.
    4) Translatome Analysis: Detailed analyses of the translatome of identified cellular subpopulations will be performed to elucidate the molecular mechanisms underlying treatment tolerance and tumor recurrence. This will involve integrating transcriptomic and proteomic data to capture the dynamics of mRNA translation. Identification of cell markers.
    5) Phenotypic Characterization: Cellular subpopulations will be phenotypically characterized using cell and molecular biology techniques to understand their functional properties and behavior.

- **Chosen Approach:**

The chosen approach involves optimizing the scRNAseq workflow using the Seurat package from R to obtain detailed insights into cellular reprogramming in response to 5-FU treatment. This approach was selected for its ability to provide high-resolution transcriptomic data at the single-cell level, allowing for the identification and characterization of heterogeneous cellular subpopulations. By leveraging advanced computational tools and experimental techniques, this approach offers a comprehensive and integrative analysis of the molecular mechanisms driving treatment tolerance and tumor recurrence in CRC.

- **Rationale:**

This approach was deemed the most appropriate as it aligns with the project's objectives of elucidating the role of translational control in cellular reprogramming post-5-FU treatment. The use of scRNAseq coupled with the Seurat package allows for the identification of subtle transcriptional changes at the single-cell level, providing valuable insights into cellular heterogeneity and plasticity. Additionally, the integration of transcriptomic and proteomic data facilitates a comprehensive analysis of the translatome, shedding light on the molecular processes underlying treatment response and resistance. Overall, this approach offers a robust framework for achieving the project's objectives and advancing our understanding of CRC treatment.

7. **Planning**:

I. Tasks:

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.

Experimental Design and Data Collection:

Duration: 3 weeks

Task: Design experimental protocols for isolating cellular subpopulations and collecting scRNAseq data before and after 5-FU treatment.

Workflow Optimization:

Duration: 4 weeks

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

Translatome Analysis:

Duration: 6 weeks

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

Phenotypic Characterization:

Duration: 4 weeks

Task: Phenotypically characterize cellular subpopulations using cell and molecular biology techniques to understand their functional properties.

II. Calendar:

III. Milestones:

Completion of Literature Review and Preliminary Analysis: Week 2

Experimental Design and Data Collection Finalized: Week 5

Completion of Workflow Optimization: Week 9

Translatome Analysis Completed: Week 15

Phenotypic Characterization Finished: Week 19

IV. Risk Analysis:

Scope Creep: Changes or additions to the project scope may lead to delays or deviations from the planned timeline.

Technical Challenges: Unexpected technical issues with experimental procedures or data analysis tools could hinder progress.

Resource Constraints: Limited availability of resources, such as laboratory equipment or computational resources, may impact the project timeline.

Data Quality Issues: Poor quality or insufficient data may necessitate additional time for troubleshooting and validation.

External Factors: External events, such as changes in regulations or disruptions due to unforeseen circumstances, could affect project progress.

Mitigation strategies include regular communication, contingency planning, and flexibility in adapting to challenges while maintaining focus on project objectives and deadlines.

8. **Expected results**:

I. Expected Results:

Work Plan: Detailed and organized plan outlining the tasks, timeline, and milestones achieved throughout the project.

Thesis Report: Comprehensive document summarizing the research background, objectives, methodology, results, and conclusions. It will also include detailed analysis, discussions, and references.

Data Analysis and Interpretation: Analyzed and interpreted scRNAseq data, including optimized workflow, translatome analysis, and phenotypic characterization results.

Final Presentation: Virtual presentation summarizing key findings, methodology, and implications of the research project.

II. Deliverables:

Work Plan: Documented plan outlining tasks, timeline, and milestones.

Thesis Report: Completed document containing all relevant sections and analyses.

Data Analysis and Interpretation: Detailed analysis and interpretation of scRNAseq data.

Final Presentation: Virtual presentation slides summarizing the research project.

III. Additional Results:

The expected additional results, if applicable, may include: Published Article: If the research yields significant findings, a manuscript may be prepared for submission to a peer-reviewed journal.

Software or Tool: If novel computational tools or software were developed during the project, they will be documented and made available for potential use by other researchers or practitioners.

Supplementary Materials: Any supplementary materials such as raw data, code, or additional analyses will be provided as necessary to support the findings presented in the thesis report.

These deliverables will collectively demonstrate the successful completion of the research project and the attainment of its objectives.

9. **Bibliography**: