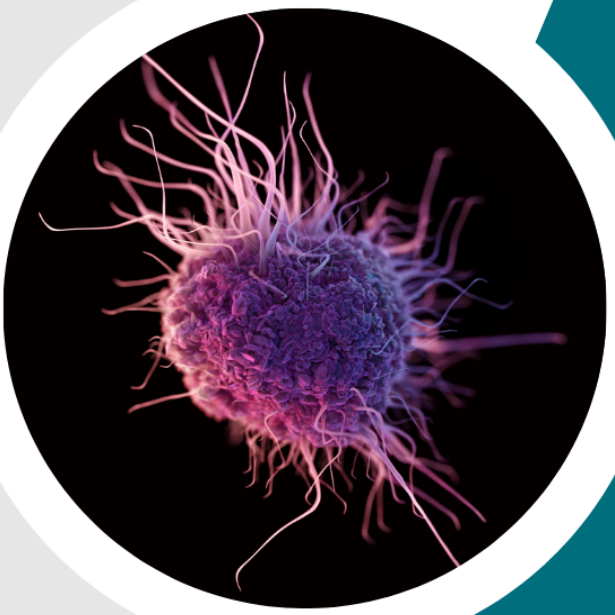
ALTERNATIVE SPLICING REMODELING IN PLASMACYTOÏD DENDRITIC CELLS IN RESPONSE TO VIRUS INFECTION

-

SPECIFICATION REPORT



PROJECT LEADER: DELPHINE GALIANA

PROJECT TUTOR: VINCENT LACROIX

RAYANN LARBI
JORDAN DUTEL
ARIANE PARADAN

Table of content

Table of content **2**
Abstract **3**
Introduction **3**
State of the art **5**
1. Biological context 5
2. RNaseq 5
3. Bioinformatics tools 5
4. Dataset 6
Project team **6**
Objectives **7**
Final product **7**
Constraints **7**
1. Cost 7
2. Deadline 7
3. Complexity of the dataset 8
4. Other constraints 8
Planning **8**
References **9**

Abstract

Plasmacytoïd dendritic cells (pDCs) are innate immune cells that can exhibit various forms and functions. These features are, in most part, shaped by transcriptomic regulations. Alternative splicing of mRNA is one of them. In this study, to understand this mechanism, researchers have collected transcriptomic data (RNAseq) of pDC exposed to a viral infection. In our project, we will confirm that quality and coverage of NGS (Next Generation Sequencing) data are usable and acceptable. Our investigation will be based on bibliography, and we will assess the accuracy of possible analysis. Finally, our project should give an answer about a potential alternative splicing remodeling in pDC after a virus exposition.

Introduction

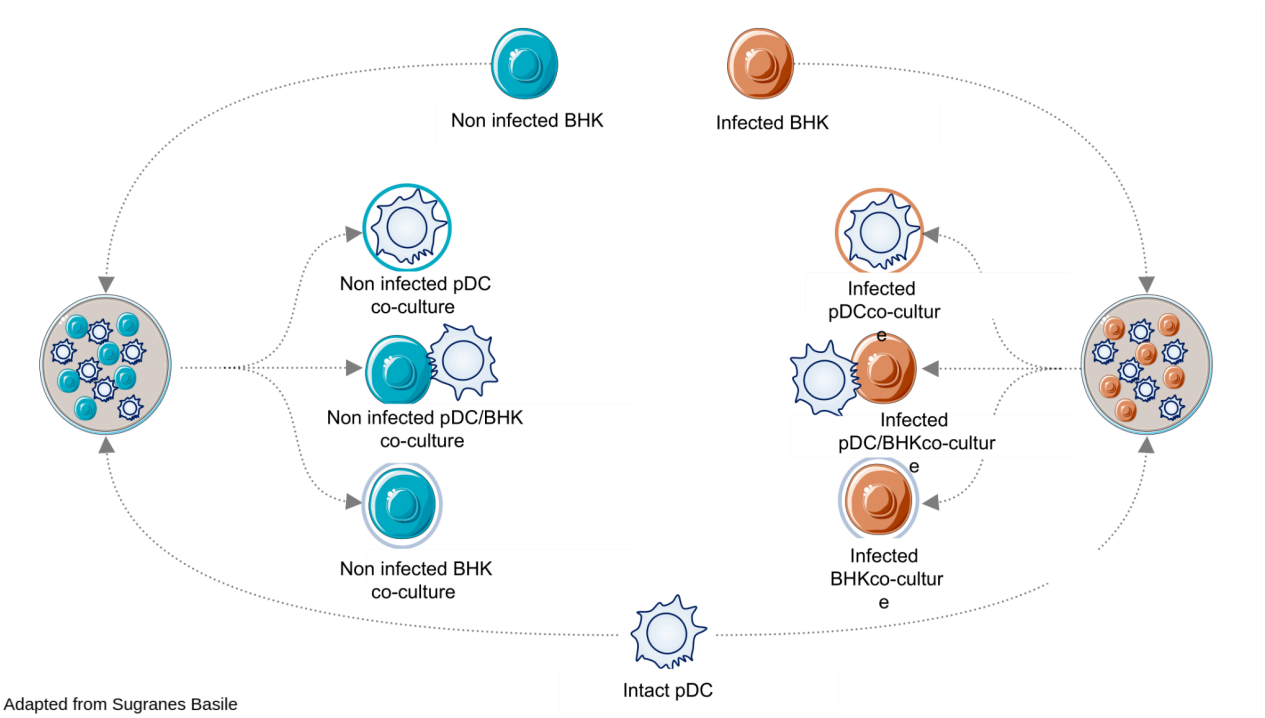
Humans have two primary immune systems: the innate and adaptive immune systems. Among the key players in innate immunity, plasmacytoid dendritic cells (pDCs) are pivotal in orchestrating the antiviral response. These "sentinel cells", although comprising only a minor fraction of circulating blood cells (0.1 to 0.5%) ***[8],*** excel in immune defense, especially during viral infections. Upon encountering a virus-infected cell, pDCs initiate their response through the recognition of viral nucleic acids. Those specific DNA or RNA are foreign to the human cellular environment. pDCs are equipped with pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs), particularly TLR7, which are adept at detecting these viral patterns ***[2]***. Once these receptors bind to viral nucleic acids, a cascade of intracellular signaling is set off, leading to the activation of the pDCs. A critical aspect of pDC activation is the transcriptional upregulation of genes coding for type I interferons, predominantly interferon-alpha (IFN-α) ***[1][3]***. Virus-infected cells are also able to produce IFN-α. Type I interferons are potent antiviral molecules that play a dual role in the immune response. First, they act in an autocrine manner, further activating infected cells itself and enhancing their ability to present virus antigens in order to be recognized by T-CD8 cells ***[1] [2]***. Secondly, when mostly secreted by pDC, they function in a paracrine way, alerting neighboring cells of the viral threat. These neighboring cells, upon interfacing with interferons, upregulate antiviral genes that heighten their resistance to viral infection and replication ***[1]***. Cells put themselves in an antiviral state. Additionally, through the production of various cytokines and chemokines, pDCs recruit and activate other immune cells, thereby amplifying the coordinated immune attack against the virus. The establishment of this antiviral response by pDCs involves a finely regulated transcriptional activity to produce the most suitable cytokines and chemokines for the situation. This plasticity in gene expression and the transcriptional regulatory mechanisms are partially governed by alternative splicing mechanisms of transcripts. ***[3]***

According to the NIH, alternative splicing is a "cellular process in which exons from the same gene are joined in different combinations, leading to different, but related, mRNA transcripts. These mRNAs can be translated to produce different proteins with distinct structures and functions all from a single gene" ***[10]***. In the context of pDCs, alternative splicing could significantly influence their function, particularly in response to viral infection. Thus, changes in the alternative splicing patterns of pDCs under infectious conditions could have a major impact on their ability to produce interferons, alert others cells, express particular receptors and recognize infected cells.

This project aims to employ advanced bioinformatic tools to detect and analyze alternative splicing in pDCs, both under infectious conditions and in their absence. When a pDC initiates prolonged contact with an infected cell, it involves adhesion proteins, the cytoskeleton, and the endocytosis machinery, allowing for the transfer of viral nucleic acids from the infected cell to the pDC. This corresponds to the establishment of the interferogenic synapse ***[2]***. This leads to the activation of the pDC and the secretion of type I interferon by it. To model this condition, BHK (Baby

Hamster Kidney) cells were previously infected for 48 hours with Dengue virus and then were put in contact with the plasmacytoid dendritic cells for 4 to 5 hours (*Figure below*). However, pDCs can also be activated by free viruses in the environment or by interferon released by their counterparts. In order to represent this condition, another experiment was realized where pDC were directly activated by viruses (*Figure below*). Then, the transcriptome of cells in different conditions was sequenced by RNA sequencing method. Data generated will be analyzed and allow us to investigate alternative splicing in pDC according to contact types between cells and activation state of cells. The goal is to understand how changes in alternative splicing can affect pDC function and, by extension, the overall immune response during viral infection. We plan to utilize bioinformatic tools to identify specific splicing patterns and their correlation with pDC activity.

This work may face challenges, particularly in terms of the complexity of splicing data and the interpretation of results. Nonetheless, the potential discoveries offer promising prospects for a better understanding of immune mechanisms and could eventually contribute to the development of new therapeutic approaches for combating viral infections.



Adapted from Sugranes Basile

State of the art

1. Biological context

In immune cells, alternative splicing plays a crucial role to tightfully orchestrating immune responses. As an example in innate immune cells, STING (STimulator of INterferon Genes) proteins, could undergo exon skipping on transcripts, leading to another STING isoform ***[3]***. By doing this, cells can deal with complementary signaling pathways and are up to lead adequate IFN-associated responses according to the pathogen they are sensing. In the same manner, pathogens can induce immune escape. As a matter of fact, SARS-CoV-2 by entailing alternative splicing and interfering with immune cells spliceosomes can reduce production of IFN upon viral infection ***[3]***. Despite this, our understanding of how alternative splicing regulates the transcriptome of pDC, especially under pathogen exposure, remains limited. This is depicted by the lack of extensive data on alternative splicing changes during the IFN response in pDCs upon detecting infected cells. Given the increasing frequency of epidemics linked to zoonoses and climate change, it is imperative to deepen our understanding of anti-viral immunity and the mechanisms underlying it, including alternative splicing ***[9]***.

2. RNaseq

RNAseq is a tool used in transcriptomics, the study of all RNA transcripts produced by the genome. Unlike qPCR which only studies one gene at a time, or DNA microchips which study only known and pre-selected genes, RNAseq allows for the study of the entire genome at once with no prior information. Hence RNAseq enables for a variety of discoveries, such as the existence of alternative splicing events, gene fusions, single nucleotide polymorphisms [SNP], or even rare transcripts.

The way it works can be summarized into steps. Messenger RNA is extracted from a biological sample. This mRNA is then fragmented into shorter chunks, and is made to go through reverse transcription using a library of complementary DNA with adaptors. The result can be put through Next Generation Sequencing [NGS] to obtain short sequence reads. ***[4]*** This data is stored in the FASTQ file format, and it is with this data that analysis may be performed.

3. Bioinformatics tools

Several tools are available to perform RNAseq analysis, and each year, new tools appear. In the context of this project, we have selected a few tools. Among these are the following: FastQC, MultiQC and STAR. These tools were selected based on pertinence, as well as our prior experience with them. In addition, two additional softwares were proposed: MISO (Mixture of Isoforms) ***[6]*** and rMATS ***[7]*** (replicate Multivariate Analysis of Transcript Splicing), the latest will be employed, but will require a self-formation period.

We will exploit FastQC and MultiQC to assess the quality of our data. FastQC generates a data quality report from a SAM BAM or FASTQC file. The report is provided in an html file, allowing a quick review of various features useful to detect problems in a library.***[11]*** MultiQC will combine and compare the different reports analysis from each of our samples in a unique html file in order to help us determine any necessary additional steps for our dataset like trimming or filtering.***[12]***

Before mapping cleaned and trimmed reads against the reference human genome, we need to remove RNA of hamster species that come from BHK cells in the condition pDC+BHK. In order to do so, we will carry out a quasi-mapping of these reads against the human transcriptome. This step will be realized with Salmon ***[15]***. This tool will allow us to perform a fast alignment while conserving reads that only belong to pDC.

STAR [Spliced Transcripts Alignment to a Reference] will be used to map sequencing data from all pDC to the human genome and assign a gene for each read of our dataset leading to the detection of alternative splicing events and the quantification of transcripts isoforms. STAR fits well for this task given its splice-aware alignment and its consideration of possible introns during mapping.

To carry out a differential alternative splicing analysis we choose to use rMATS [replicate-based Multivariate Analysis of Transcript Splicing]. rMATS detects and analyzes the events of major types of alternative splicing patterns (such as SE,A5SS,A3SS, MXE, RI) from replicated RNA-seq data. ***[13]*** Preliminary tests need to be carried out with a reduced dataset to ensure good results. Given that rMATS is sensitive to biological variability in the samples,firstly we must ensure that there are enough samples to enable rMATS to have a better estimate of it. Subsequently, we need to ensure that there is enough depth to reduce the uncertainty of the results. ***[14]*** Finally rMATS has several interesting associated software to use. Such as rMAPS and rmats2sashimiplots to visualize alternative splicing on selected genes. Or even rMATS - STAT to estimate the p_values for each alternative splicing event identified by rMATS and thus determine the statistical significance of the differences observed between samples. ***[13]***

4. Dataset

The dataset consists of 50 files in a compressed FASTQ format. 44 of those files are part of the 2023 dataset, while the remaining 6 form the 2020 dataset. Those files are raw data and no quality control has been processed. In these files, several types of RNA are present and derive from either : BHK, pDC or pDC+BHK. For each type of RNA, there are two conditions : infected or not infected. Finally, we deal with several replicates from three to four, in each condition. Certainly, the dataset also contains controls. We have a sample of non infected and non co-cultured pDC's RNAS, allowing us to have a reference condition to compare with. Another control is RNA from infected but non co-cultured BHK cells. This specific control will be determinant to cut away RNA from BHK cells in condition "pDC+BHK".

Dataset contain also contains a TSV format file containing information on each library, including name, cells, infection, data and replicates.

Project team

The team leading the project is the VIV Team [Vesicular trafficking, Innate response and Viruses], from the CIRI [International Center for Infectiology Research]. The CIRI is a joint research unit of the Claude Bernard University, the Inserm [National Institute for Health and Medical Research], the CNRS [National Center for Scientific Research] and the Lyon ENS [Superior Normal School]. They are a gathering of research communities around three related fields: bacteriology, immunology and virology. ***[5]***

The VIV team is part of the virology field, which focuses on both emerging, highly pathogenic viruses as well as more chronic ones; but they also rely on immuno-virology and cell biology. Their team studies the biological response to viral infections, most specifically the interferon signaling pathways. They aim to understand how viruses circumvent the antiviral response, and discover the alternative sensing pathways, such as how cell types specialized in the interferon pathways recognize viral infection in neighboring cells. ***[5]*** Among these specialized cells are the pDCs, the subjects of this project.

Objectives

There are two main goals for this project. To begin with, we must assess the quality of the data. Therefore, to carry out narrow analysis and accurate deductions we need to ensure that the quality and the coverage of the dataset is sufficient. This can be achieved with the use of FastQC and MultiQC. Through bibliographical research, we will also estimate the analysis accuracy achievable with our data.

Then, potential differences between the transcriptomes of naive and activated pDCs' suggesting a rearrangement due to alternative splicing will be investigated. Thus we will perform a differential expression analysis between the two transcriptomes and a structural comparison with IGV [Integrated Genome Viewer].

Finally, we will try to highlight the crucial role of post-translational regulation in the efficacy of antiviral innate immunity.

Final product

The final product must include a detailed and popularized report on the dataset's quality check. This report must specify if the dataset quality is considerable enough to carry out studies on alternative splicing. Should it pass the quality check, the report must also include the results of analysis on potential differences between transcriptomes of naive and activated pDCs.

Constraints

1. Cost

The project does not have an associated budget, and no software or hardware will be lent to our team. All the work must be carried out using our own equipment. However, the software required for analysis is available for free, so there are no real financial constraints.

2. Deadline

The final product must be delivered before the deadline, which is on March 28th, 2024. A project presentation will be held on the following day, March 29th, 2024. Six weeks will be freed and dedicated to project work, as well as for preparation for the presentation. Updates must be shared with both the tutor and the project leader, the frequency of which will be established after the project is initiated.

3. Complexity of the dataset

Parts of the dataset are made up of RNA that derive from both pDC and BHK cells. An initial step of data pre-processing is mandatory to accurately analyze alternative splicing specifically in pDC. Without this, BHK's RNA could potentially skew our analysis and allocate, erroneously, effects from BHK to pDC cells. In order to avoid this situation, we plan to realize quasi-mapping on human transcriptome. In this way, BHK's RNA, that comes from hamster species, will not be able to map on the human genome and will be excluded. Thus, through this preliminary step, alternative splicing events could be attributed exclusively to RNA that come from pDC.

4. Other constraints

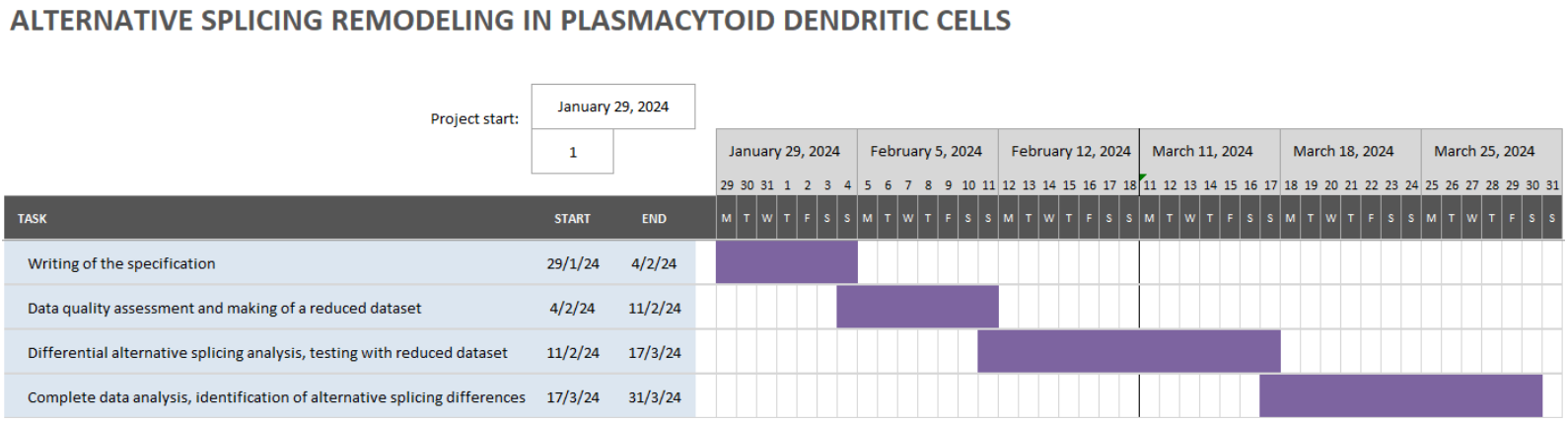
The product must be understandable and explainable to an uninitiated audience, for both the analysis and the following presentation. This requires a popularization of the work results.

Furthermore, a certain amount of time will be necessary to fully learn MISO and rMATS, how to effectively use these sophisticated bioinformatics tools but also understanding their underlying algorithms and interpreting the results they provide.

Planning

Overall, we have six weeks to conduct our project:

- First week: Writing of the specification.
- Second week: Data quality assessment and making of a reduced dataset.
- Third and fourth week: Implementation of a pipeline to carry out a differential alternative splicing analysis and testing with the reduced dataset.
- Fifth and sixth week: Analysis of the complete data and identification of alternative splicing differences between the samples.



References

- **[1]** Janeway Charles. *Janeway's Immunobiology*. 10th ed. W.W. Norton and Company 2022.
- **[2]** Assil, Sonia et al. "Plasmacytoid Dendritic Cells and Infected Cells Form an Interferogenic Synapse Required for Antiviral Responses." *Cell host & microbe* vol. 25,5 (2019): 730-745.e6. doi:10.1016/j.chom.2019.03.005. https://pubmed.ncbi.nlm.nih.gov/31003939/
- **[3]** (Not Published) Daligault, Camille, "Alternative splicing in the activation of the interferon expression during a viral infection". *Master BioSciences, Département de Biologie, École Normale Supérieure de Lyon*. (2023).
- **[4]** Wang, Zhong et al. "RNA-Seq: a revolutionary tool for transcriptomics." *Nature reviews. Genetics* vol. 10,1 (2009): 57-63. doi:10.1038/nrg2484
- **[5]** https://ciri.ens-lyon.fr/
- **[6]** Pai, AA, G Baharian, A Pagé Sabourin, JF Brinkworth, Y Nédélec, JW Foley, JC Grenier, et al. "Widespread Shortening of 3' Untranslated Regions and Increased Exon Inclusion Are Evolutionarily Conserved Features of Innate Immune Responses to Infection." PLoS Genet 12, no. 9 (September 30, 2016): e1006338. doi: 10.1371/journal.pgen.1006338. https://pubmed.ncbi.nlm.nih.gov/27690314/
- **[7]** Shen, Shihao et al. "rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data." *Proceedings of the National Academy of Sciences of the United States of America* vol. 111,51 (2014): E5593-601. doi:10.1073/pnas.1419161111. https://pubmed.ncbi.nlm.nih.gov/25480548/
- **[8]** Ye, Yishan et al. "Plasmacytoid dendritic cell biology and its role in immune-mediated diseases." *Clinical & translational immunology* vol. 9,5 e1139. 26 May. 2020, doi:10.1002/cti2.1139
- **[9]** Use of generative artificial intelligence (ChatGPT) to rephrase this sentence
- **[10]** National Human Genome Resaearch Institute - Genomic and Genetic Terms https://www.genome.gov/genetics-glossary/Alternative-Splicing
- **[11]** : FastQC website and documentation - https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
- **[12]** : Philip Ewels, Måns Magnusson, Sverker Lundin, Max Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics*, Volume 32, Issue 19, October 2016, Pages 3047–3048, https://doi.org/10.1093/bioinformatics/btw354
- **[13]** : Multivariate Analysis of Transcript Splicing (MATS) Xing Lab, Children's Hospital of Philadelphia: https://rnaseq-mats.sourceforge.io/
- **[14]** : Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A. 2014 Dec 23;111(51):E5593-601. doi: 10.1073/pnas.1419161111. Epub 2014 Dec 5. PMID: 25480548; PMCID: PMC4280593.
- **[15]** : Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. https://www.nature.com/articles/nmeth.4197