

**Papatheodorou Group**Supervisors: Yuyao Song, Dr. Irene Papatheodorou

Internship funded by the Embassy of France

**Anaëlle Cossard**

Paris-Saclay University

**Cross-species integration of single cell RNA-seq data from the primary motor cortex between human and mouse using expiMap and SATURN**

This internship takes place at the European Bioinformatic Institute (EBI), which is a part of the European Molecular Biology Laboratory (EMBL) and is permitted by the French Embassy in London partnership with the EBI. This partnership consists of offering some paid internships to French master's student, who are studying either computer-science, statistics, or bioinformatics. With a background in mathematics and computer science and currently doing a master's degree in bioinformatics and biostatistics, I thought this program would be a great enriching experience for me. Having shown a particular interest to cross-species analysis, I am grateful to have join Dr. Irene Papatheodorou's Gene Expression Team at the EBI for 3 months.

An aim of cross-species analysis is to study differences between the cell types across species, this could allow to understand cell types evolution over time and would be useful to transfer knowledge between different species. In the case of a disease, if a drug is tested on the mouse and has been proven effective, we could know if it would possibly be effective on the human by looking at the similarity between the targeted cell types. However, before being able to analyze those differences, an integration of cross-species data must be performed. Or to do so, we need a reliable tool able to map the same cell types from several species together while keeping the species-specific information.

Following this objective, Yuyao performed a benchmarking of different cross-species integration strategies of single cell RNA sequencing data, by testing different tools and parameters. It turns out that the tools based on neural networks were delivering the best results. That is why the purpose of my internship is to test a new tool called expiMap designed to integrate data from diseased and healthy patients, based on a neural network, which outperformed the most robust tool to perform cross-species integrations according to Yuyao's benchmarking.

In order to do so, we first selected 3 different datasets of scRNASeq data from the primary motor cortex of the human, mouse and fly. Second, I constructed an object making the link between the homologous genes of the human and the mouse, required to perform expiMap integration. Unfortunately, the results obtain with expiMap were not satisfactory suggesting that expiMap is not strong enough to perform cross-species integration, so we did not go any further. Moving on, I used SATURN, another new tool designed especially for cross-species, to performed new integrations giving this time great results.

I would like to thank my supervisors, Irene Papatheodorou and Yuyao Song, for allowing me to do this internship in a such supportive environment. I have learned a lot in various subjects: scRNASeq data integration, homologous genes, neural network and how to use new tools. However, the most important knowledge for me is what I learned about research in general: I am now aware that everything takes time, it is always helpful to talk with other people and it is normal to make mistakes, more we can learn a lot about it. Those are what I will keep in mind for, perhaps, my future PhD. I am very grateful to Yuyao Song for her attention and her precious advice throughout the entire internship. I would also like to thank Nadja Nolte, Anna Vathrakokoili-Pournara, Francesca Nadalin and Paula Cardenas for our enriching conversations. Finally, I am truly thankful to the Embassy of France in London for the financial support, that allowed me to get the most of this internship in Cambridge, hopping this partnership will benefit many other students that are looking to get into the research world.