



## Master project 2024-2025

### Personal Information

**Supervisor** Oleksandra Soldatkina

**Email** oleksandra.soldatkina@bsc.es

**Institution** Barcelona Supercomputing Center

**Website** <https://www.bsc.es/discover-bsc/organisation/scientific-structure/transcriptomics-and-functional-genomics-lab-tfgl>

**Group** Transcriptomics and Functional Genomics Lab

### Project

## Computational genomics

### Project Title:

Machine learning applications on biomedical research and image processing

### Keywords:

Computer vision, histopathology images, convolutional neural networks, disease, smoking, diabetes, gene expression.

### Summary:

Summary: The candidate will join Marta Melé's Transcriptomics and Functional Genomics lab in the Life Sciences Department at the Barcelona Supercomputing Center. The lab is interested in understanding how variation in the activation of our genes can explain differences between individuals such as differences in disease risk. In this project, we will perform computer vision techniques on hundreds of human tissue and organ images from diseases and healthy donors to classify individuals. We will start doing image feature extraction and classic machine learning approaches to confirm that specific clinical traits, such as diabetes and smoking, can be correctly classified using histopathology images from different human tissues. We will then train convolutional neural networks with our image data with the aim of obtaining a more accurate classifier. This project will complement and confirm some of the previous findings obtained in the lab using high-throughput transcriptomic analysis. As a future step, this project could be extended to combine high-throughput transcriptomic data with tissue images to associate genes' behavior with tissue morphology. In this way, we could pinpoint the responsible genes for particular disease symptoms, like immune infiltration. To do so, we would extract image features to correlate with gene expression matrices. What you will learn: Development of computational pipelines to analyze biological images working in High-Performance Computing environments. Effective communication of research findings to internal and external meetings, scientific writing, collaboration, and critical thinking.

### References:

References 1. Garcia-Perez R, Ramirez JM, Ripoll-Cladellas A, Chazarra-Gil R, Oliveros W, Soldatkina O, Bosio M, Rognon PJ, Capella S, Calvo M, Reverter F, Guigó R, Aguet F, Ferreira PG, Ardlie KG, & Melé M§. The landscape of expression and alternative splicing variation across human traits. § corresponding author. Cell Genomics 0: 100244.

### Expected skills:

Strong programming skills in bash, python, R, C++, or similar Some experience working in HPC clusters Excellent communication skills in spoken and written English Capacity to contribute to research projects with novel research ideas and analysis Capacity to work as a team in a highly collaborative and diverse environment Interest in biological and medical driven research Availability to start in July 2024 is preferred

### Possibility of funding:

Yes

**Possible continuity with PhD:**

To be discussed

**Comments:**

We are located at the 4th floor of the Barcelona Supercomputing center which has great views over Barcelona. We all go to work at least twice a week - most students go every day -but working from home is allowed a few days per week. For a list of selected publications from the lab see: <https://www.bsc.es/mele-marta/publications> All master students accepted in the lab will be offered a work contract as well as a practice agreement with the university.