Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Marta Mele
Email	marta.mele@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 systems biology

Project Title:

Unraveling the role of long non-coding RNAs in response to Ebola infection

Keywords:

Transcriptomics, long non-coding RNA, viral infections, differential gene expression, de novo transcriptome assembly, Ebola, alternative splicing, linear models.

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 the information encoded in our genome determines gene expression variation and splicing patterns across all cell types in our body. To address this question, we use large-scale transcriptomic analysis and the latest single-cell sequencing technologies combined with methods development to study gene expression, splicing and cell type composition variation across human tissues and phenotypes. We ultimately want to understand how these expression changes in both coding and non-coding genes are associated with specific disease states1,2. The candidate will focus on developing an integrative analysis of gene expression variation in non-human primate tissues that have been infected with Ebola virus. Specifically, he/she will focus on the analysis of long non-coding RNAs that have been shown to participate in immune system activation but have not yet been studied in depth in the context of viral infections. This is a unique unpublished dataset recently generated in collaboration with the Sabeti lab (Broad Institute of Harvard and MIT). Viral hemorrhagic fevers, such as Ebola, are infamous for their astonishingly high fatality rates, yet the precise mechanisms and pathways altered during infection still lack a comprehensive characterization. The few transcriptional studies carried out to date have been limited to a few cell types and to protein-coding genes3,4. A recent study has explored the transcriptional changes that occur upon Ebola infection at the single-cell level4, shedding light on Ebola tropism, its replication dynamics, and the elicited immune response. Although long non-coding RNAs (lncRNAs) are emerging as a key class of immune response regulators5, the role of the non-coding transcriptome in response to Ebola infection remains elusive. In our lab, we have studied the landscape of lncRNA expression in Chinese rhesus macaque (Macaca mulatta) over the course of Ebola infection6. Through a full de novo annotation, we have discovered over 10,000 novel lncRNAs which, along with previously annotated lncRNAs, serve as an atlas to study the non-coding transcriptome in macaque. Using PBMC single-cell data, we have found that lncRNAs are dysregulated upon Ebola infection across multiple immune cell types, in a cell type and disease stage specific manner, and that dysregulated lncRNAs are found close to protein-coding genes involved in immune response. In this project, the candidate will computationally analyze hundreds of bulk transcriptomes from 12 different tissues (adrenal gland, brain, kidney, liver, lung, lymph node, ovary, skin, spinal cord, spleen, testis, and whole blood) of macaques infected with Ebola at several time points. First we aim to detect lncRNA up-regulated or down-regulated in response to the viral infection, and identify early, mid and late onset lncRNA that could be the key players in the immune response to Ebola virus. Second, we want to discern between systemic and tissue-specific responses, and build gene expression networks that connect protein-coding genes with their lncRNA regulators. Ultimately, the question that we want to tackle in this project is what is the role of lncRNA in mediating the immune response elicited upon Ebola virus infection. What you will learn: Development of computational pipelines to analyze and interpret large datasets, especially from bulk RNA-sequencing and single-cell RNA-seq as well. Working in a High Performance Computing environment. Interpretation of multi-omics data. Scientific collaboration, effective communication of research findings into internal and external meetings, scientific writing and critical thinking.

References:

1. Melé, M. et al. The human transcriptome across tissues and individuals. Science (80-.). 348, 660–665 (2015). 2. Melé, M. et al. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. Genome Res. 27, 27–37 (2017). 3. Caballero, I. S. et al. In vivo Ebola virus infection leads to a strong innate response in circulating immune cells. BMC Genomics 17, 707 (2016). 4. Kotliar, D. et al. Single-Cell Profiling of Ebola Virus Disease In Vivo Reveals Viral and Host Dynamics. Cell, 83, 1383–1401, (2020). 5. Chen, Y. G., Satpathy, A. T. & Chang, H. Y. Gene regulation in the immune system by long noncoding RNAs. Nat. Immunol. 18, 962–972 (2017). 6. Santus, L. et al. Single-cell profiling of long non-coding RNAs during Ebola virus infection in non-human primates reveals dysregulation of the non-coding transcriptome in key immune-related pathways. In preparation.

Expected skills::

Strong programming skills in bash, python, R, perl, or similar Some experience working in HPC clusters Some experience with Next Generation Sequencing data analysis 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 Availability to start in July 2021 is preferred

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed