Project 7

Project title: Integrating single-cell transcriptional and epigenetic data to analyse plasticity in cancer

Lead group: Nicassio group, CGS- IIT

Partner group: Marioni group, EMBL-EBI

Rationale

In cancer, heterogeneity has been frequently observed. Cells from the same tumour can display very different behaviour, including the ability to resist conventional treatments or to promote tumour relapse or distant metastases (Meacham et al., Nature 2013). Recent developments of genome-wide technologies have provided us with tools to dissect at high-resolution the transcriptional and epigenetic blueprints of human tumours, thus identifying genes and pathways involved in tumour initiation and progression or those that are responsible for aggressive cancer phenotypes. Nevertheless, mechanisms at the basis of cancer cell plasticity and cancer evolution still remain largely unexplored. Intra-sample heterogeneity, rarity of the sub-population of interest and transcriptional noise are among the main obstacles in this quest. The recent development of single-cell technologies represents a powerful opportunity to fill this gap of knowledge. Single-cell transcriptomic profiling (**sc-RNA-seq)** and chromatin accessibility (**sc-ATAC-seq**) allow the transcriptional and epigenetic landscape to be characterised at high-resolution, even in rare sub-populations. Moreover, integrating data from these two technologies provides the potential to understand how changes in transcription are regulated, albeit computational methods for this are still in their infancy. Additionally, analytical tools to analyse and interpret single cell epigenetic profiles are underdeveloped relative to those used for the analysis of single cell RNA profiles [e.g. Seurat3, Monocle2, scran (Stuart et al., Cell 2019; Pliner et al., Molecular Cell 2018, Lun et al., Genome Biology, 2016)].

Aims

The aim of this project is to develop an innovative computational framework able to integrate single-cell transcriptional (**sc-RNA-seq)** and epigenetic (**sc-ATAC-seq**) data performed in parallel from the same sample. This computational approach will be tailored to the analysis of transcriptional heterogeneity, with a particular focus on human cancer using a model of breast cancer that mimics cancer cell plasticity developed in the host lab.

Objectives

The primary objective of this fellowship is the development of a toolkit for single cell genomics that allow the combined analysis of sc-RNA-seq and sc-ATAC-seq datasets. Such a method will represent an extremely valuable tool to study the regulatory framework of complex samples (i.e. heterogenous samples) by integrating different regulatory elements (transcription factors, enhancer, promoter) with transcriptional output. It will have great potential in discovering novel biological relevant mechanisms in a broad spectrum of contexts.

The core functionality will be the ability to bi-directionally transfer information across the different single-cell platforms (i.e. transcription to chromatin-accessibility and *vice versa*), facilitating the definition of cis- and trans- regulatory maps. Other features will also be implemented, depending on the fellow's scientific interests, background and expertise. Some examples are: i) integrative visualization tools, to jointly explore graphically sc-RNA-seq and sc-ATAC-seq profiles, ii) 'gene activity matrixes' (based on enhancer and promoter activity at chromatin level) iii) *cis*- and *trans*- regulatory networks (based on integration of gene, non-coding RNA expression levels, transcriptional factor motifs and footprints); iv) analysis of pooled CRISPR-mediated perturbations (implementing the feature barcoding technology).

We propose to train and validate the method focusing on cancer cell plasticity, by characterizing the population of cancer cells that evade chemo-treatment. The lead group has generated an *in-house* breast cancer model that mimics some aspects of the adaptive response and transcriptional reprogramming occurring upon chemotherapy treatment. A vast collection of genomics data has been already produced, using both bulk genomic analyses and single-cell technologies (drop-seq with the 10X Genomics Chromium platform), including sc-RNA-seq and sc-ATAC-seq profiles of more than 50.000 cells.

Integration of expertise of partners

The Nicassio lab based at Center for Genomic Science of IIT (located in Milan) has broad expertise in modern genomic technologies applied to the study of complex biological processes and diseases, with a particular emphasis on cancer. Related to the proposed project, the lab has generated the model of

chemo-adaptation that will be used. A vast number of genomic analysis have been performed and are already available, thus greatly facilitating the initial steps of the project.

The Marioni group, based at EMBL-EBI and at the CRUK Cambridge Institute within the University of Cambridge, have substantial expertise in the analysis of single-cell transcriptomics data, both in terms of methodological development (e.g., Achim et al., Nat Biotechnol, 2015; Vallejos et al., Nat Methods, 2017; Haghverdi et al., Nat Biotechnol, 2018) and in applications focused on cell fate decisions (e.g., Scialdone et al., Nature, 2016; Martinez-Jimenez et al., Science, 2017; Richard et al., Nat Immunol, 2018; Pijuan-Sala et al., Nature, 2019; Atzekan et al., Science, 2019).