Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Renee Beekman in collaboration with Francois Serra and Alfonso Valencia
Email	renee.beekman@crg.eu; francois.serra@bsc.es; alfonso.valencia@bsc.es
Institution	Centre for Genomic Regulation (CRG) in collaboration with the Barcelona Supercomputing Center (BSC)
Website	https://www.crg.eu/en/programmes-groups/beekman-lab; https://www.bsc.es/discover-bsc/organisation/scientific-structure/computational-biology
Group	Single Cell Epigenomics and Cancer Development (CRG) in collaboration with Computational Biology Life Sciences (BSC)

Project

Computational genomics

Project Title:

Unveiling differences between the 3D chromatin structure of homologous chromosomes in the context of translocations in tumour cells

Keywords:

3D Chromatin Structure, Chromosome/Allele-specific Read Mapping, De Novo Genome Assembly, Normal Cells and Tumour Cells, Lymphoma.

Summary:

Each mammalian cell has two copies of each chromosome. For the vast majority of computational analyses, the two copies of each chromosome are combined. However, it is generally known that many biological processes can show differences between the two copies of the chromosomes. A specific gene can for example be expressed from only one of the two chromosomes (allele-specific expression). Or, genetic mutations in tumour samples occur on only one of the two chromosomes (heterozygous mutations). In our group, we aim to distinguish information of the different copies of the chromosomes to better understand the development of cancer. We will do this in the context of lymphomas, which are tumours that originate from normal immune cells. While the chromosomes are considered linear structures, they are actually folded at the three-dimensional (3D) level in a highly organised way (Dekker et al. Nat Rev Genet. 2013). This organisation is needed for the chromosomes to regulate gene expression. Importantly, in lymphoma cells the 3D chromatin structure is altered in comparison to normal cells (Vilarrasa-Blasi & Soler-Vila et al. BioRxiv 2019). In our group, we aim to study the 3D chromatin structure in lymphoma cells in comparison to normal cells. More specifically, we study how genetic translocations (=a piece of one chromosome fuses to another chromosome) in lymphoma cells affect the 3D chromatin structure. In this project, we will use Hi-C data generated to study the 3D chromatin structure in normal cells (Rao et al. Cell 2014) and lymphoma cells (Vilarrasa-Blasi & Soler-Vila et al. BioRxiv 2019 and unpublished data). Hi-C is a molecular technique coupled to next generation sequencing that allows to reconstruct the 3D folding of the genome in the nucleus (Lieberman-Aiden et al. Science 2009). First, we will computationally separate the two homologous copies of chromosome 14. We focus on this chromosome as in lymphomas one of the copies of chromosome 14 is affected by a genetic translocation we aim to study. More specifically, we will use the variation in the genetic code between the two copies of chromosome 14 to distinguish them. To that end, we will use the genomic sequencing data of this chromosome in these samples and perform a de novo assembly to create a reference sequence for the two copies separately. Next, we will use this reference sequence to map the Hi-C reads representing the 3D chromatin structure to one or the other chromosome. Finally, we will reconstruct the 3D-chromatin structure using these separated sets of reads in TADbit (Serra et al. PLoS Comput Biol. 2017). From these reconstructed copies of chromosome 14 we will analyse the differences in the 3D chromatin structure in order to understand the effect of this genetic translocation specific to lymphoma on the 3D chromatin landscape surrounding it. What will you learn: • Computational biology:

basics on network analysis; collaborative software development using GIT; to design and use of computational pipelines for high performance computing (in the 30th most powerful supercomputer in the world). • Structural Genomics: to process and analyse data from Chromosome Conformation Capture techniques (mostly Hi-C and Capture-C). • Genomics and Epigenomics: to explore available data at the interface of genomics and epigenomics; to understand the basics of gene regulation mechanisms and to postulate hypotheses about deregulation of these mechanisms in cancer and test them by analysing the data. • Tumour Biology: to understand the genetic and epigenetic mechanisms underlying the development of lymphomas. • Scientific Dissemination: to present in lab meetings and to write a research article resulting from your work.

References:

Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet. 2013 Jun;14(6):390-403. doi: 10.1038/nrg3454. Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M. V, Ragoczy, T., Telling, A., et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 2009 Jul; 326, 289–93. doi: 10.1126/science.1181369 Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014 Dec 18;159(7):1665-80. doi: 10.1016/j.cell.2014.11.021. Serra F, Baù D, Goodstadt M, Castillo D, Filion GJ, Marti-Renom MA. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput Biol. 2017 Jul 19;13(7):e1005665. doi: 10.1371/journal.pcbi.1005665. Vilarrasa-Blasi R, Soler-Vila P, Verdaguer-Dot N, Russiñol N, Di Stefano M, Chapaprieta V, Clot G, Farabella I, Cuscó P, Agirre X, Prosper F, Beekman R, Beà S, Colomer D, Stunnenberg HG, Gut I, Campo E, Marti-Renom MA, Martin-Subero JI. Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. bioRxiv 764910

Expected skills::

A strong background in UNIX command line tools as well as in python or R programming, in combination with creative thinking and enthusiasm to work in a multi-disciplinary team with wet lab and bioinformatic experience.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Our lab in the CRG can be divided into two branches: a wet lab branch and a bioinformatic branch. A key aspect of our group is that these two branches are intermingled, whereby the different team members can interact on a day-to-day basis and during weekly lab meetings. On top of that, most team members will have shared wet lab-bioinformatic projects. Moreover, this project in particular will be conducted in collaboration with the lab of Alfonso Valencia in the Barcelona Supercomputing Center, giving a significant boost in computational power and bioinformatics expertise. We strongly believe that both wet lab and bioinformatic analyses and especially the interaction between these two fields are critical to better understand biological phenomena.