PROJET LONG

Sujet: Investigation bio-informatique du syndrome ICF par analyse de génome

Project proposal - Team of Claire Francastel (contact: Costas Bouyioukos - costas.bouyioukos@univ-paris-diderot.fr)
CNRS UMR7216 Epigenetics and Cell Fate – University Paris Diderot

Our genome contains all the information necessary for the development of a healthy subject. However, the correct exploitation of the information carried by our genome is insured by chemical modifications that are collectively defined as the "epigenome". DNA methylation is among the best-studied epigenetic modification in vertebrates. It is essential for normal embryonic development and perturbed DNA methylation patterns are hallmarks of many human diseases. Studies of inherited diseases with perturbed DNA methylation landscapes has recently identified new candidate players in DNA methylation pathways.
Work in our team focuses on a rare disease, the ICF syndrome (Immunodeficiency with Centromeric instability and Facial anomalies), caused by a remarkable loss of DNA methylation associated with genomic instability. Logically, mutations in a DNA methyltransferase were the first reported causes of the disease. More recently, mutations were found in factors with virtually unknown function. Importantly, these new "ICF factors" represent new candidate players to consider in pathways to DNA methylation, normal development and maintenance of genome integrity. This discovery raises important questions that still need to be addressed with regard to their function, but also to where and how these factors are targeted on the genome.
The project that we propose in the frame on the Master in Bioinformatics is to identify a sequence signature in the genomic regions affected by abnormal methylation levels as a consequence of mutations in ICF factors. The analysis will be based on the genome-wide DNA methylation analysis that we generated in cells from a mouse model. As a proof of concept, we will focus on one of these factors in the context of the mouse genome. We identified the genomic regions that loose or gain DNA methylation compared to healthy subjects. Webservers like MEME-ChIP already exist but they are not well-suited for this type of analysis. Hence, the student will have to propose a script which will take genomic sequences from regions of interest (i.e. differentially methylated compared to normal cells), compare them (GC content, search for known motifs, alignment with synthenic regions from other species) and try to identify de-novo genomic signatures i.e. significant motifs/patterns that are over-represented in the dataset. We will test the functionality of the script on available data for factors with known binding sequences. We will then apply this script for the "discovery" of binding sites for a factor for which we virtually have no information but which is important for genome stability.