JOB OPENING AT UPF BARCELONA

PhD/technical position at the Structural Bioinformatics group of GRIB

The Structural Bioinformatics group (SBI) is integrated in the Research Programme on Biomedical Informatics (GRIB), a joint research programme of the Hospital del Mar Medical Research Institute (IMIM) and the Department of Experimental and Health Sciences of the Universitat Pompeu Fabra, located at the Barcelona Biomedical Research Park (PRBB) in Barcelona (Spain).

Research Outline of SBI group (http://sbi.imim.es)

Knowledge of transcription factor (TF) -binding sites, the locations at which TFs bind to DNA in the genome, is key to understand how genes are regulated. Yet, the binding preferences of most eukaryotic TFs remain unknown. In this scenario, the development of computational tools as a complement to experimental procedures is fundamental. A major difficulty in studying TF-DNA binding specificity and in evaluating models for representing this specificity has been scarcity of structural data. We are involved in the development of a homology modeling-based approach that combines structural information and protein binding microarray (PBM) data to predict the binding preferences of TFs and model TF-DNA interactions.

For more information, visit our web at http://sbi.imim.es

We are seeking for a research student, interested in obtaining the PhD in bioinformatics, with capacity to offer technical support for the improvement of programmes generated in the group and the development of new ideas and services within the scientific subject of this announcement (see further).

DUTIES

- ✓ The successful candidate will be involved in the improvement, development and update of the tools of structural bioinformatics, with particular attention on the analysis of protein-protein and protein-DNA interactions in network analyses, implying new aspects of research within the topic of this opening (see further)

SKILLS AND EXPERIENCE

- ✓ Knowledge of structural bioinformatics techniques (docking, template-based homology modelling, sequence and structure alignment, etc.)
- ✓ Oral and written communication skills.
- ✓ Fluent spoken and written English.
- ✓ Excellent programming/scripting skills in at least one of each of the following bullet points:
    - Script/Programming languages: Python
        - Specificities: BioPython, Numpy, SciPy, TkInter.
    - Programming languages: C, C++, Java.
    - Web programming knowledge on at least one of the following points:
        - Angular JS, JavaScript, PHP.
        - Django, Bottle.
        - HTML5, CGI.
    - Databases in: SQL, MySQL, SQLite, MongoDB

- ✓ Graduate/engineer in one of the following areas: mathematics, physics, chemistry, biochemistry, informatics, human biology, biomedical engineering.
- ✓ Master or equivalent in bioinformatics or computational biology, or in the process of obtaining it.

WORK CONDITIONS

- ✓ Two-year contract with possible extension of one more year, up to the third year required to defend the thesis.
- ✓ Annual gross salary of 16,500€ with the possibility to add collaborations with SMEs in the field of Bioinformatics.
- ✓ Starting on February-April 2022

SUPPORT

The work is supported by grant: PID2020-113203RB-I00 from the Spanish Ministry of Economy (MICIN).

SCIENTIFIC TOPIC

Summary: Transcriptional regulatory elements in complex genomes are key players of the genome during development, cell and tissue homeostasis, responses to external stimuli, and disease. High-throughput experiments of genomics and proteomics have provided a plethora of activating regulatory elements, promoters and enhancers, of the genome. However, very few proteins in human occupy most of their motif matches under physiological conditions, which highlights the importance of the balance between the co-operativity of transcription factors (TFs), their strength upon binding and the environment (such as chromatin state and epigenetic marks). Co-operative recognition of DNA by multiple TFs defines unique genomic positions on the genome and confers a robust regulation. Most eukaryotic TFs recruit cofactors as ‘‘coactivators’’ or ‘‘corepressors’’ forming large complexes formed by protein-protein and protein-DNA interactions to regulate transcription. Of particular interest is the co-operation of “pioneer factors” (the first to engage target sites in chromatin, culminating in transcription by displacing nucleosomes). Protein-DNA binding strength has been of extraordinary importance for understanding co-operativity and transcription regulation. Variable TF-DNA interactions are increasingly considered as key drivers of phenotypic variation. Overall, about 19% of the human TFs are currently associated with at least one phenotype and a significant number of mutations associated with diseases are within or near genes encoding TFs. In the last years, different models for representing TF motifs, such as machine learning models combining multiple layers of genomic, transcriptomic, DNase hyper-sensitivity (DHS) and epigenomic information, were developed to score new putative regulatory sequences or calculate the 'gainability' and 'disruptability' of Single Nucleotide Variant (SNV) alleles. The strength of binding of TFs may not only be affected by mutations, but also by epigenetic modifications of chromatin, such as methylation of cytosine bases. With the emergence of new methods to outline DNA binding, methylation was found to be abundant throughout the genome except at active regulatory regions bound by TFs, such as gene promoters and distal enhancers. There are also few examples where modification of a single CpG dinucleotide directly affects transcription factor binding and regulation of a target gene in vivo.

Therefore, understanding the features that determine the binding strength between TFs and TF Binding Sites and the co-operation between TFs is crucial to explain gene cis-regulation and the genotype-phenotype relationship. The main objective of our proposal is to understand, model and predict the effect of changes, genetic and/or epigenetic, in the DNA sequence. Our study will help us to understand the mechanisms of pioneer factors and on the re-design of DNA-binding proteins, not only TFs but also integrases, transposases and nucleases, with increasing relevance in genome-editing. Some examples are the re-design of homing endonuclease I-MsoI or the potential use of PiggyBac (PB) cut-and-paste DNA transposase in genome-editing. The lack of a DNA footprint left behind after piggyBac transposition has boosted its exploitation for genome

engineering, while nucleases have been used since the early applications in genome-editing and gene therapy in combination with Zinc-fingers or CRISPR.

HOW TO APPLY

Applications for this opening should include a CV, a motivation letter and the names of two potential referees. Send them by **e-mail** to baldo.oliva@upf.edu with the subject “SBI position”

Reference: **SBI494.2**

For more information, contact Dr. Baldo Oliva (baldo.oliva@upf.edu) or visit our web site at http://sbi.imim.es