Postgraduate student: **George Georgakilas**

Thesis Title:

Prediction of microRNA transcripts from genomic data

Abstract:

microRNAs are sort length (~22 nucleotides) endogenously produced RNA molecules which regulate gene transcription by binding, in a sequence related way, on 3' UnTranslated Region of messenger RNA. There are lots of microRNAs in eukaryotic cells, regulating a wide variety of gene-targets. During the past few years, miRNAs have been related with the regulation of many biological processes.

This thesis is part of a larger project, whose goal is to identify mechanisms that regulate microRNA transcription and therefore place them in a wider biological pathway of gene transcription and regulation. The transcription start sites of these microRNA genes are largely unknown. In order to discover this currently unknown scientific area, certain models need to be developed, whose training is based on already largely explored scientific areas with common features with the area that is explored, test these models and finally evaluate them on novel experimental data.

The goal of the thesis is to develop a model, on data based on known protein coding gene transcription start sites. For this, we use information related to the methylated version of a histone (protein), called H3K4me3, which has the ability to bind on DNA regions that are transcription start sites. Out of these data, the appropriate features are extracted, and the model is trained and tested.

The results are very good, since even the simple additive algorithm accomplishes a high prediction performance level (precision up to 70%) while on the same time by using a more advanced machine learning algorithm this performance level is enhanced even more.

Examining Committee

Artemis Hatzigeorgiou, Researcher B', BSRC Alexander Fleming (Supervisor)

Ioannis Emiris, Professor, Department of Informatics and Telecommunication, UoA

Spyridon Garbis, Researcher B,' Biomedical Research Foundation, Academy of Athens