

Research Topics - Computational Genomics

1. DNA Sequencing

a. An Artificial Sequencing Laboratory.

DNA sequencing is of vital importance for (medical) diagnostics and understanding the working of the genome. Sequencing technology, however, is not flawless; it consists of several steps, and mistakes in previous phases may have a knock-on effect on later stages of sequencing. Systematic studies to gauge the impact of technological defects and varying parameters of the procedures on the quality of the sequencing output are difficult and costly to run. Therefore, it is worthwhile to do experiments with a simulated sequencing procedure. Currently, a PhD project on the effects of library preparation (including fragmentation of DNA and the cloning of such fragments) in such an “Artificial Laboratory” is carried out and we are looking for a follow-up to address issues of actual sequencing.

b. Spiking Neural Networks for Motif Detection in DNA.

Many computational methods to detect motifs in the sequence of nucleotides in DNA have been developed, but the application of spiking neural networks has received almost no attention. This project addresses this omission by investigating the pros and cons of a spiking neural network and its possible extension to deep learning for identifying motifs in DNA.

2. Gene Regulatory Networks

a. Neural Network Models for Gene Regulation.

The temporal (in-) activation of genes rather than their mere presence is essential for the working of the genome. The machinery regulating gene activation takes the form of a network of genes, which are activated by particular proteins (Transcription Factors) produced by the genes themselves and by those to which they are connected. Although much has been done to recognise patterns in the structure of these networks, there is a growing need for stronger theoretical constructs to understand their dynamics. This project aims to deliver these and novel insights by expanding analogies between real, biological neural networks and gene regulatory networks, use them as the basis for models and reflecting on their implications (e.g. can genomes “learn” and if so what does this mean?).

b. Dynamics of Gene Regulatory Networks underlying Epigenomic Landscapes.

It is increasingly acknowledged that other genomic processes than just mutations are shaping the development of a phenotype. The iconic explanatory framework is Waddington’s depiction of development as a pathway through the valleys of a canalised “epigenetic landscape”.

This project is about linking the topography of the landscape to the topology of an underlying gene regulatory network by viewing a lattice of genes beneath the surface of the landscape, of which the depth at each point represents the number of connections the gene (at the matching point of the lattice) has with other genes. The sequence of activations of genes traces a route through the network corresponding to the valleys in the landscape and of which the direction is determined by a series of Hopfield networks.

All projects require a sufficient background in biology (or the willingness to acquire it) and strong computational capacities from the candidate. The latter include programming, mathematics and statistics for all projects and dynamical systems theory and machine learning in particular for respectively projects (2b) and (1b, 2).