

CS84300 : Computational Molecular Biology

Course Description

Genome sequencing technology is now affordable and widely accessible for clinical applications, drug discovery research, and for independent researchers running small laboratories. Furthermore, with the broad availability of genome sequencing technologies and reduction in cost, tens of PetaBytes of genomic data have been generated and deposited in public databases during recent years. Currently, while complexity metrics such as clustering, centrality and other techniques have been applied to epigenetic interaction networks using the publicly available NGS datasets, no study has been performed for analysis of the networks from an algorithmic, computability or information theory perspective. A key hypothesis we will explore within this course, is whether there is a minimal set of epigenetic modules that catalyze the formation of the complete network, and which can be considered as information processing hubs that code the smallest biological programs or algorithms evolved as key factors of epigenetic control. These minimal modules can be identified as the non-repetitive, non-compressible network structures, when algorithmic information theory and Kolmogorov complexity metrics are applied to epigenetic network datasets.

Dr. Konstantinos
Krampis

Phone

212-896-0461

Email

kk104@hunter.cuny.edu

Office Location

413 E 69th Street
BRB-467F,
Belfer Building,
4th floor
New York, NY 10021

Office Hours

Thurs. 4-6pm

Course Topics

Part 1 Introduction and Basic Bioinformatics Analysis

- 1.1 Introducing genome sequencing technology and comparison of platforms
- 1.2 Genome sequencing applications in health and disease
- 1.3 Meta-genomics, synthetic biology & designer genomes
- 1.4 Genome assembly algorithms, graph data structures, core assembly statistics and evaluation
- 1.5 Big Data visualization and interpretation of genome assembly
- 1.5 Data types, sizes and files formats, computational and data storage requirements

Part 2 Computability and Complexity Modeling of Biological Systems using NGS data.

- 2.1 Sequential Dynamical Systems and Cellular Automata for cellular network modeling.
- 2.2. Information content and Kolmogorov complexity of chromatin conformation networks.
- 2.3. Comparative analysis of biological networks through category theory.
- 2.4. Biologically Inspired Algorithms, Novel DNA and Cellular Computing Models.
- 2.6 Computing with engineered or unmodified gene, signaling and epigenetic networks in the cells.
- 2.7 Computability and complexity of DNA computing, decidability of biological algorithms.

Grading Plan

The final grade of the class consists of the following: 70% weekly homework (consisting of weekly data analysis or research / literature review assignment) and participation in the class discussion and presentations, 30% final project.