CS196-1: Algorithmic Foundations of Computational Biology
CIT 368 · Tuesday and Thursday, 9:00-10:20

Course Syllabus

1. **Introduction**. Comparative genomics: genomes (DNA and protein sequence), protein structures (geometry), gene regulation (logic, systems), immunology (systems). The nature and complexity of bio-molecular data. The intertwining of algorithms and statistics in the design of genomics tools. The "Gold-Bug" – a metaphor for Bioinformatics.

2. **Genomics**.

   - **Alignment of two bio-molecular sequences**. Local and global alignment. Dynamic Programming algorithms. Edit graph theory and visualization of alignments. The fundamental Dynamic Programming recurrence. The Smith-Waterman algorithm. Probability and statistical significance. Evolutionary models. Information theory and the genetic code. The PAM matrices of Margaret Dayhoff, the "mother and father" of Bioinformatics. Statistical assumptions for bio-molecular data. Statistics hypothesis testing. How Sir R.A. Fisher caught Mendel "cheating."
   - **BLAST**. An outline of the BLAST statistical theory. Algorithmic speed up: a linear time approximation of the quadratic Smith-Waterman algorithm.
   - **Gene prediction. Hidden Markov Model algorithms**.
   - **Genome Assembly**. Assembly algorithms. Comparing assemblies: Of Mice and Dogs and Chimps and Men.

3. **Genomic Regulation**. Regulatory motifs. Transcription factors. Position weight matrices algorithms. Sea urchin - the First Genome of genomic regulation. A visit to the Sea Urchin Assembly. Suffix trees data structure and algorithms. Compressing genomic regulatory information. Designing DNA arrays.

4. **Protein folding**. The computational protein folding problem. Secondary structure prediction algorithms. Classification of protein folds. Protein structure alignment algorithms. Protein missfolding and the Mad Cow Disease.

5. **Genetic variation**. Single Nucleotide Polymorphism. Haplotypes. Informative SNPs. The Minimum Informative Subset Problem. Guilt by association. Statistical power and disease associations.

6. **Systems Biology**.

   - **Biological complexity**. Complex systems and Herbert Simon.s Hora and Tempus problem.
   - **Human and pathogens**. Comparative immuno-peptidomics of human and their pathogens. A tale and a tour of two genomes: the virus genome and the bacteria genome. Do pathogens evolve their proteome to evade the human immune system?
   - **Cancer genomics**. Tumor complexity.
   - **Gene regulatory networks**. Logic functions of genomic cis-regulatory code. Davidson vs. von Neumann: an information processing parallel between the genomic regulatory system and the nervous system.