CS123A Bioinformatics

Module 1 – Week 1 – Presentation 1

Leonard Wesley

Computer Science Dept

San Jose State Univ

Agenda

- Intro To Course
- Intro of Instructor
  - Background
  - R&D Projects: Current & Future
  - Student Funding Opportunities
- What Topics The Course Covers & What Will You Learn
- Greensheet
- Canvas
- Q&A

Introduction To Course

FROM THE CATALOG:

Introduction to the main public domain tools, databases and methods in bioinformatics. Analysis of algorithms behind the most successful tools, such as the local and global sequence alignment packages, and the underlying methods used in fragment assembly packages. Solution of complex biological questions requiring modification of standard code.

Introduction of Instructor

- My education and background
- Northeastern U., Boston MA – BA Physics
- U. Massachusetts, Amherst, MA – MS & CS Computer Science (PhD Computer Vision)
- SRI International (AI Center), SJSU (CmpE Dept)-> (Gen Eng) -> (CS Dept)
- Johns Hopkins U., Baltimore, MD (Online) – MS Bioinformatics (expected ~2020).
- Current R&D Interests:
  - ML for small molecule-protein drug affinity prediction
  - Protein structure scoring
  - ML & Evidential Reasoning to predict multifocal disorders & MCI (Alzheimer's)
  - ML prediction of pancreatic cancer
  - Characterizing and comparing power of ML algorithms.
  - AI-based search for Bioinformatics related literature
  - Web and mobile-based interfaces to some of the above applications
  - HPC-based implementation of above.

Student Funding Opportunities

- Current (Spring 2020): Paid
  - Liquid biopsy biomarker pancreatic cancer prediction
  - MRI & ML-based analysis of dementia, MCI, & Alzheimer's related diseases

- Future:
  - HPC related work, e.g., predicting small molecule protein affinity activity based on chemical structure & solution environments.
  - Building Web-based and mobile app for trans-cranial stimulation
  - Any of the R&D projects previously listed

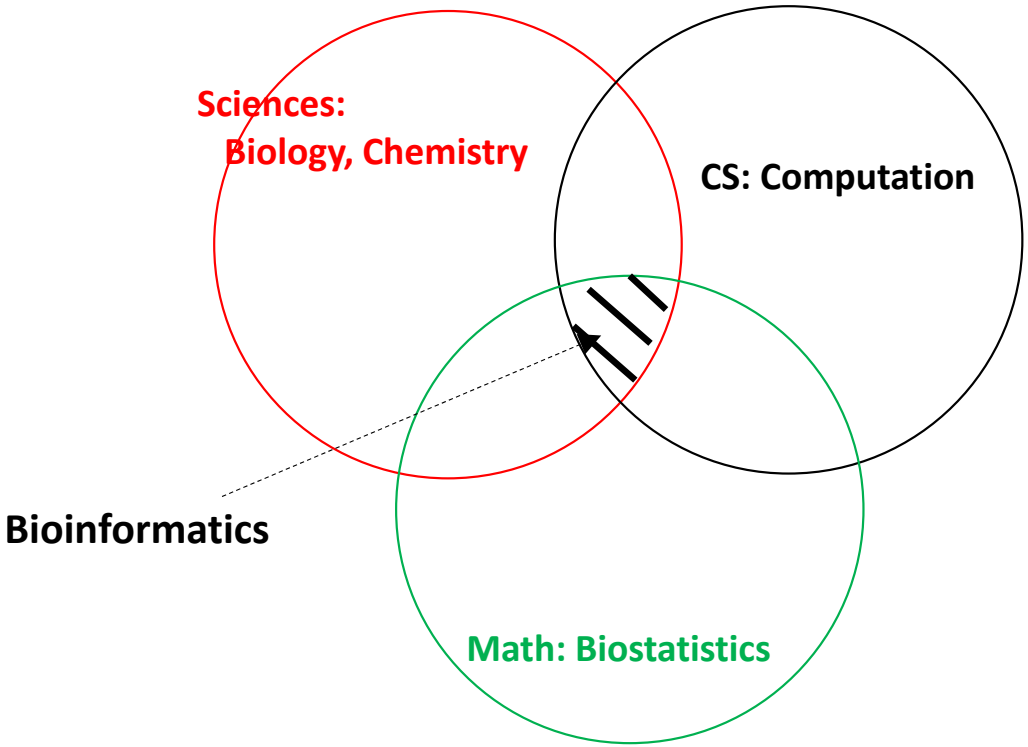
What Topics The Course Covers & What Will You Learn?

- SLO-1 BIOLOGY BASICS & BIOINFORMATICS DBs: Describe the structure of bioinformatics-related DBs and how they function to analyze sequence and related biological data. Navigate through various DBs to research and answer questions of interest, identify genes, and analyze complex genomes.
- SLO-2 ALIGNMENT: Describe and use pairwise and multiple sequence alignment algorithms to conduct local, global, and semiglobal alignments. Understand and use BLAST and advanced DB searching.
- SLO-3 PHYLOGONY: Build, understand, and use molecular phylogenetic trees. Understand and answer questions about evolution using molecular phylogenetic trees.
- SLO-4 PROTEINS & FUNCTIONAL GENOMICS: Understand protein analysis, proteomics, and functional genomics.
- SLO-5 NGS: Describe, understand, and analyze state-of-the-art technologies such as next-generation-sequencing (NGS) and genome assembly.

Course Introduction (cont.)

- Greensheet
- Canvas
- Q&A

The Three Pillars Of Bioinformatics



Bioinformatics Can Help Address Many Different Problems

- What are the causes of diseases
- Possible causes of a disease
- Move toward personalized medicine
- Role of genes/proteins in disease pathways

Bioinformatics Can Help Address Many Different Problems (cont.)

- Population genetics
- Organization of biological data
- Improve speed and accuracy of computations
- Identify new (reclassify) species
- …. *and more*

Identifying Genomes Is A Huge Current Activity

- ~30-40K human genes → hundreds of thousands of proteins
- Fruit Fly has ~13K genes
- Roundworm ~18K genes
- ~98% of human DNA identical to chimpanzee DNA

Drug Discovery & Development Is Another Active Area

- Takes 10-15 years and > $1 Billion to make one successful drug
- 95+% of candidate drug compounds and biologics fail at some point along the drug development pipeline.
- Bioinformatics helps along the entire pipeline

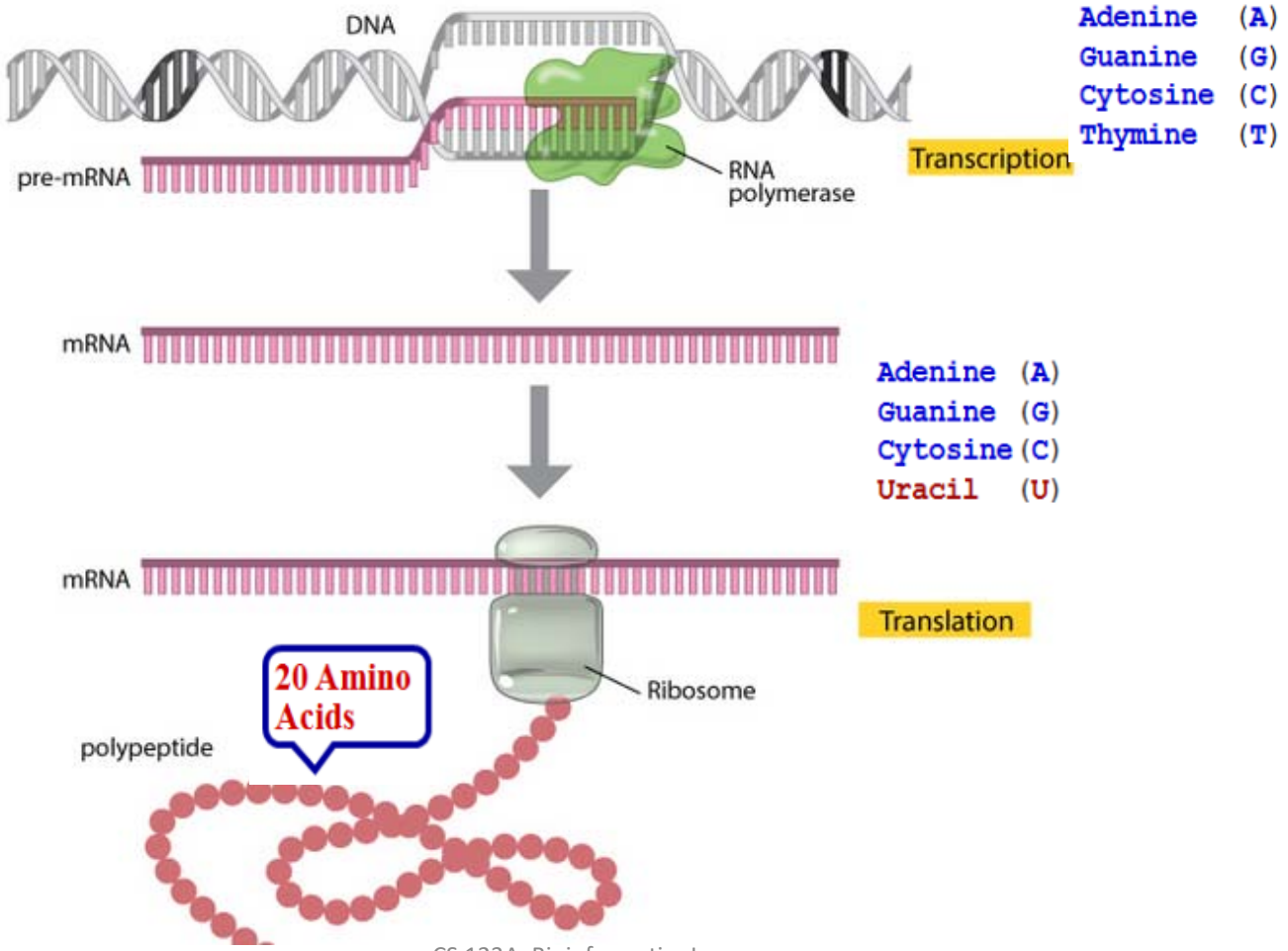
Definition of Bioinformatics

- Bioinformatics –

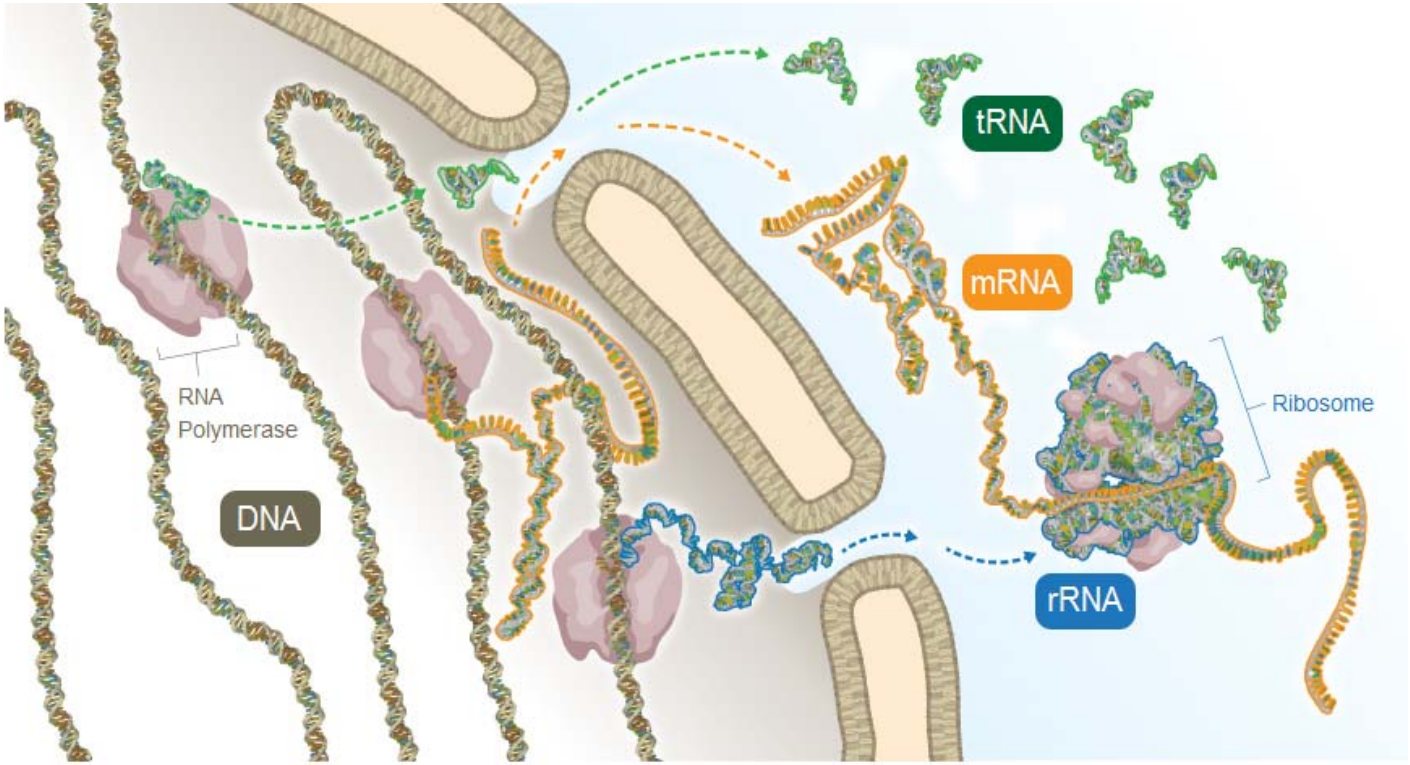
  is an applied science field that involves the management, manipulation, analysis, and interpretation of biological related information.

Review

DNA Is The Blueprint For Life



Central Dogma Of Molecular Biology



Proteins are the workhorse of life: structure, regulation, transport, defense, enzymes

Prokaryote vs Eukaryote

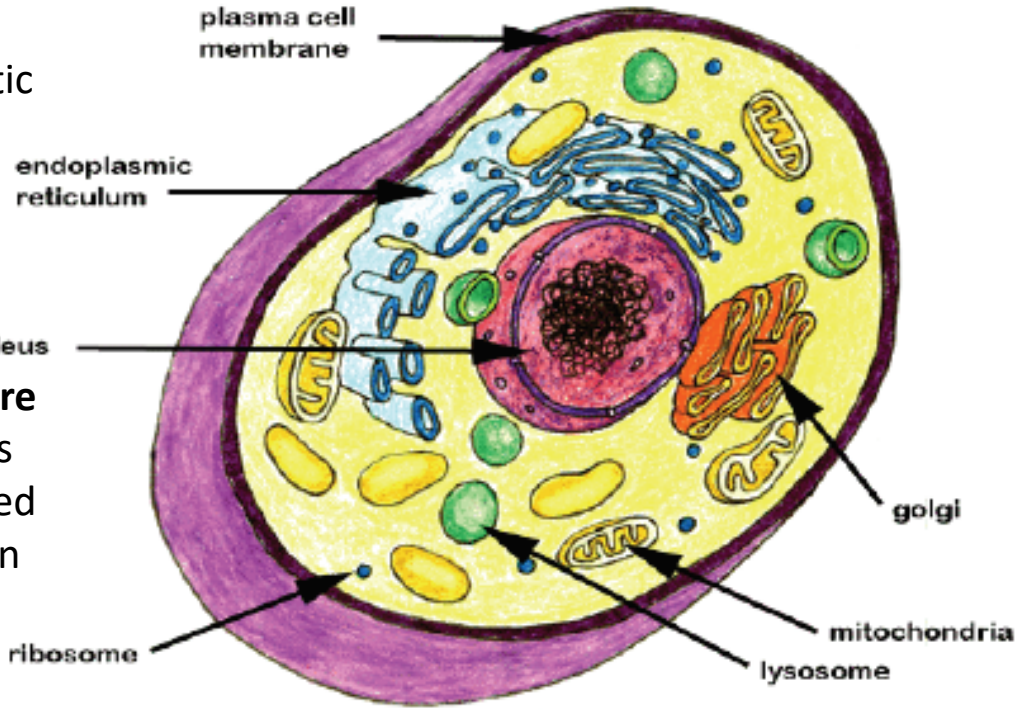
- The cell is the “atomic” unit of life

- Two types of cells
  - **prokaryotes, which are single-celled**
    organisms with no cell nucleus: archea and bacteria.
  - **eukaryotes, which are higher level organisms,** and their cells have nuclei: animals and plants.

Eukaryote Cell

A **cell carries the** entire set of genetic instructions: the genome, that makes an entire organism.

The **instructions are** encoded in DNA as genes and packaged as chromosomes in the nucleus.



Proteins & Nucleic Acids

- All living organisms have a similar molecular chemistry.

  The main actors in the chemistry of life are molecules:
  - – **proteins: which are responsible for what a living being is and** does in a physical sense.
  - "We are our proteins" R. Doolittle.

- – **nucleic acids: which encode the information necessary to**
  - produce proteins and are responsible for passing the "recipe" to subsequent generations.
  - Living organisms contain 2 kinds of nucleic acids:

- – **Ribonucleic acid (RNA)**
- – **Deoxyribonucleic acid (DNA)**

STOP HERE

See Reading & Homework Assignment #1 On Canvas