Introduction to Bioinformatics
CSC 314

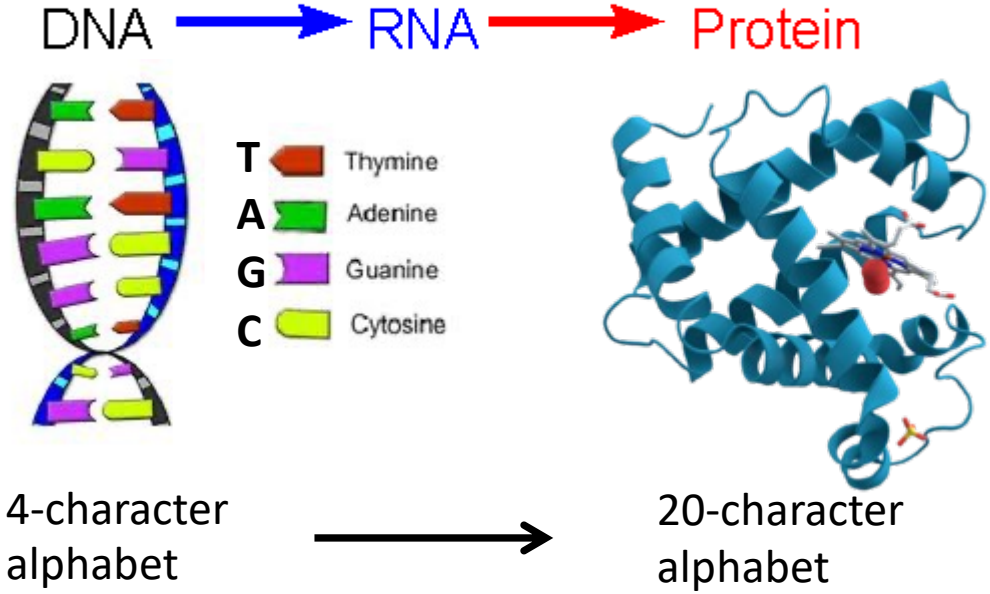
Spring 2025

Dr. Garrett Dancik

Course notes: https://gdancik.github.io

What is bioinformatics

- Bioinformatics:
  - Biology + information
  - the study and utilization of methods for storing, retrieving and analyzing biological data

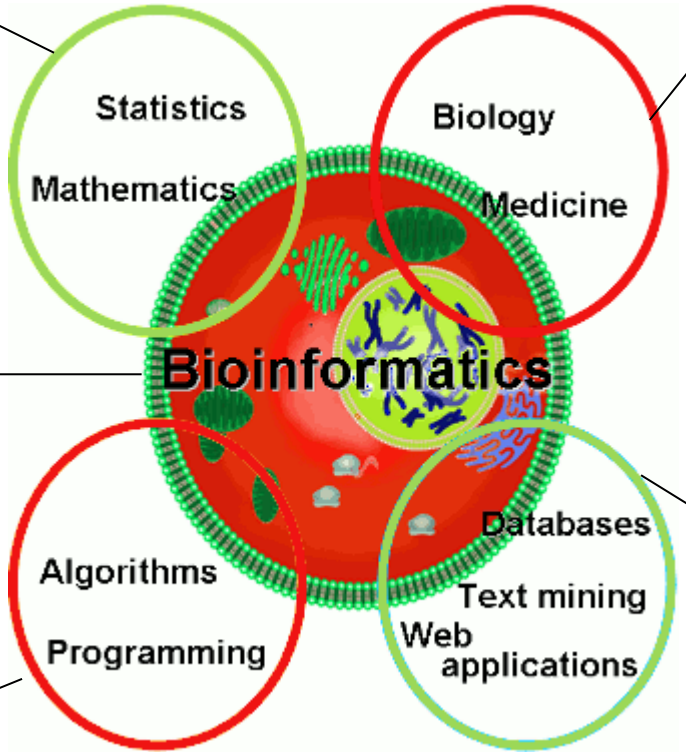


- How much information:
  - Human genome: 3 billion nucleotides
  - ~20,000 genes
    - many more when considering “junk DNA” and alternative splicing
  - >10 million sites of DNA variation
  - Countless possible interactions between DNA, RNA, and proteins

Bioinformatics is interdisciplinary

Calculus
Probability
Hidden Markov Models

Biology
Genetics
Chemistry
Physics



Applied bioinformatics
Methods in bioinformatics
Genomic data analysis

Database design
Database applications
Database searching

Computer programming
(Python/Biopython)
Computer algorithms
(Dynamic programming)

Source: http://www.amberbio.com/

What is this?

```
// Java code to output “Hello, World!”
public class helloWorld {
   public static void main(String[] args) {

        System.out.println("Hello World!");
   }
}
```

```
// Python code to output “Hello, World!”
print('hello world!')
```

What is this?

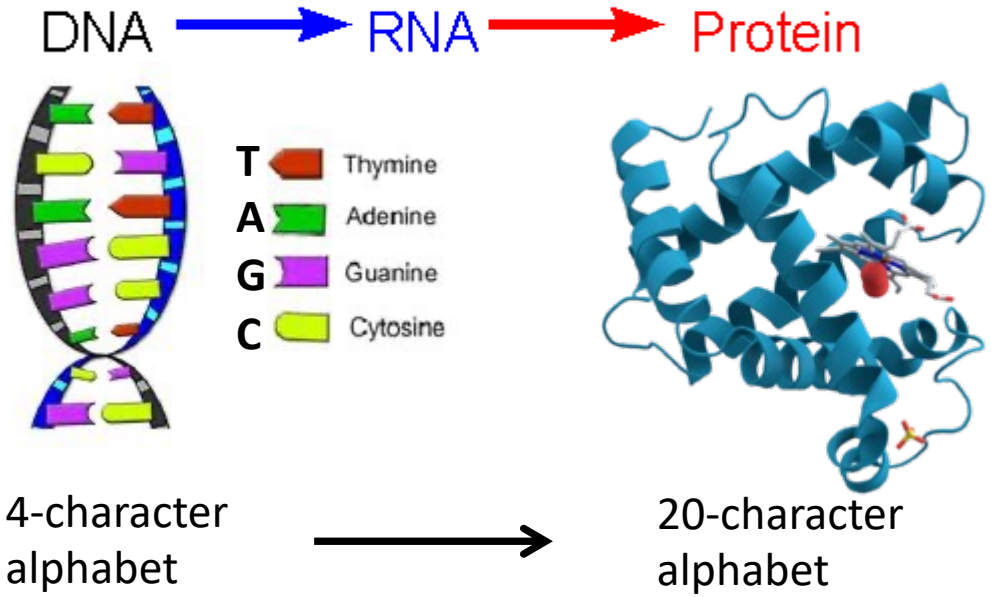
>NG_007114.1 Homo sapiens insulin (INS), RefSeqGene on chromosome 11

GGCGGCCAGGGAAGGTCTCTGCCGCCAGGGAAGTGTCCCAGAGACCCCTGGAGGGGCTGCTGACACCCCC
GGTGCCCCCACCTCGAGCATGACCCAGGGCTGCCTCTCCCCATCCTTCATCCTCCCTGCTCCACAGGACA
TTGGCCTGGCGTCCCTGGGGGCCTCGGATGAGGAAATTGAGAAGCTGTCCACGGTGGGTTGACCCCTCCC
TGCAGGGCCTGGGGTGTGGGTTTGGGGGTCTGAATCCAGGCCTCACCCTCTTGCCGTCCAGGCTGAGGCC
TCTCCTTCCACCCACGAATTGTGACCCTCACCCTGGCCTGCCTGCATCCTGGCCTGGCCTCCCTGGGGGT
GGTATCCTGGTCACGGGTGACCAGGGGCTGCCCGGTGGGCGGCAGCTGTCTCTGGGCTGATGCTGCCCGG
CTTCCCCGCAGCTGTACTGGTTCACGGTGGAGTTCGGGCTGTGTAAGCAGAACGGGGAGGTGAAGGCCTA
TGGTGCCGGGCTGCTGTCCTCCTACGGGGAGCTCCTGGTGAGAGTCTCTCCTTGCTGCAGCCCCCAGCAG
AGGGGCAGGGCTGGGGGACGGTGCAGGGAGGGGACAGGCTCCCAGTGGGAGGAAACTGAGGCCTGGACCT
CCAGGACTCAGGCTCTGTTTGGGAGAAGGCTTGTCTCTGCCCAGTCCTCACCCCACATTATCCCAGGCCT
CCGAAGGCCCGGCGGGGGAGATGGGGGTGACTCTACCCAAGGAACCCACCCAGCGTCAGGCCACGGTGCC
CCAGTTCCCTCGGGGACCTGGGTGCAGTGGAGTCAGTGATGCCATTGGCCTCCTGCCAGCACTGCCTGTC
TGAGGAGCCTGAGATTCGGGCCTTCGACCCTGAGGCTGCGGCCGTGCAGCCCTACCAAGACCAGACGTAC
CAGTCAGTCTACTTCGTGTCTGAGAGCTTCAGTGACGCCAAGGACAAGCTCAGGTGGGCTAGGCTGCTAG
GGCAAGCCCCCCATGGTGCCCCCAAACTGGGCCAGCCAGGCCTTCCTTCTGGCCTTGAGCAGGGCTGGAC
CTGTGAGCCCAGGTCACAGATGAGAAAACCGACCCCTGGTTGCAGCAGCCCCCACACAGCAGGGACACCA
TCCGTGAGAAGGACCCCAGCGTCTGGGGAGGGGCAGACCTACAGGACTGGGGGCTGCTGGGTGGCCGGGT
CAAGGCCAGTCTTGGAGGTGCTGACAGAGCCTGAGCTTTGTGAGGACGTCCTGTGGAACCTGTCCCGGCC
CCCTGCCCTGGGATGGGGAGAAGTCAGGGGGATAGACAGAGTCAAGGTGGGGGACAGGGCGGGAGTGGGG
TCCCCAGGGCTGGGGGCCTTTGGTGCAGTGACCAGAGTGTCAGGAGAGGGGAGCAAAGCCCTCTAGCCTC
ATCCTCATAAAAGGTCTCATCATTTTCCCTCCAGCCTCTTATGCACTGGGGAAACTGAGGCCAGGGGCTA…

Bioinformatics is an information science

- Computer code is a *set of instructions* that tells a computer how to process data and output results
- The genetic code is also a set of instructions, that tells a cell how to produce a molecule (RNA/protein) from DNA
  - Information flows from DNA → RNA → protein
  - The DNA information determines the structure/function of the RNA and protein

Central Dogma of Molecular Biology



- The function of a protein can be predicted from its DNA or protein sequence
- Just like Java (or Python) is a language for computers, genetics is the language of life (DNA is the alphabet)
  - This is a fundamental concept in bioinformatics

Intro to Genetics (Genetics 101)

- What are genes?
  - http://www.youtube.com/watch?v=ubq4eu_TDFc
  - Genes are part of what molecule?
  - How many possible bases (characters) are found in DNA?
  - How are genes organized?
  - How many *pairs* of chromosomes do humans have?
- What are SNPs?
  - http://www.youtube.com/watch?v=tJjXpiWKMyA
  - What is the human genome?
  - What is a SNP?

Intro to Genetics (Genetics 101)

- Where do your genes come from?
  - http://www.youtube.com/watch?v=-Yg89GY61DE
  - Where do your genes come from?
  - What are homologous chromosomes?

Bioinformatics Preview

- Let's look briefly at the genome of SARS-Cov2, the virus that causes COVID-19:
  - https://www.ncbi.nlm.nih.gov/nuccore/NC_045512
  - This is a preview and will make much more sense by the end of the semester