

Introduction to Bioinformatics

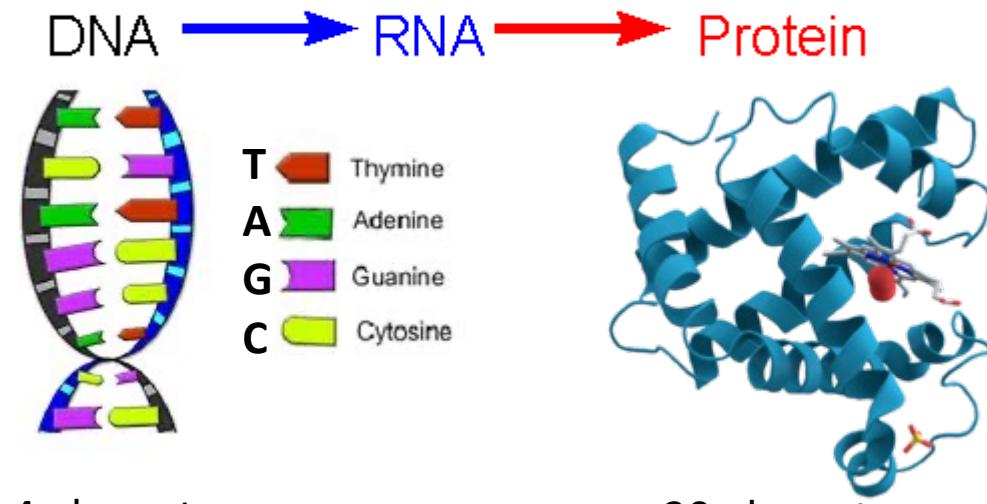
CSC 314

Spring 2026
Dr. Garrett Dancik

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

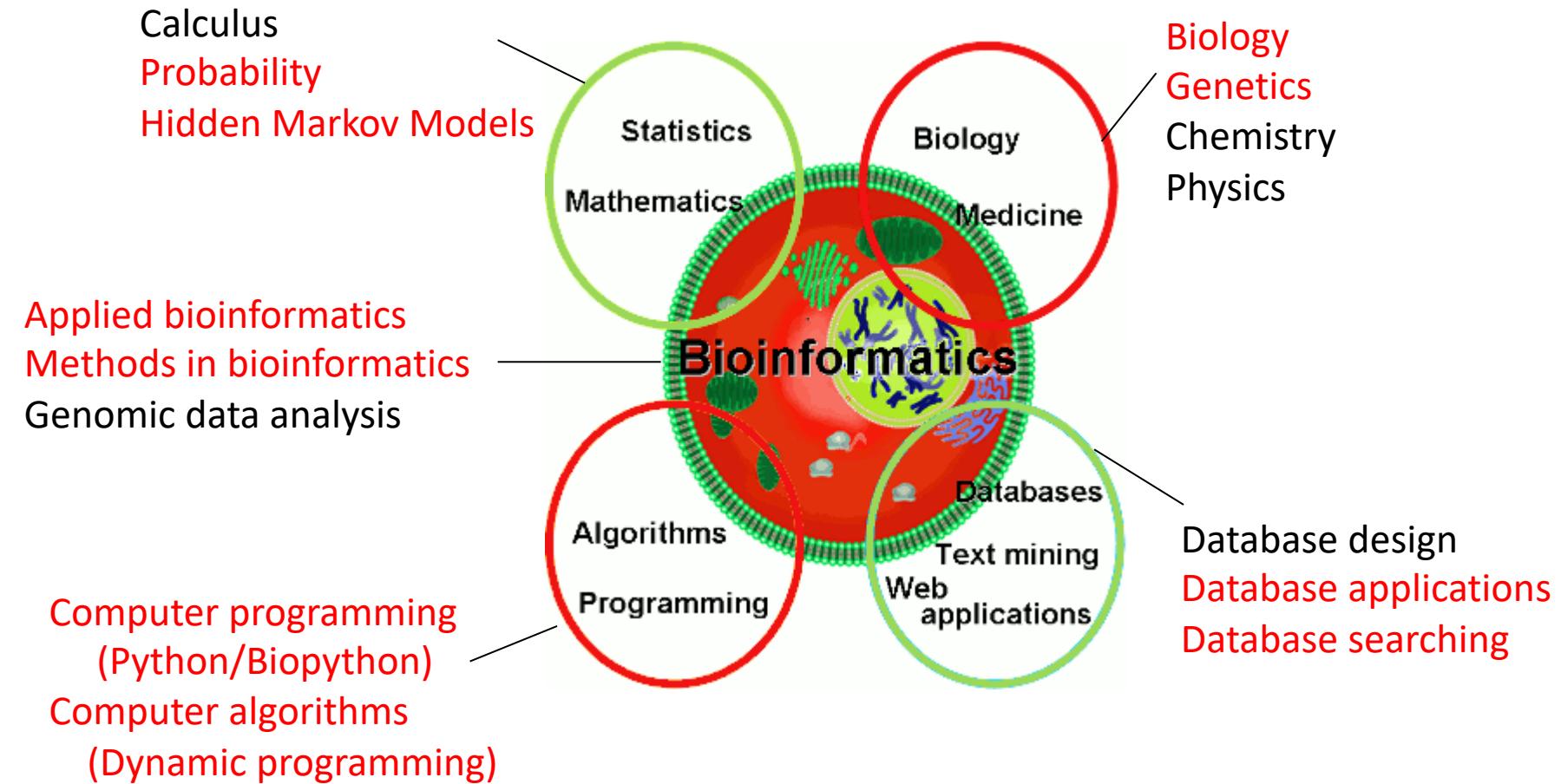
What is bioinformatics

- 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



What is this?

```
// Java code that instructs the computer  
// to output "Hello, World!"  
  
public class helloWorld {  
    public static void main(String[] args) {  
  
        System.out.println("Hello World!");  
    }  
}
```

```
// Python code that instructs the computer  
// to output, "Hello, World!"  
  
print('hello world!')
```

What is this?

// A “genetic code” that instructs a
// cell to produce the protein insulin

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

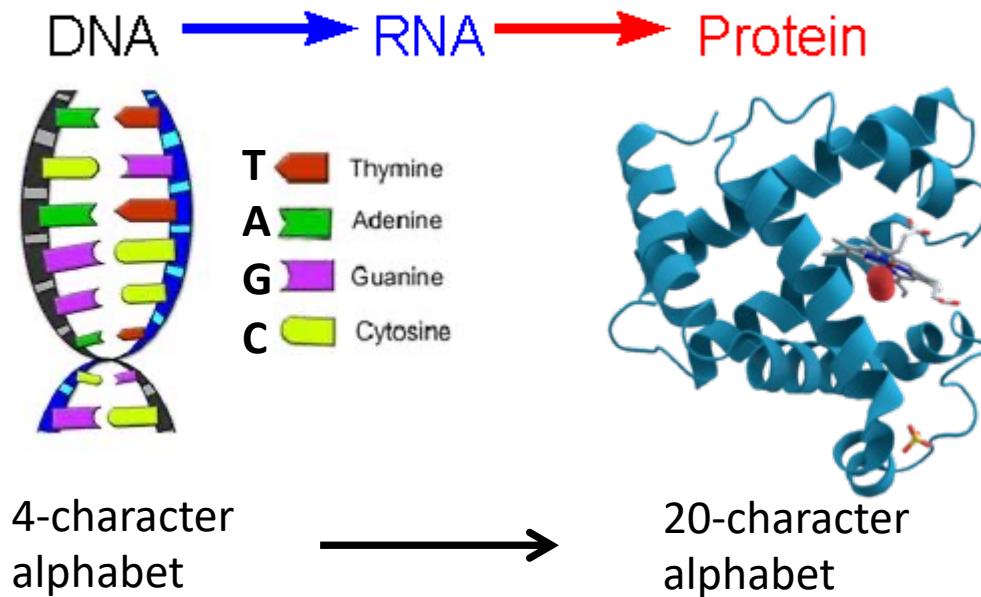
```
GGCGGCCAGGGAAGGTCTCTGCCGCCAGGGAAGTGTCCCAGAGACCCCTGGAGGGGCTGCTGACACCCCC  
GGTGCCCCCACCTCGAGCATGACCAAGGGCTGCCTCTCCCCATCCTCATCCTCCCTGCTCCACAGGACA  
TTGGCCTGGCGTCCCTGGGGCCTCGGATGAGGAAATTGAGAAGCTGTCCACGGTGGGTTGACCCCTCCC  
TGCAGGGCCTGGGGTGTGGGTTGGGGTCTGAATCCAGGCCTCACCCCTTGCCTCCAGGCTGAGGCC  
TCTCCTTCCACCCACGAATTGTGACCCTCACCTGGCCTGCCTGCATCCTGGCCTGGCCTCCCTGGGGT  
GGTATCCTGGTCACGGTGACCAGGGCTGCCCGGTGGCGCAGCTGTCTCTGGGCTGATGCTGCCCGG  
CTTCCCCGAGCTGTACTGGTCACGGTGGAGTTGGCTGTGTAAGCAGAACGGGGAGGTGAAGGCCTA  
TGGTGCCGGGCTGCTGTCCTACGGGGAGCTCTGGTGAGAGTCTCTCCTGCTGCAGCCCCCAGCAG  
AGGGGCAGGGCTGGGGACGGTGCAGGGAGGGACAGGCTCCAGTGGAGGAAACTGAGGCCTGGACCT  
CCAGGACTCAGGCTCTGTTGGGAGAAGGCTTGTCTGCCAGTCCTCACCCACATTATCCAGGCCT  
CCGAAGGCCGGCGGGGAGATGGGGTGACTCTACCCAAGGAACCCACCCAGCGTCAGGCCACGGTGCC  
CCAGTTCCCTCGGGGACCTGGGTGCAGTGGAGTCAGTGATGCCATTGGCCTCTGCCAGCAGCAGCTGTC  
TGAGGAGCCTGAGATTGGGCCTCGACCCCTGAGGCTGCCAGGCTACCAAGAACGACGTAC  
CAGTCAGTCTACTCGTGTCTGAGAGCTTCAGTGACGCCAAGGACAAGCTCAGGTGGCTAGGCTGCTAG  
GGCAAGCCCCCATGGTGCCCCAAACTGGGCCAGCCAGGCCTCTGGCCTTGAGCAGGGCTGGAC ...
```

(Technically, a cell uses the *genetic code* to
synthesize a protein from a gene)

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=tJjXpiWKMMyA>
 - 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 gene for CPS1, which is responsible for eliminating waste products from protein metabolism:
 - https://www.ncbi.nlm.nih.gov/nuccore/NG_008285.1
 - This is a preview and will make much more sense by the end of the semester