

# Introduction to Bioinformatics

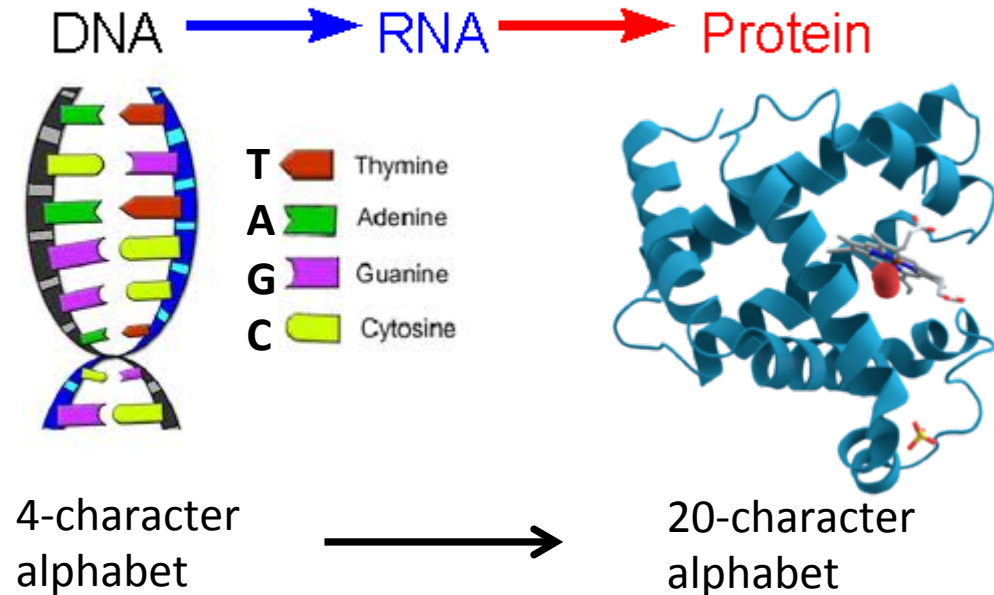
## CSC 314

Spring 2017

Dr. Garrett Dancik

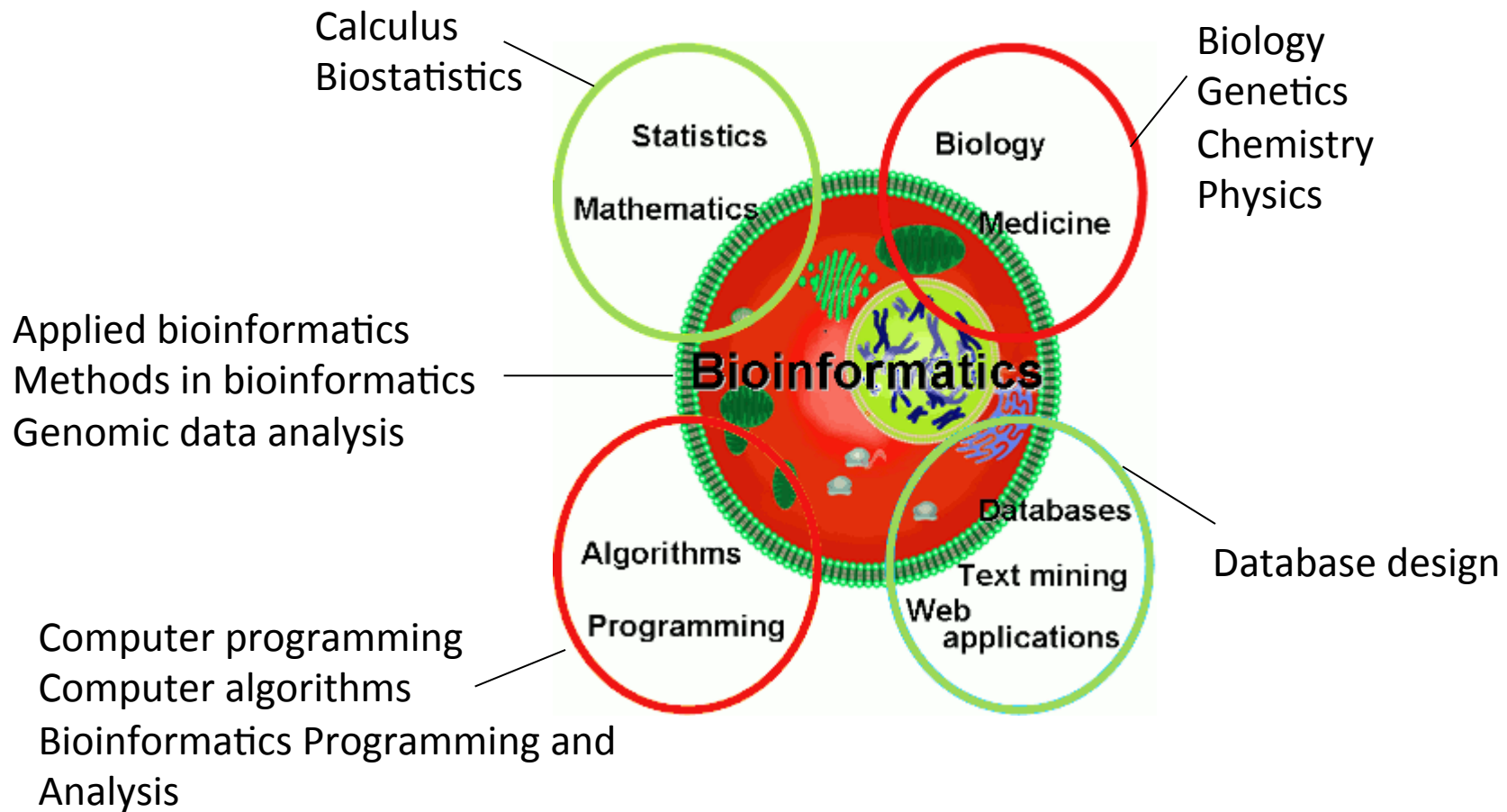
# What is bioinformatics

- Bioinformatics:
  - Biology + information
  - the study and utilization of methods for storing, retrieving and analyzing biological data
  - Falls under the category of Big 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 Big Data?

- Datasets that are "large" or at least complex
- Generated by scientific studies, technology, and commerce
- From technology
  - <http://www.domo.com/learn/infographic-data-never-sleeps>
  - In one day, FB will receive 500 terabytes of data
    - 2.7 billion likes
    - 300 million photo uploads
  - From business
    - On Black Friday (2013), Walmart processed 10 million transactions in 4 hours

# Big Data Examples

- What does FB know about you?
  - <http://www.nbcnews.com/science/gay-conservative-high-iq-your-facebook-likes-can-reveal-traits-1C8805606>
- What does Target know about you?
  - <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>
- What information does your genome reveal about you?
  - Gender
  - Paternity and ancestry
  - Eye color
  - Relative Risk of breast cancer
  - Relative Risk of alcoholism

(Note that many genetic predictions are complex, not well understood, and can be difficult to interpret; genetic traits are often not black/white, and can be influenced by environmental factors)

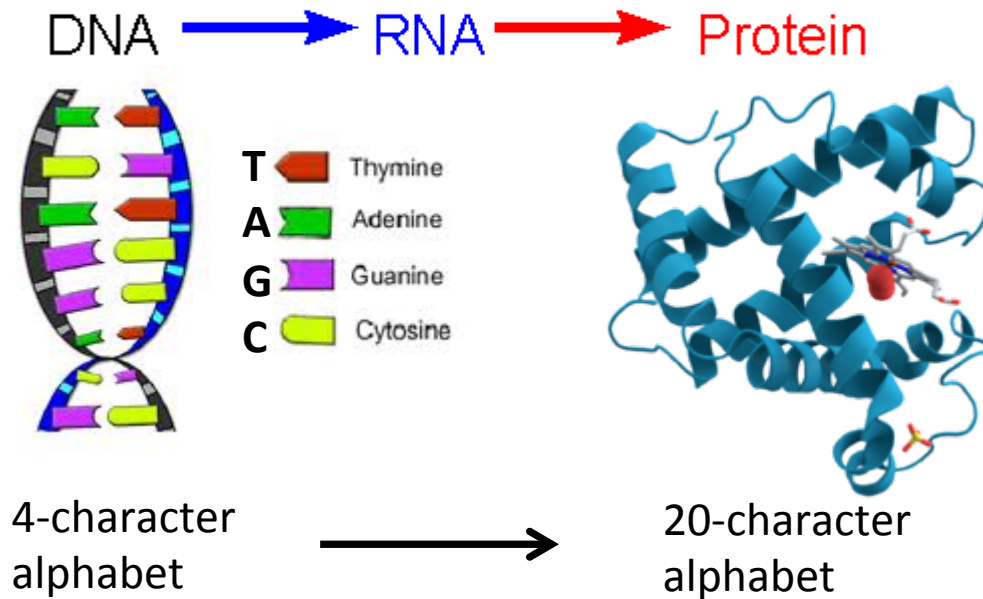
# What is this?

```
#include <iostream>
using namespace std;
int main {
    cout << "hello world\n";
    return 0;
}
```

# 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 (such as a protein)
  - Information flows from DNA → RNA → protein
  - This information determines the structure/function of RNA and protein

# Central Dogma of Molecular Biology



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



# Bioinformatics example

- Let's look briefly at the gene *BRCA1* (breast cancer type 1 susceptibility protein)
  - <http://www.ncbi.nlm.nih.gov>
  - Search Nucleotide for BRCA1
  - Click on the Genomic reference sequence
- This is a preview and will make much more sense by the end of the semester

# Intro to Genetics (Genetics 101)

- What are genes?
  - [http://www.youtube.com/watch?v=ubq4eu\\_TDFc](http://www.youtube.com/watch?v=ubq4eu_TDFc)
  - Genes are part of what molecule?
  - How many bases is DNA composed of?
  - 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?
  - What determines your sex?