

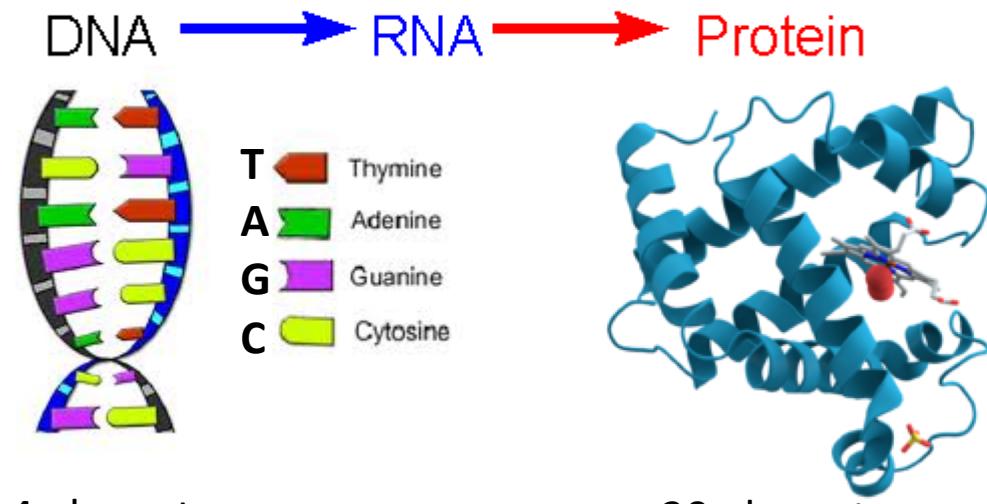
Bioinformatics Wrap-Up

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



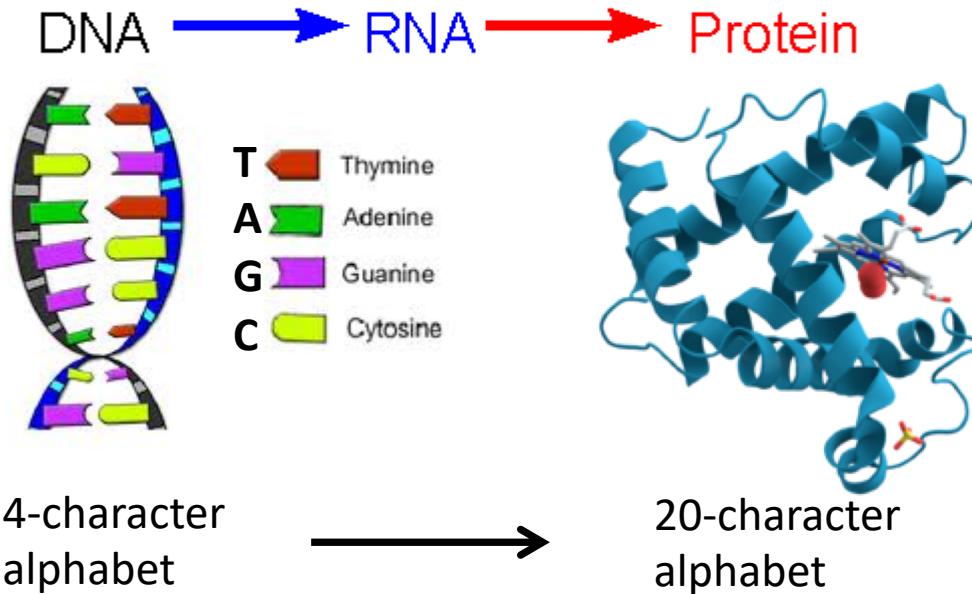
- How much information:
 - Human genome: 3 billion nucleotides
 - ~20,000 genes
 - many more when considering “junk DNA” and alternative splicing
 - >10 million sites of DNA variation
 - Countless possible interactions between DNA, RNA, and proteins

What is this?

```
public class helloWorld {  
    public static void main(String[] args) {  
        System.out.println("Hello World!");  
    }  
}
```

- 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 Java (or Python) is a language for computers, genetics is the language of life (DNA is the alphabet)
 - This is fundamental concept in bioinformatics

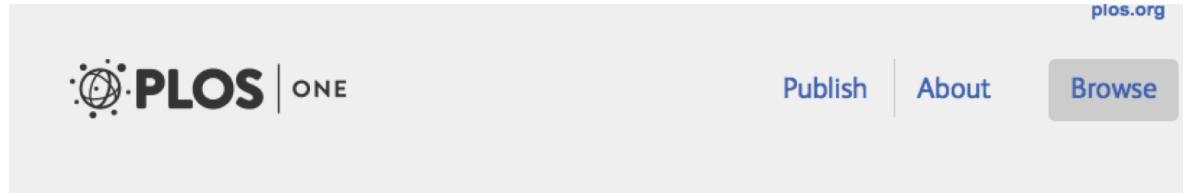
Bioinformatics example

- 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 Transcript reference sequence
- Hopefully this now makes sense!

Why do we need bioinformatics?

- To identify genetic mechanisms of diseases and other inherited (or acquired) conditions
 - Nature via nurture
- For personalized treatment of disease

Why do we need bioinformatics?



- BLAST lab

A thumbnail of a scientific article from PLOS ONE. The title is "Bioinformatics Analysis of the Complete Genome Sequence of the Mango Tree Pathogen *Pseudomonas syringae* pv. *syringae* UMAF0158 Reveals Traits Relevant to Virulence and Epiphytic Lifestyle". The authors listed are Pedro Manuel Martínez-García, Pablo Rodríguez-Palenzuela, Eva Arrebola, Víctor J. Carrión, José Antonio Gutiérrez-Barranquero, Alejandro Pérez-García, Cayo Ramos, Francisco M. Cazorla, and Antonio de Vicente. The article was published on August 27, 2015, with the DOI <http://dx.doi.org/10.1371/journal.pone.0136101>.

- Python lab

A thumbnail of a scientific article from GENOME RESEARCH. The title is "A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples". The authors listed are Samia N. Naccache, Scot Federman, Narayanan Veeraraghavan, Matei Zaharia, Deanna Lee, Erik Samayoa, Jerome Bouquet, Alexander L. Greninger, Ka-Cheung Luk, Barrett Enge, Debra A. Wadford, Sharon L. Messenger, Gillian L. Genrich, Kristen Pellegrino, Gilda Grard, Eric Leroy, Bradley S. Schneider, Joseph N. Fair, Miguel A. Martínez, Pavel Isa, John A. Crump, Joseph L. DeRisi, Taylor Sittler, John Hackett, Jr., Steve Miller, and Charles Y. Chiu. The article is from Genome Res. 2014 Jul; 24(7): 1180–1192. doi: [10.1101/gr.171934.113](https://doi.org/10.1101/gr.171934.113). The PMCID is PMC4079973.

- Group project

Regulation of developmental rate and germ cell proliferation in *Caenorhabditis elegans* by the p53 gene network

WB Derry^{*,1,2,3}, R Bierings¹, M van Iersel¹, T Satkunendran^{2,3}, V Reinke⁴ and JH Rothman^{*,1}

p53 consensus binding sites in CEP-1-regulated genes. In an effort to identify putative direct targets of CEP-1, we analyzed CEP-1-activated genes for p53 consensus DNA-binding sites.²³ Vertebrate p53 binds two copies of the palindromic DNA consensus sequence RRRCWWGYYY, and CEP-1 binds an oligonucleotide containing the human p53 consensus motif *in vivo* and *in vitro*.^{14,15} We searched the *C. elegans* genome for pairs of p53 consensus binding sites and found that ~25% of CEP-1-activated genes contain potential p53-binding sites in their promoters and introns (Table S2). These sites were

The genomic landscape of bladder cancer



Additional Databases and Tools

- The Cancer Genome Atlas
 - <http://cancergenome.nih.gov>
 - Comprehensive genomic characterization of tumors
- cBioPortal for Cancer Genomics
 - <http://www.cbioportal.org/>
- Genomics of Drug Sensitivity in Cancer
 - <http://www.cancerrxgene.org>
 - Screen > 1000 cell lines with drug
 - Genomic characterization of cell lines
- And many more (current research)
 - <http://bioinformatics.oxfordjournals.org>
 - <http://www.biomedcentral.com/bmcbioinformatics>