

Bioinformatics Wrap-Up

A stylized DNA double helix structure is depicted against a blue background. The structure is composed of numerous spheres of varying sizes and colors, including red, purple, and green, arranged in a helical pattern. The spheres are connected by thin lines, creating a sense of depth and movement. The overall aesthetic is clean and modern, typical of scientific or educational presentations.

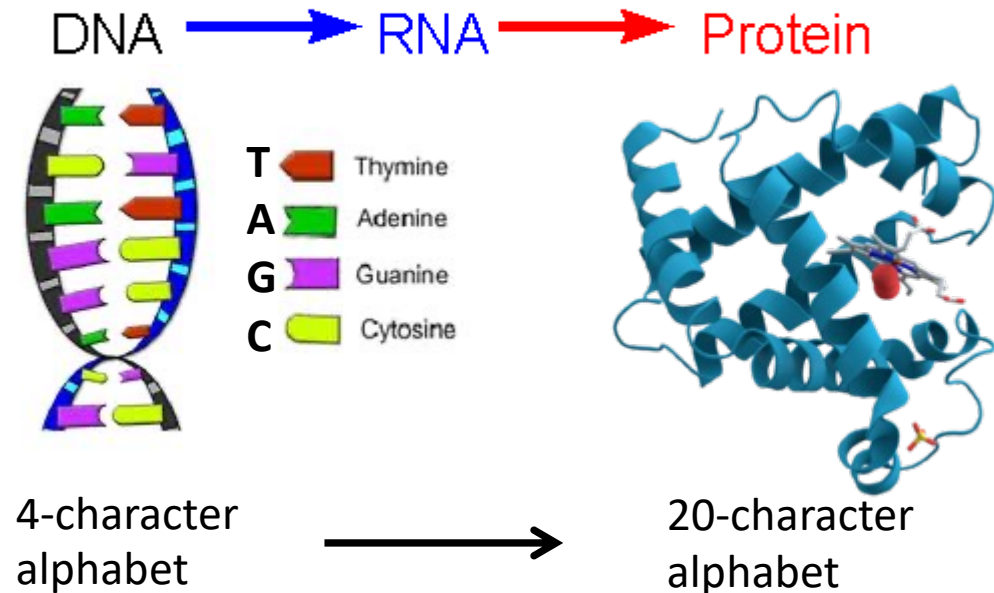
Stanislaw Pytel/Getty Images

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:

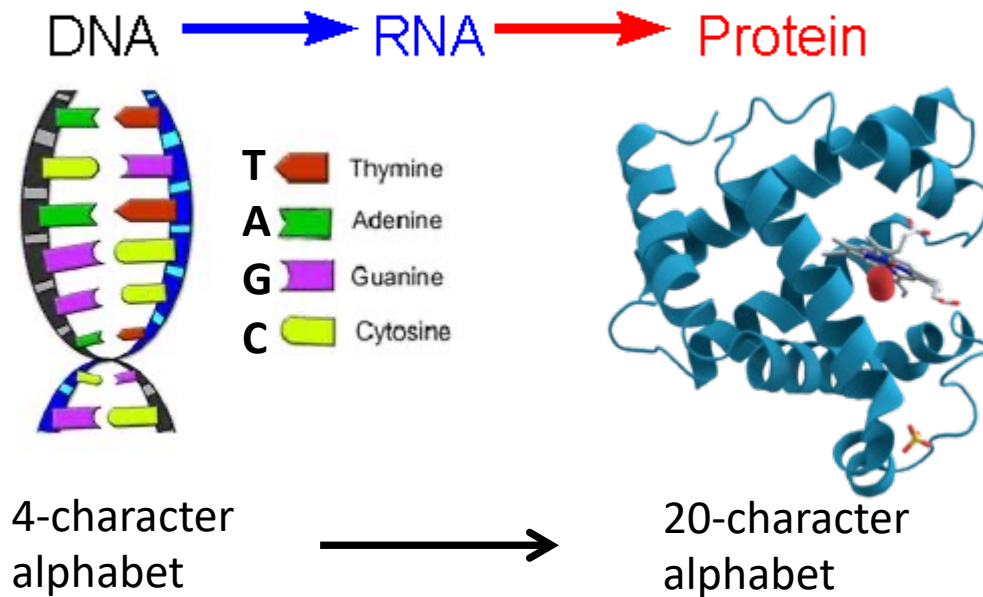
- Human genome: 3 billion nucleotides
- ~20,000 genes
 - many more when considering “junk DNA” and alternative splicing
- >10 million sites of DNA variation
- Countless possible interactions between DNA, RNA, and proteins

What is this?

```
print('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 protein from a DNA molecule
 - 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 Genomic reference sequence
- Hopefully this now makes sense!

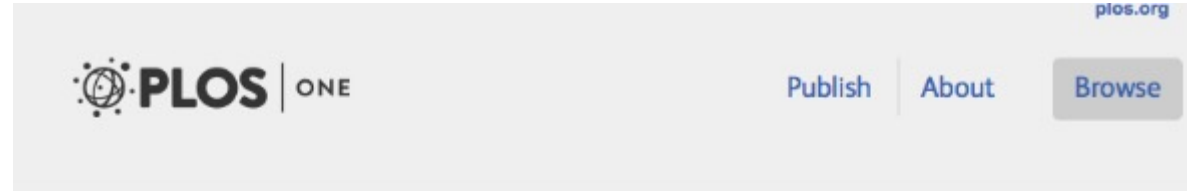
Why do we need bioinformatics?

- 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




- Sequence Alignment



OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

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

Pedro Manuel Martínez-García , Pablo Rodríguez-Palenzuela , Eva Arrebola, Víctor J. Carrión, José Antonio Gutiérrez-Barranquero, Alejandro Pérez-García, Cayo Ramos, Francisco M. Cazorla, Antonio de Vicente 

Published: August 27, 2015 • <http://dx.doi.org/10.1371/journal.pone.0136101>



CSHL Press | Journal Home | Subscriptions | eTOC Alerts | BioSupplyNet

Genome Res. 2014 Jul; 24(7): 1180–1192.

PMCID: PMC4079973

doi: [10.1101/gr.171934.113](https://doi.org/10.1101/gr.171934.113)

A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples

Samia N. Naccache,^{1,2} Scot Federman,^{1,2} Narayanan Veeraraghavan,^{1,2} Matei Zaharia,³ Deanna Lee,^{1,2} Erik Samayoa,^{1,2} Jerome Bouquet,^{1,2} Alexander L. Greninger,⁴ Ka-Cheung Luk,⁵ Barryett Enge,⁶ Debra A. Wadford,⁶ Sharon L. Messenger,⁶ Gillian L. Genrich,¹ Kristen Pellegrino,⁷ Gilda Grard,⁸ Eric Leroy,⁸ Bradley S. Schneider,⁹ Joseph N. Fair,⁹ Miguel A. Martínez,¹⁰ Pavel Isa,¹⁰ John A. Crump,^{11,12,13} Joseph L. DeRisi,⁴ Taylor Sittler,¹ John Hackett, Jr.,⁵ Steve Miller,^{1,2} and Charles Y. Chiu^{1,2,14,15}

[Author information](#) ► [Article notes](#) ► [Copyright and License information](#) ►

The genomic landscape of bladder cancer



Bioinformatics and coronavirus

- Coronavirus genomic analysis suggests early spread in the United States
 - <https://www.the-scientist.com/news-opinion/coronaviruss-genetics-hint-at-its-cryptic-spread-in-communities-67233>
- Advanced bioinformatics rapidly identifies existing therapeutics for patients with coronavirus disease-2019 (COVID-19)
 - <https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-020-02430-9>
- Combination of Biodata Mining and Computational Modelling in Identification and Characterization of ORF1ab Polyprotein of SARS-CoV-2 Isolated from Oronasopharynx of an Iranian Patient
 - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7171442/>
- Computational Strategies to Combat COVID-19
 - <https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbaa232/5955939>

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>