

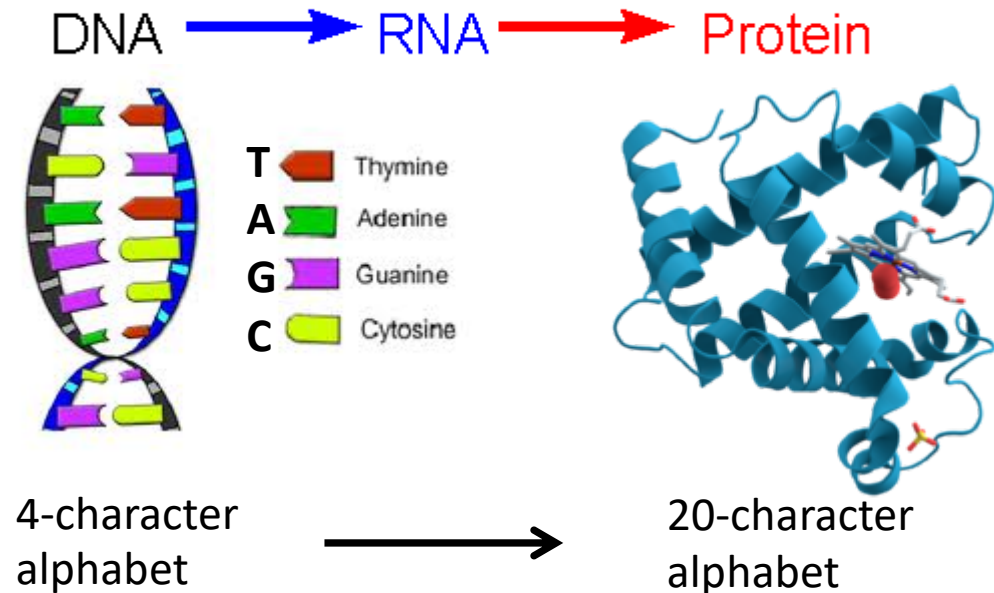
# Bioinformatics Wrap-Up

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:

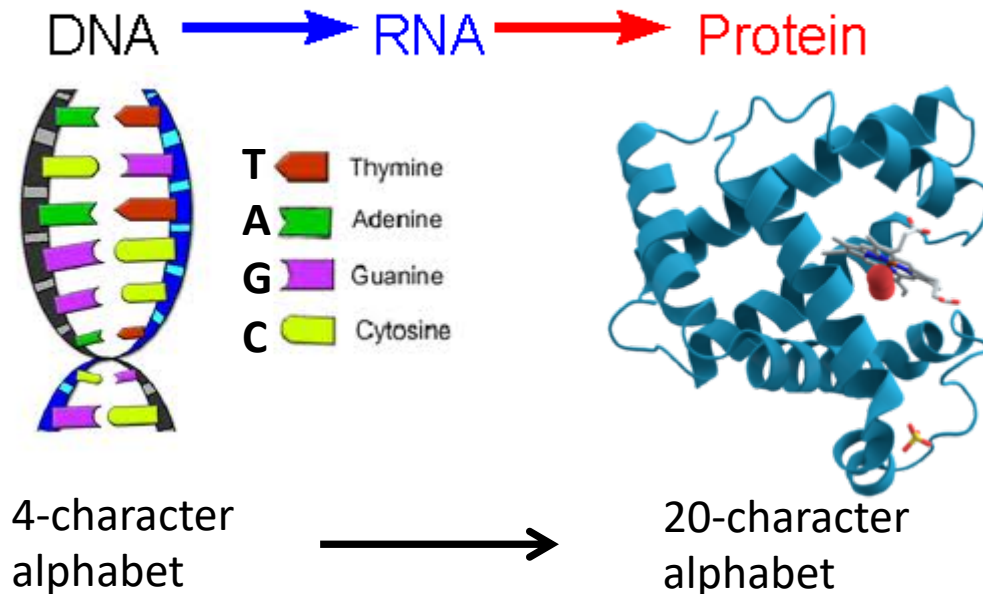
- Human genome: 3 billion nucleotides
- ~20,000 genes
  - many more when considering “junk DNA” and alternative splicing
- >10 million sites of DNA variation
- Countless possible interactions between DNA, RNA, and proteins

# What is this?

```
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 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

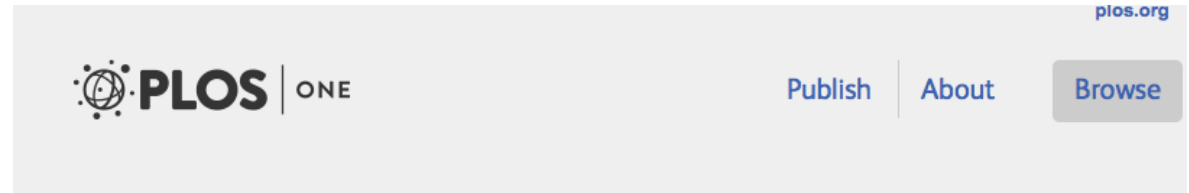
- Let's look briefly at the gene *BRCA1* (breast cancer type 1 susceptibility protein)
  - <http://www.ncbi.nlm.nih.gov>
  - Search Nucleotide for BRCA1
  - Click on the Transcript reference sequence
- Hopefully this now makes sense!

# 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



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.

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

PMCID: PMC4079973

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

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

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

- Python lab

## Lab #8


## REVIEW

## Census and evaluation of p53 target genes

M Fischer<sup>1,2,3</sup>

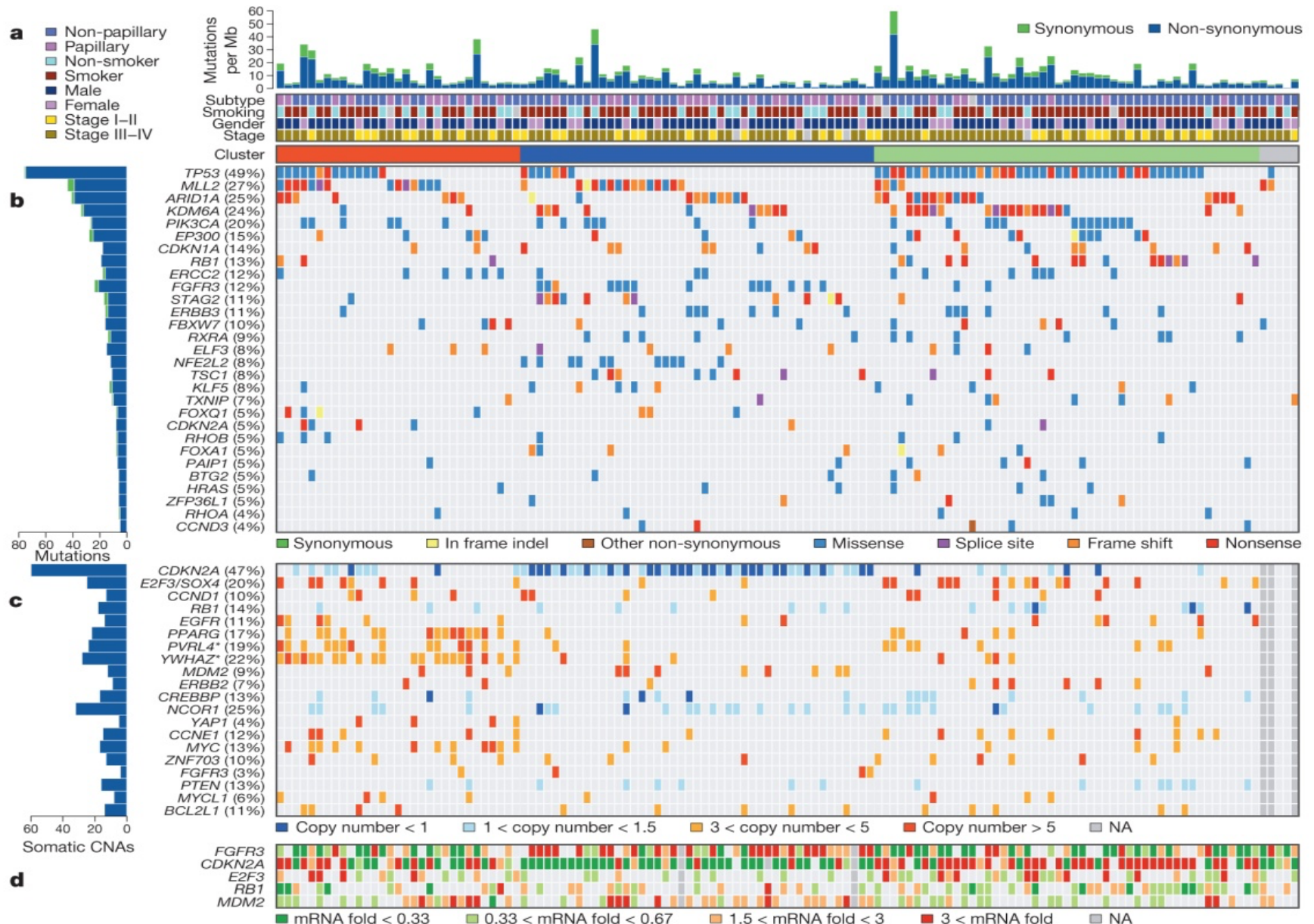
The tumor suppressor p53 functions primarily as a transcription factor. Mutation of the *TP53* gene alters its response pathway, and is central to the development of many cancers. The discovery of a large number of p53 target genes, which confer p53's tumor suppressor function, has led to increasingly complex models of p53 function. Recent meta-analysis approaches, however, are simplifying our understanding of how p53 functions as a transcription factor. In the survey presented here, a total set of 3661 direct p53 target genes is identified that comprise 3509 potential targets from 13 high-throughput studies, and 346 target genes from individual gene analyses. Comparison of the p53 target genes reported in individual studies with those identified in 13 high-throughput studies reveals limited consistency. Here, p53 target genes have been evaluated based on the meta-analysis data, and the results show that high-confidence p53 target genes are involved in multiple cellular responses, including cell cycle arrest, DNA repair, apoptosis, metabolism, autophagy, mRNA translation and feedback mechanisms. However, many p53 target genes are identified only in a small number of studies and have a higher likelihood of being false positives. While numerous mechanisms have been proposed for mediating gene regulation in response to p53, recent advances in our understanding of p53 function show that p53 itself is solely an activator of transcription, and gene downregulation by p53 is indirect and requires p21. Taking into account the function of p53 as an activator of transcription, recent results point to an unsophisticated means of regulation.

*Oncogene* (2017) 36, 3943–3956; doi:10.1038/onc.2016.502; published online 13 March 2017

Typically, p53 binds to the target genes as a tetramer, which comprises two dimers that each binds a decameric half-site with the consensus sequence RRCWWGYYY  (R = A/G, W = A/T, Y = C/T).<sup>7–10</sup> The discovery of the first p53



# 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
  - Comprehensive genomic characterization of tumors
  - <http://cancergenome.nih.gov>
- 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>