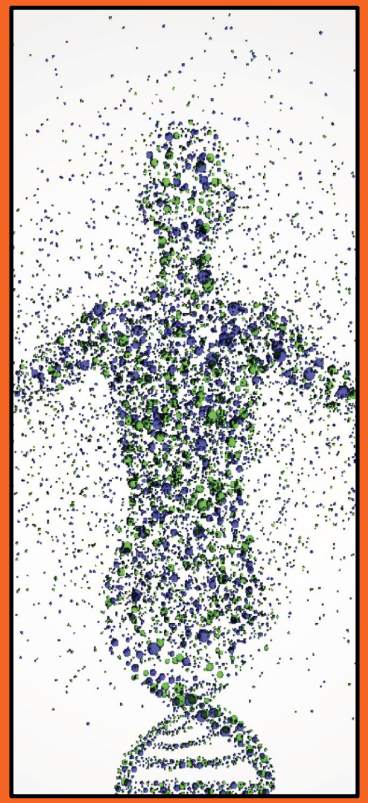
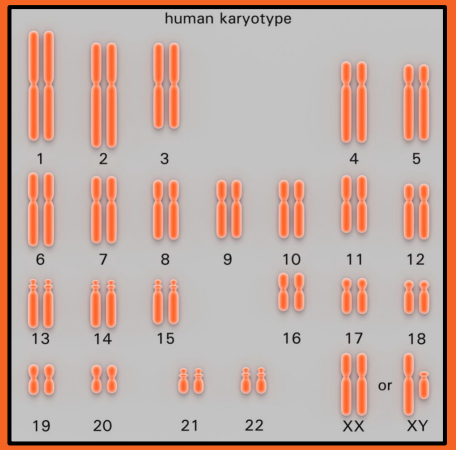


Visualizing the Human Genome!



Whitney Fee | Eric Ellestad | Angel Ortiz Nuñez

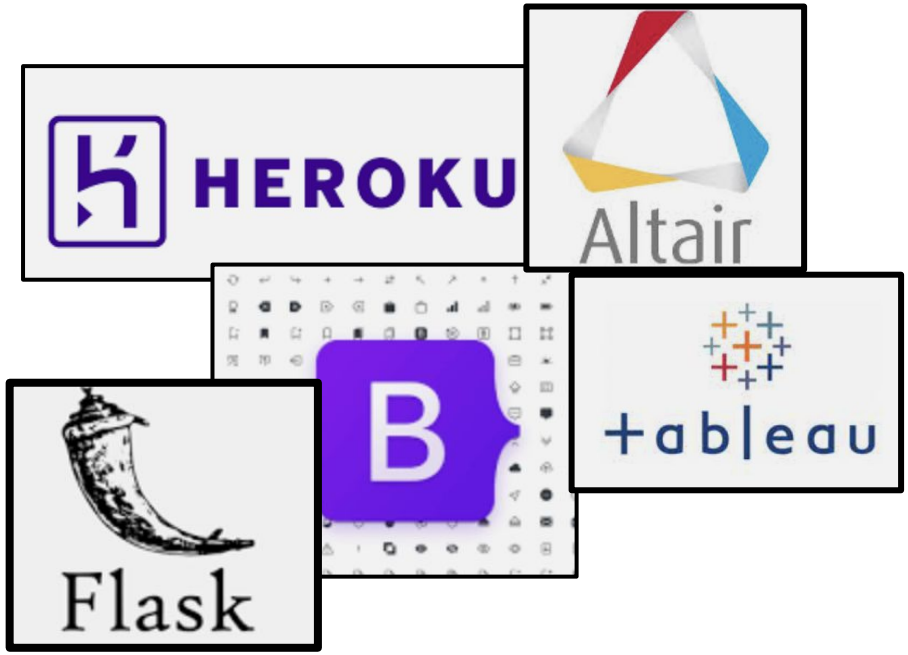
DNA → RNA → Proteins

The Central Dogma of Biology

Users

Biology Students

Architecture



- Flask - backend
- Tableau and Altair charts
  - iFrame embeds
  - Python in Flask
- Heroku
- Bootstrap

Usability Test Feedback

- Load/lag time
- Confusing definition links
- Website aesthetics/chart sizing
- Consolidation/clean-up of views
- Interactivity Instruction clarity

Lead/lag Improvement

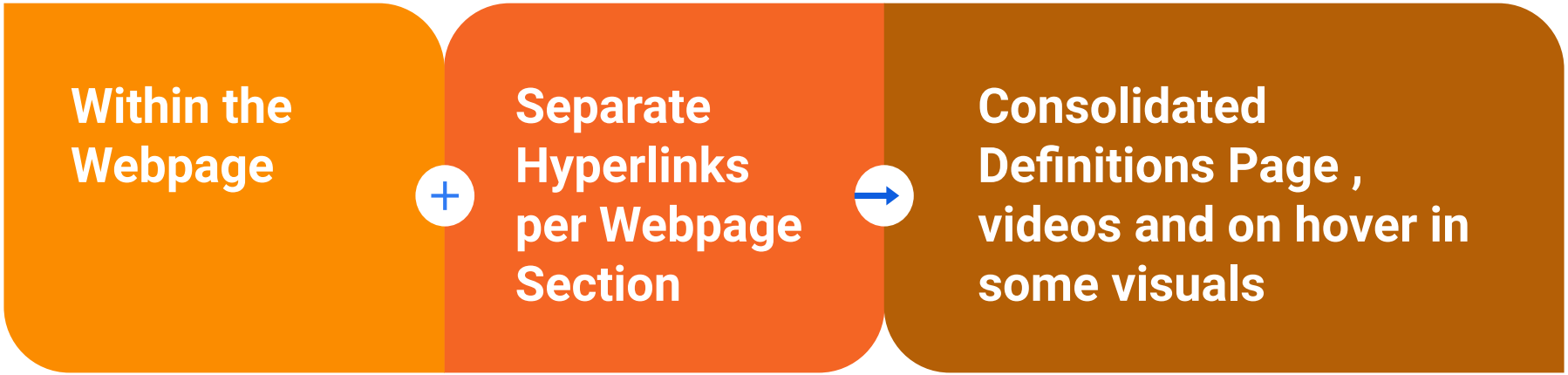
- Stored each gene in its own compressed CSV and only import the gene-subcomponent detail if/when the gene is selected for viewing

- Removed the most burdensome charts and reconstructed with lighter weight datasets

- Reduced the total number of interactive charts

- Minimized the number of charts that required dropdown selection interactivity and shared a single dropdown across all gene expression charts

Definitions Improvement



Web Aesthetics and Sizing Improvement

Visualizing the Human Genome!

Berkeley School of Information - W209 - Spring 2022 - Data Visualization Final Project

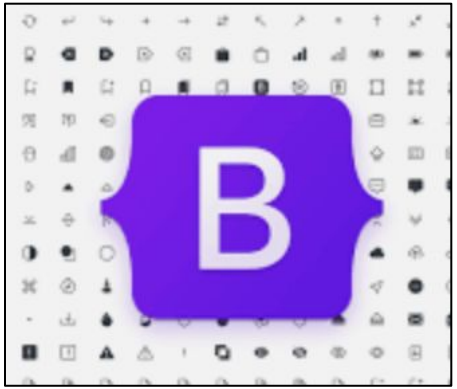
Team Members:

- Whitney Fee
- Angel Ortiz Nuñez
- Eric Ellestad

PART 1: An Introduction to the Genome and its Building Bl

Genome

The genome is the entire set of genetic instructions found in a cell. In humans, the genome cons
chromosome found in the cells' mitochondria. Each set of 23 chromosomes contains approxima



Visualizing the Human Genome

UC Berkeley School of Information

Spring 2022 - W209 Data Visualization - Final Project

Introduction

The Human Genome Project was completed in 2003 which ushered in the modern genomics era by sequencing an entire human genome for the very first time. In the following two decades, advances in high-throughput genetic sequencing technologies have made DNA sequencing faster, cheaper, and widely available. This has lead to a proliferation of genomic "big data" and the

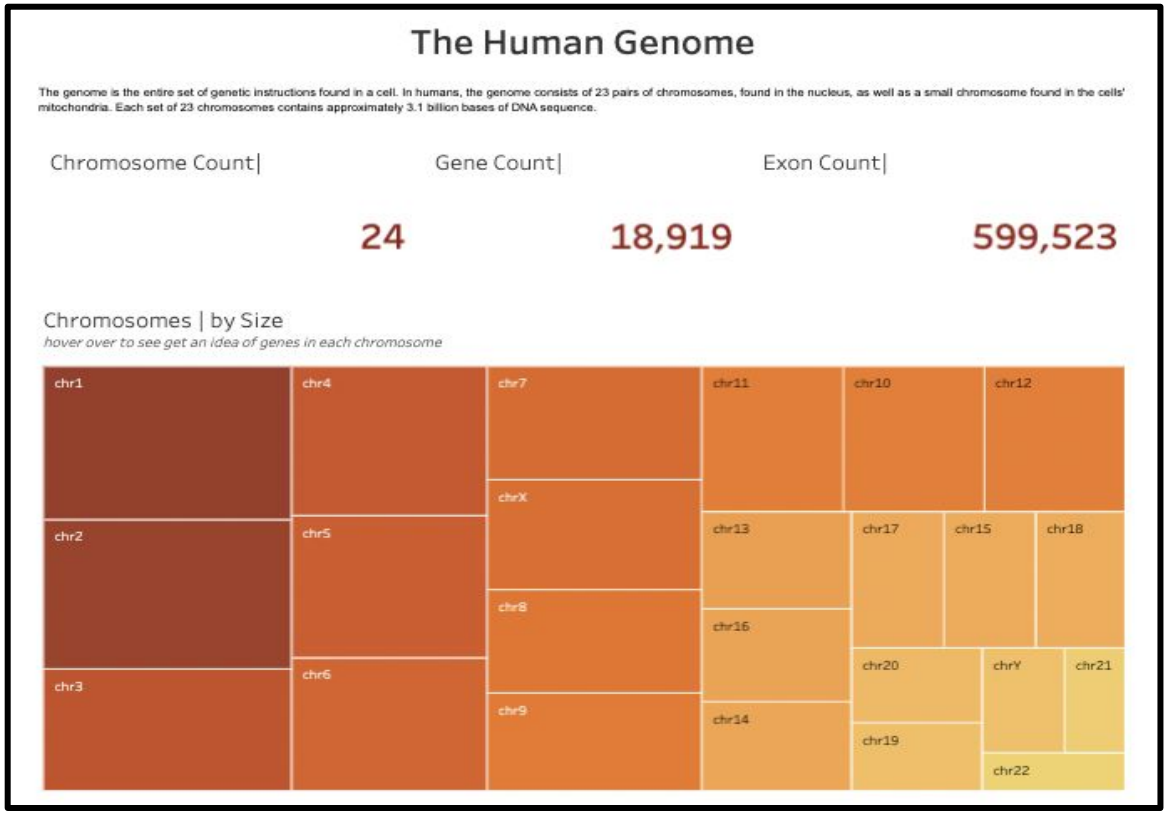
View Cleanup and Instructions Clarity Improvement

- **Will discuss visual by visual**

- **Overall Learnings:**
  - **Sometime less is more in terms of charts and increasing understanding**
    - Example: Too many charts leading to confusion on a subject

  - **Just because you can doesn't always mean you should from a viz perspective if it confuses users**
    - Example: Chromosome icons for interactivity/zooming functionality

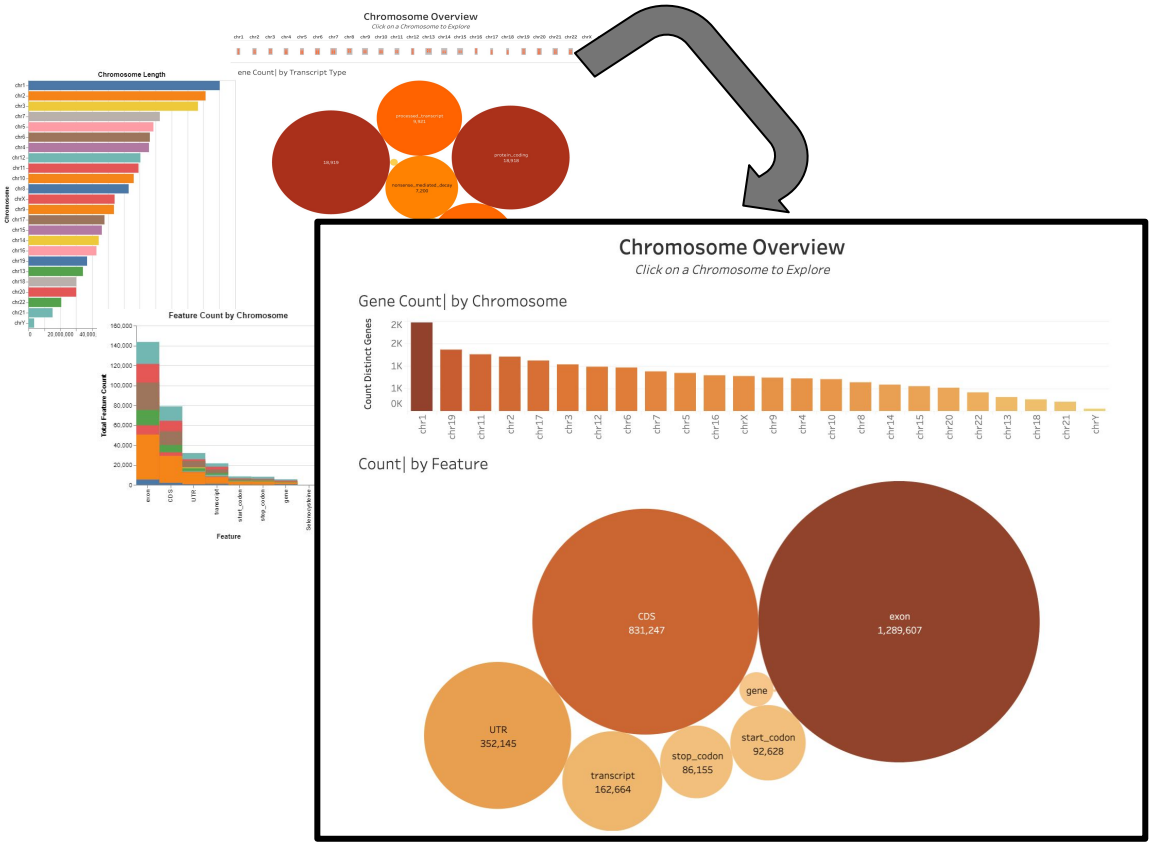
Part 1 - Overview



Tasks:

1. Count the Chromosomes
2. Count the Genes
3. Count the Exons
4. Inspect the relative visual size of the Chromosomes/Genes/Exons
5. Confirm the Chromosomes appear in pairs:
    a. Chr 1-22 are autosomal pairs (same chromosome from each parent)
    b. Sex Chromosome is either XY or XX
6. Understand scope of Human Genome at a high level and have a sense of Hierarchy - Chromosome being at the top

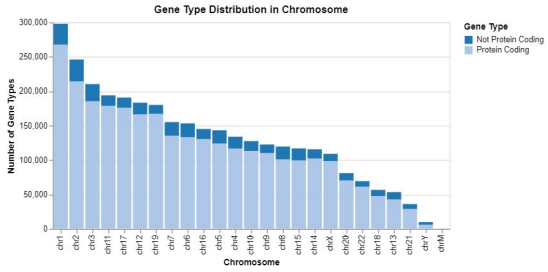
Part 2 - Looking Within a Chromosome and Locating and Identifying Genes



Tasks:

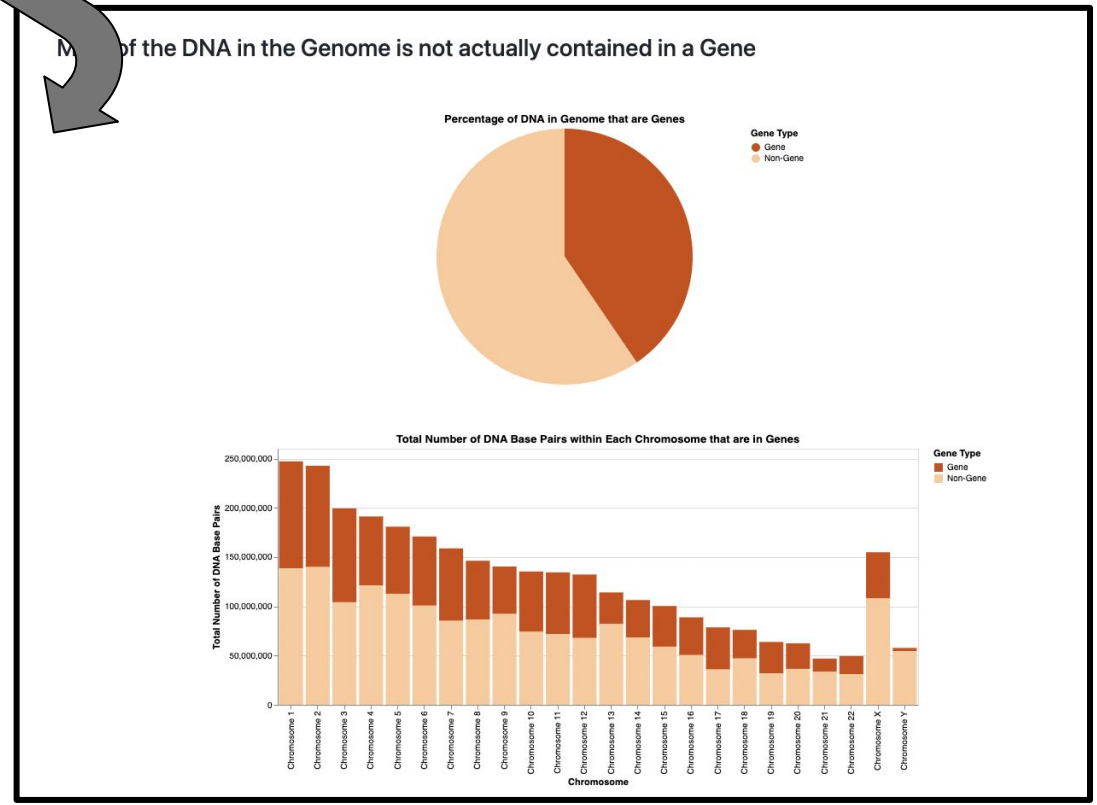
1. For each Chromosome - understand the interaction and click on a Chromosome
2. Identify distribution of Chromosome length
3. Once Chromosome is clicked glean understanding of the Counts associated with the Chromosome by their Feature
4. Begin understanding the hierarchy of the Human Genome - Chromosome -> Gene

Part 2 - Looking Within a Chromosome and Locating and Identifying Genes

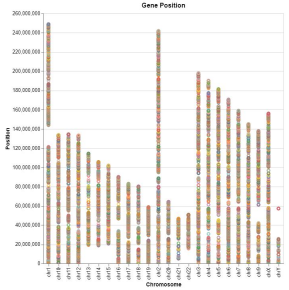


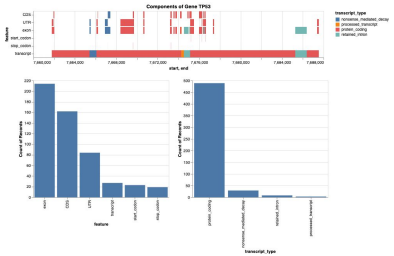
Tasks:

1. Identify how much of the human genome is considered a gene (both in amount and percentage).
2. Compare the distribution of genes/non-genes between chromosomes.



Part 3, 4, 5 ,6 - Gene Subcomponents





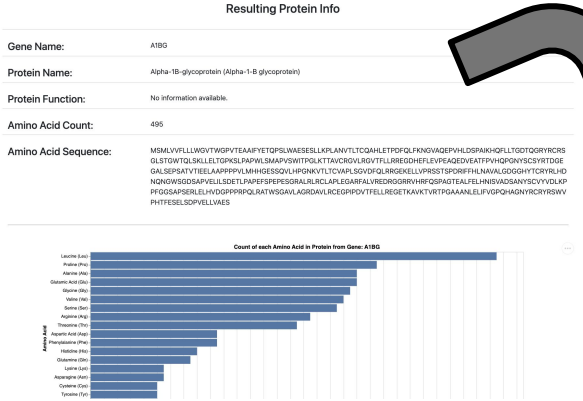
Gene Location Tasks:

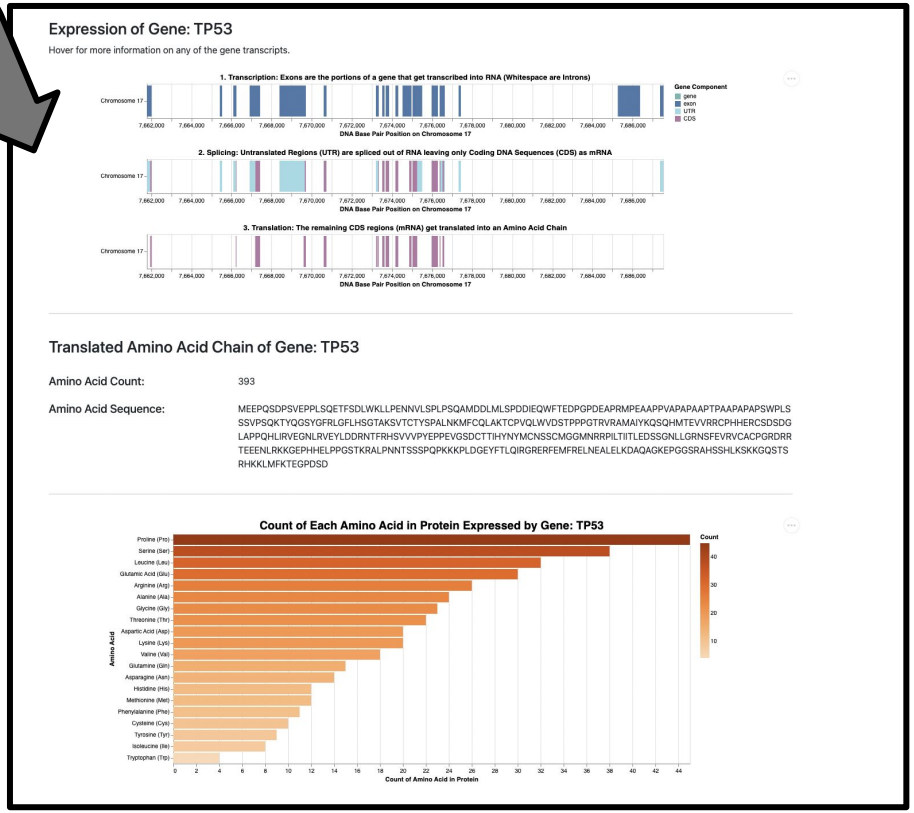
1. Identify gene position and chromosome.
2. Identify surrounding genes to a given gene.

Sub Components Tasks:

1. Identify features present in gene and their locations
2. Compare gene feature counts

Part 7 - Visualize the Protein a Gene Encodes For





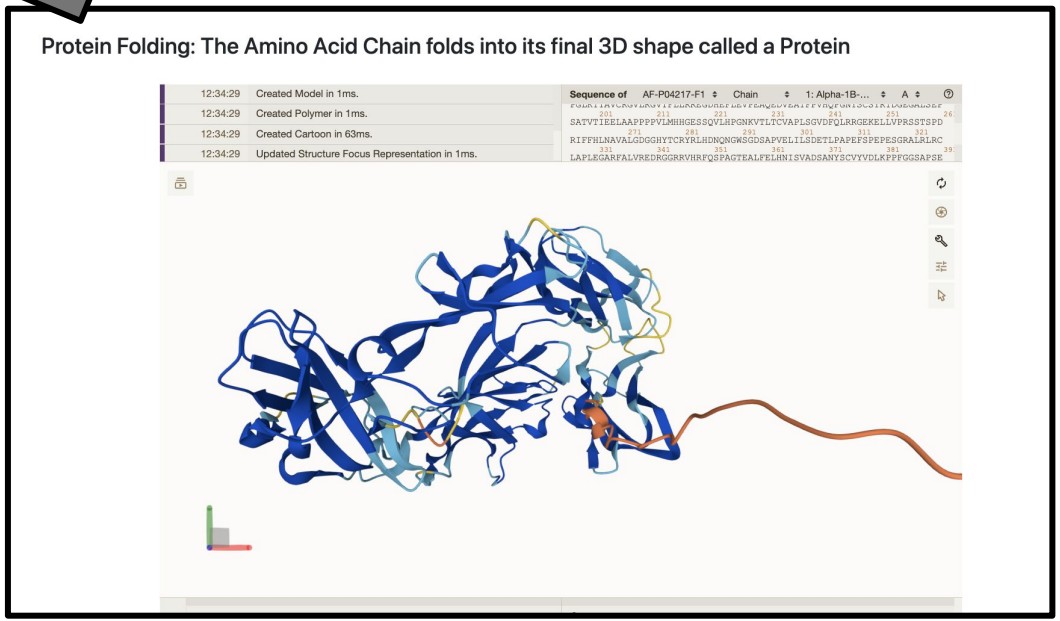
Tasks:

1. Identify the name of the Protein encoded for by a given gene.
2. Identify the Amino Acid Sequence that defines the Protein.
3. Identify the 3D shape of the protein.
4. Identify the protein's function, if known.

Part 7 - Visualize the Protein a Gene Encodes For (Improved)

Tasks:

1. See shape and structure of the Protein created by the Amino Acids of the Gene filtered for



Recap

Website Demo

Emails

Whitney Fee: whitneylynnefee@ischool.berkeley.edu

Eric Ellestad: eric.ellestad@ischool.berkeley.edu

Angel Ortiz: angelortiz@ischool.berkeley.edu

Division of Responsibilities

Whitney Fee:

- Exploratory Data Analysis
- Tableau Visualization - Genome Chromosome
- Tableau Visualization - Chromosome Overview
- CSS and Website Formatting
- Definitions

Angel Ortiz Nuñez

- Data Pre-Processing and Data Wrangling
- Altair Charts - Gene Chromosome
- Altair Charts - Protein Composition
- Altair - Gene Function Charts
- Altair - Interactivity Across Charts

Eric Ellestad:

- Data Collection and Dataset Integration
- 3D Protein Viewer
- Altair Charts - Gene Expression Charts
- Bootstrap and Website Formatting
- Flask Backend and Website Integration
- Heroku Platform Integration and Hosting