



# PeakLearner

---

Team GNomes Members: Jacob Christiansen, Allen Clarke, Yuanyuan Fu, John Jackson

Mentor: Mahsa Keshavarz

Sponsor: Dr Toby Hocking

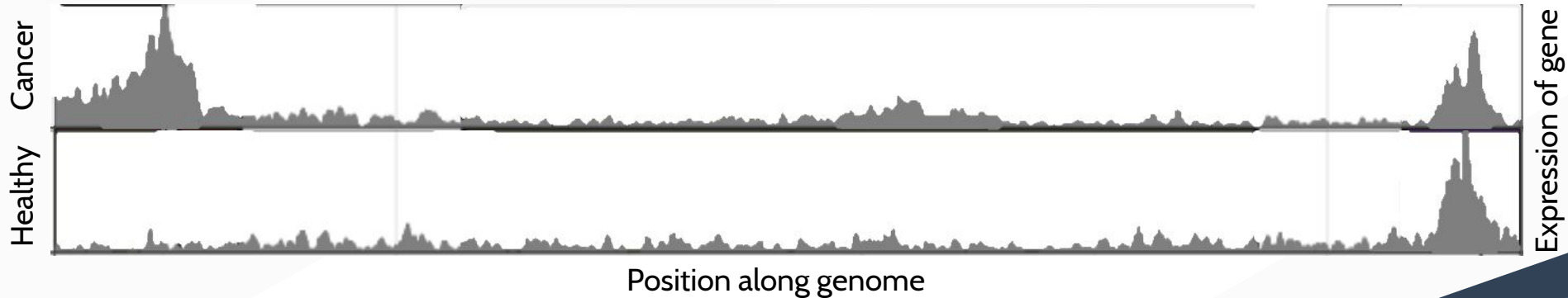
# Biology Background

- Human genome contains about 3 billion base pairs.
  - A, C, T, G
- Genes mutate over time.
- Cancer is caused by gene mutation.
- How can we detect potential mutations before they grow large enough to cause concern?

# Biology Example

- Peaks are found where a small section of genome is highly expressed
- Peaks can help determine which genes could be causing cancer

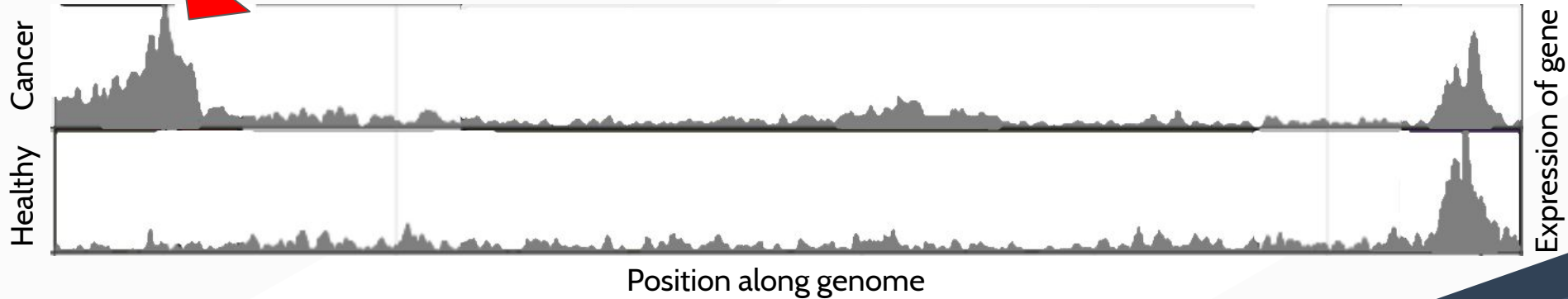
Graph of ChIP-Seq data from two different patients



# Biology Example

- Cause of cancer?

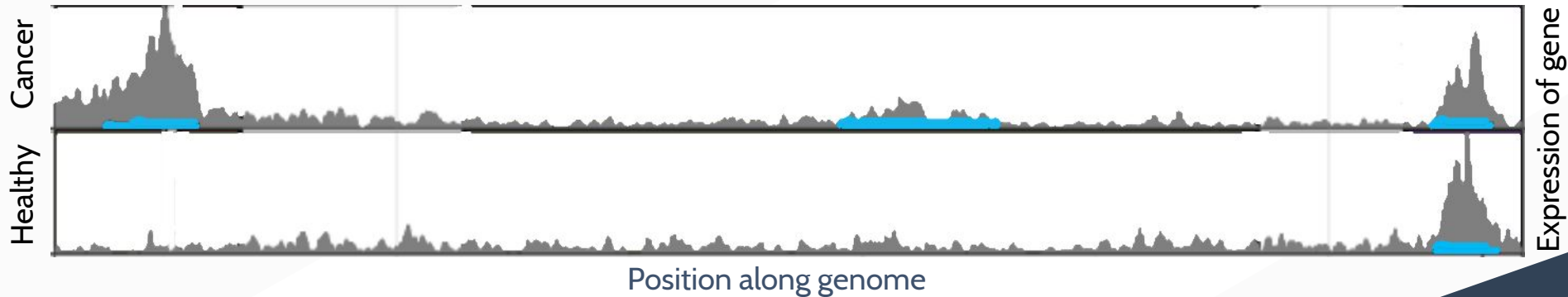
Graph of ChIP-Seq data from two different patients



# Biology Example

- Blue line represents a model stating where Peaks are found
  - Currently models are not very accurate

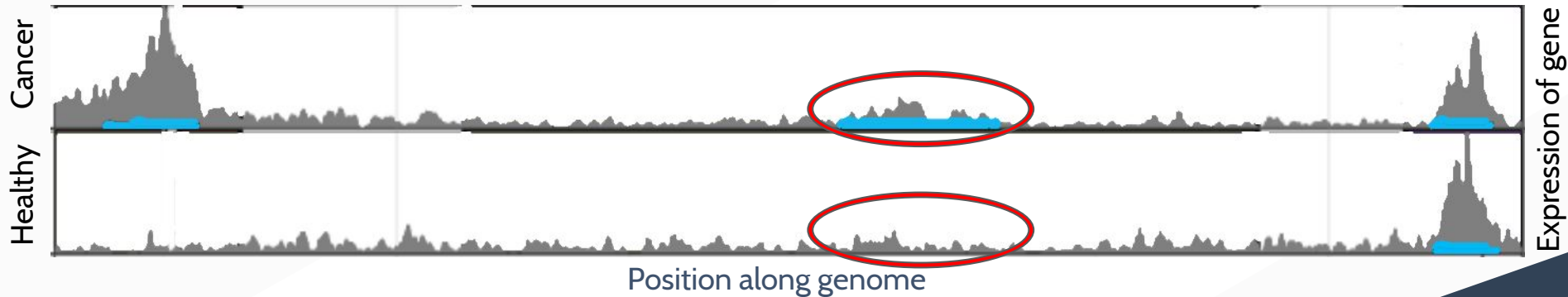
Graph of ChIP-Seq data from two different patients



# Biology Example

- Blue line represents a model stating where Peaks are found
  - Currently models are not very accurate

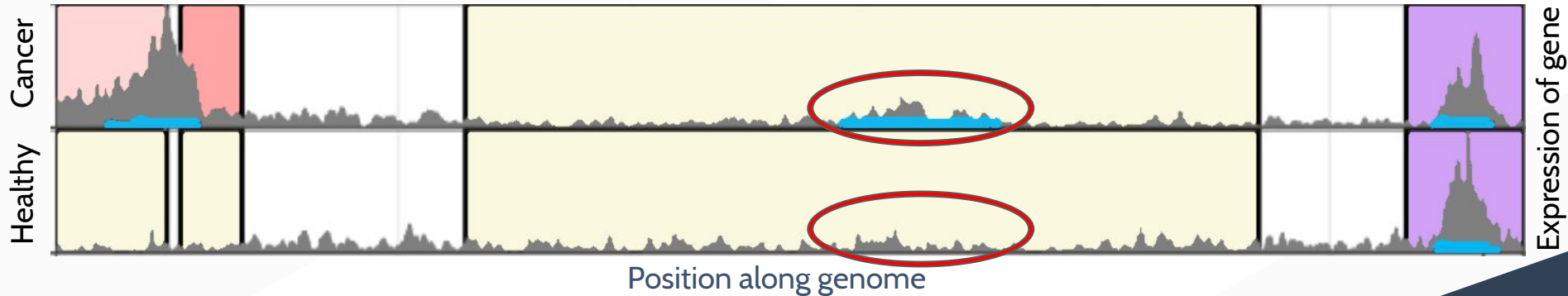
Graph of ChIP-Seq data from two different patients



# Biology Example

- Adding Labels to the picture gives a clearer idea of what is expected of the model in a given region of the genome

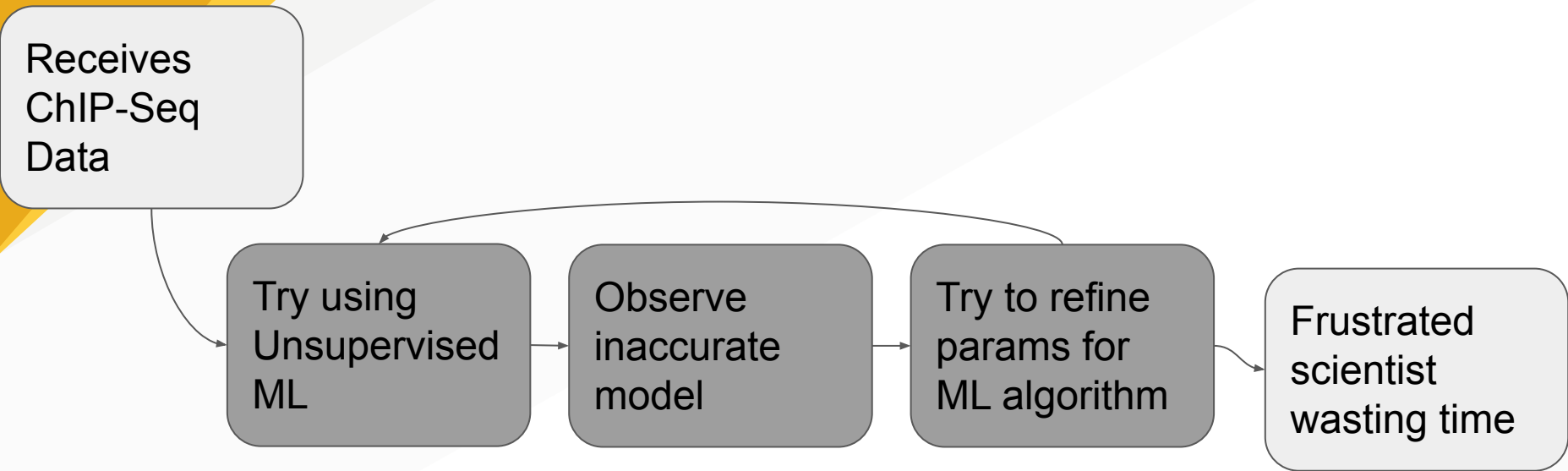
Graph of ChIP-Seq data from two different patients



Yellow - No Peak  
Purple - Peaks

Pink - Start of Peak  
Red - End of Peak

# Current Scientist Workflow





# Problem

- Scientists are frustrated because they can clearly see peaks in data but can not label them accurately
- Current tools
  - Guess and check unsupervised ML models
  - No clear graphical representation
  - No way to fix a model after it is made

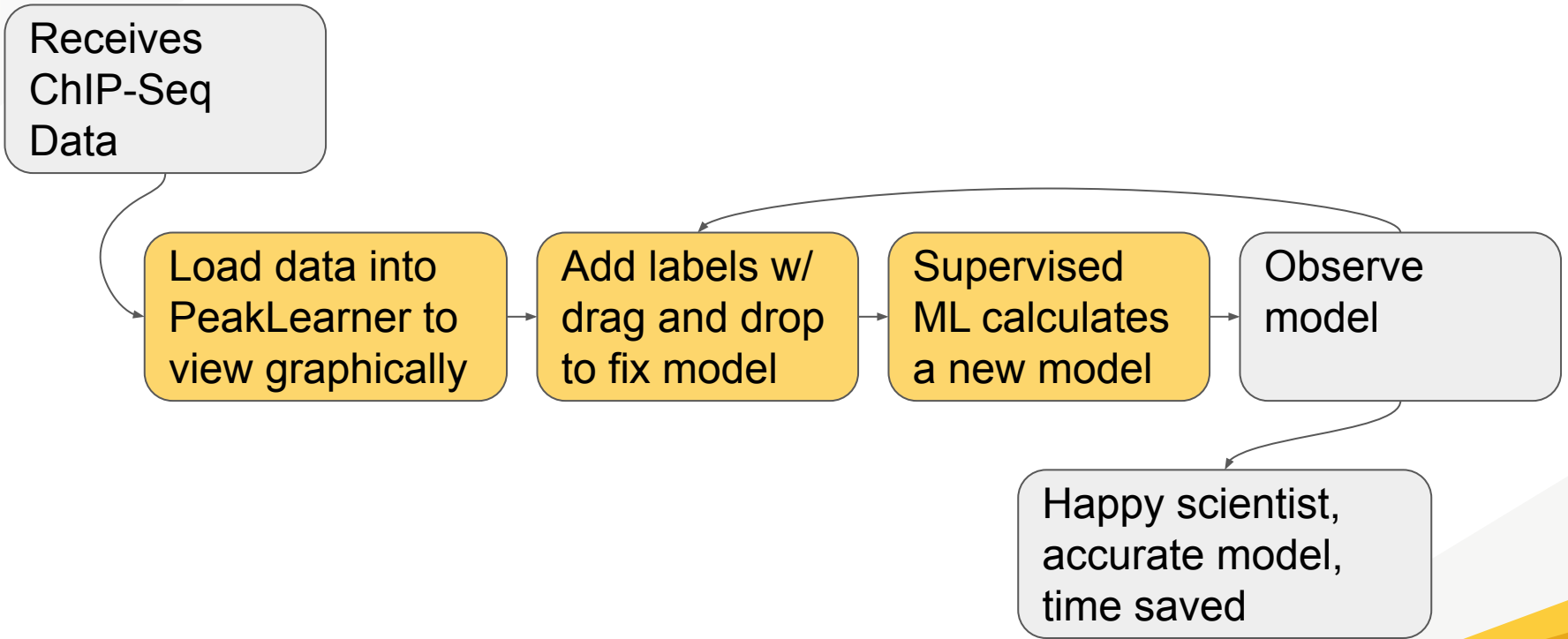
# Solution

- Build a GUI webapp for scientists
- Genome Browser
  - Tool already used in industry
- Allow them to click and drag over model to fix inaccuracies by adding labels
  - Eliminates guess and check of unsupervised ML
  - Replaces with supervised ML
- User sees model change immediately after adding label
  - Saves time getting accurate model

# Proposed Solution



# New + Improved Workflow

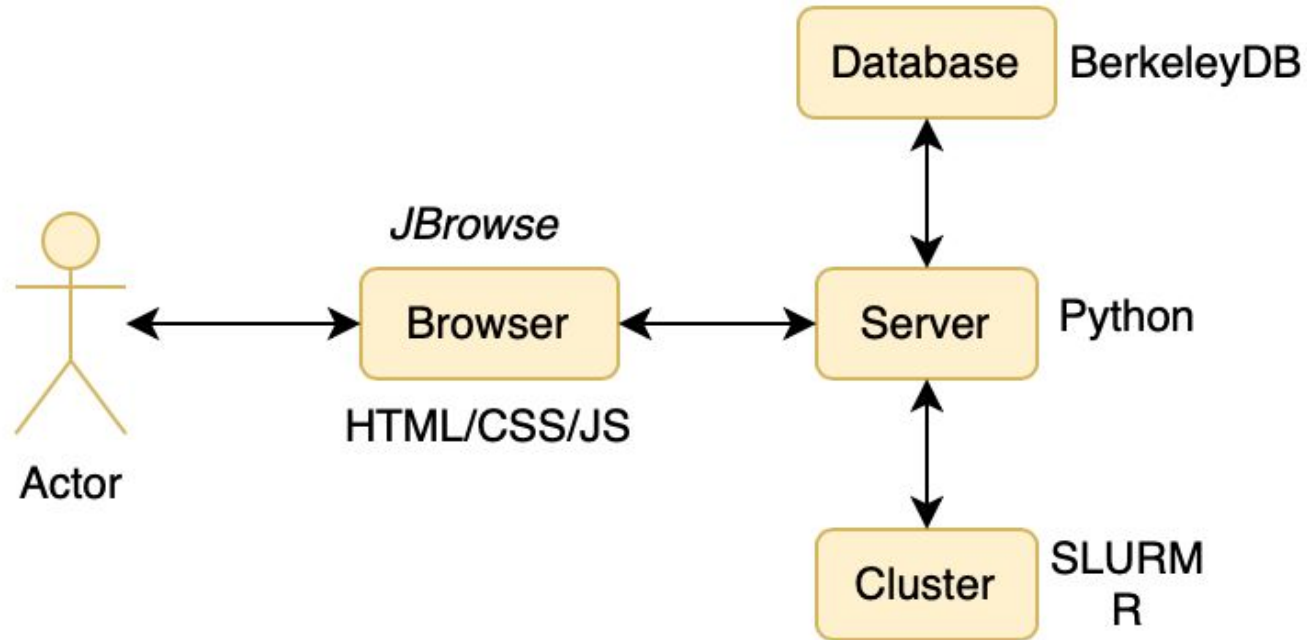


# Implementation Overview

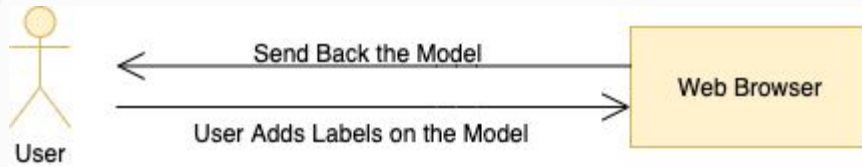
## *Key Functional Requirements*

- Interactive web interface
- Users can add/ modify labels
- Generated peak model superimposed on data
- Users can download data & model

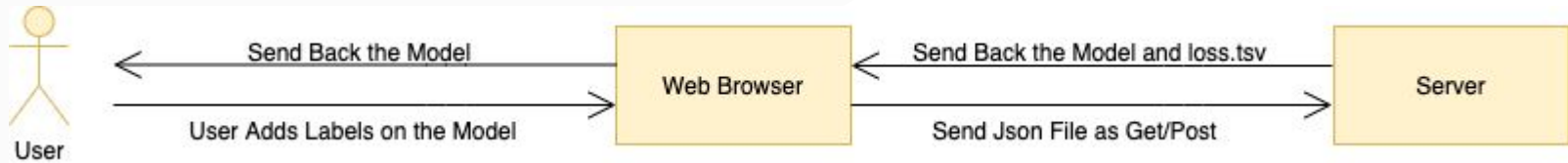
# Implementation Overview



# Architecture Overview

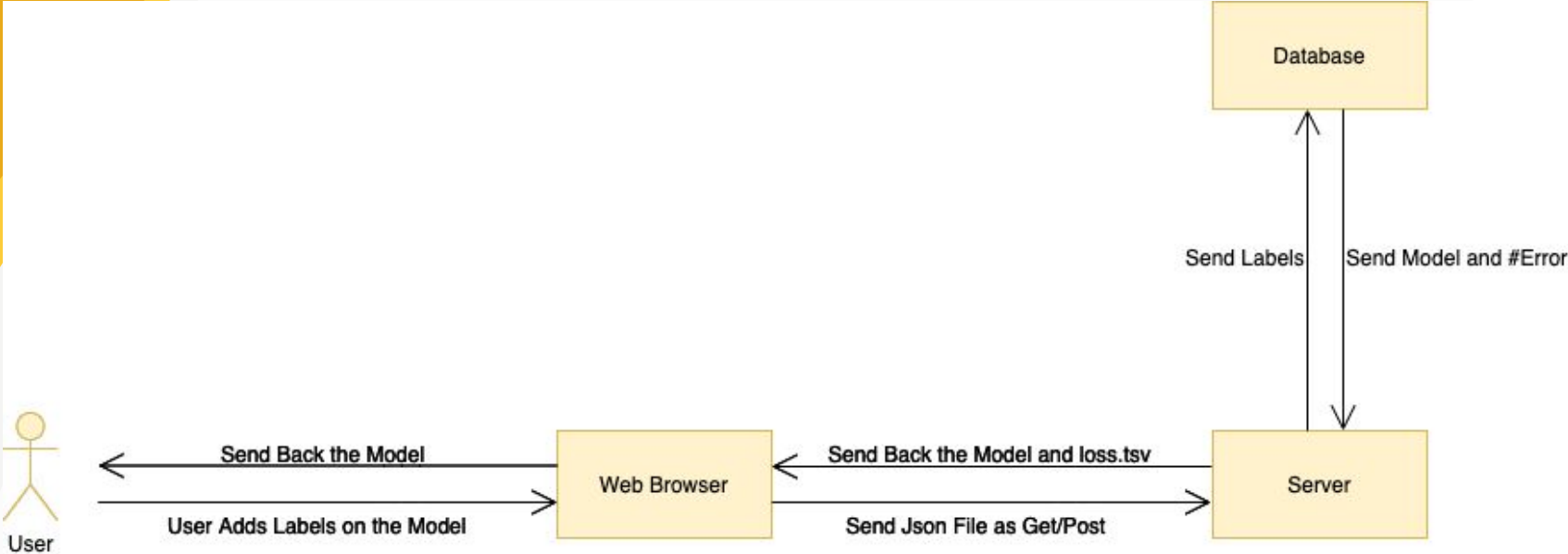


# Architecture Overview

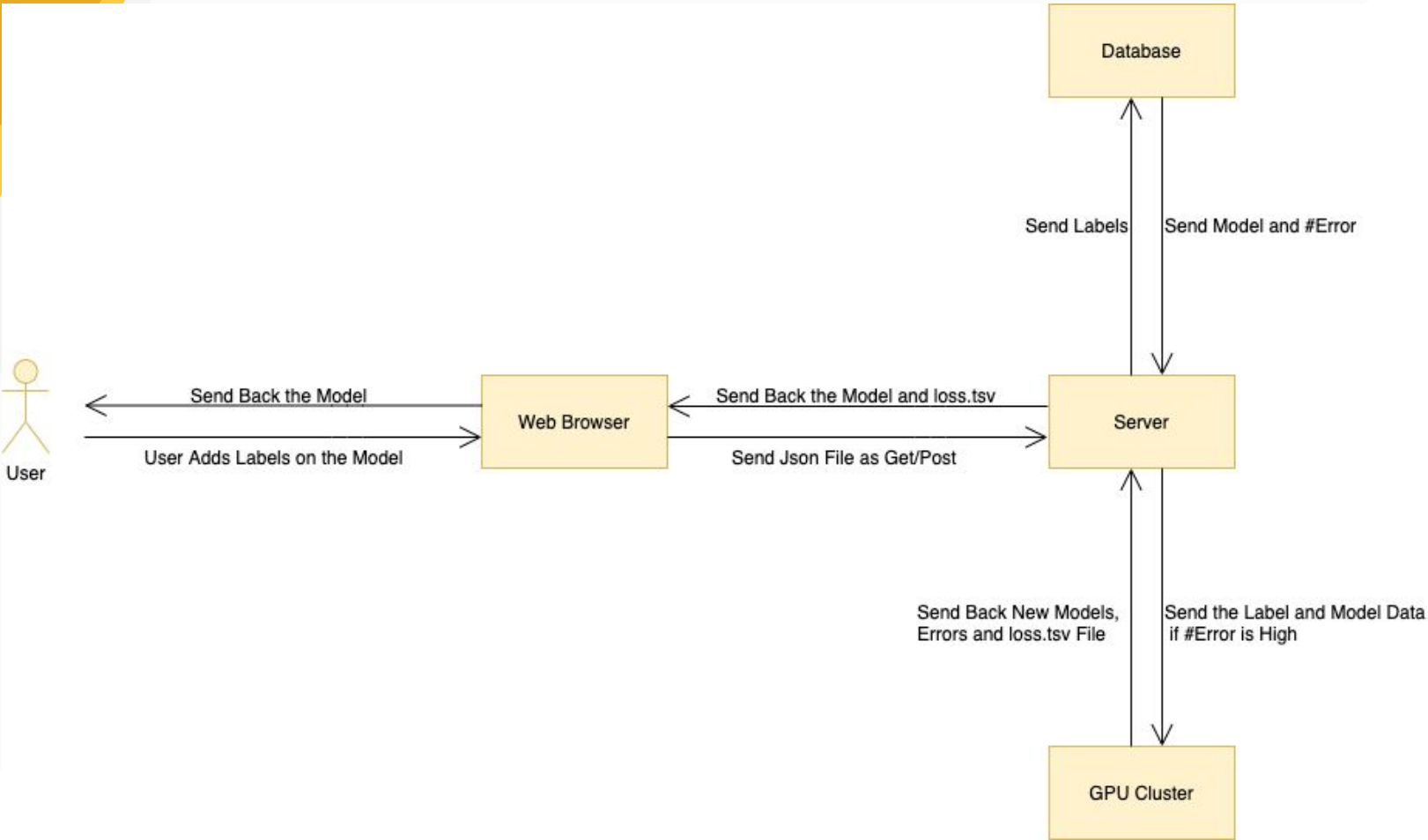




# Architecture Overview



# Architecture Overview



# Challenges

- **BigWig => range requests**
- **BedGraph => color callbacks**
- **Parsing Dr. Hocking's code**
- **Debugging**

# Challenges

- BigWig => range requests
- **BedGraph => color callbacks**
- Parsing Dr. Hocking's code
- Debugging

# Challenges

- BigWig => range requests
- BedGraph => color callbacks
- **Parsing Dr. Hocking's code**
- Debugging

# Challenges

- BigWig => range requests
- BedGraph => color callbacks
- Parsing Dr. Hocking's code
- **Debugging**

# Gantt Chart

Now

## PeakLearner

### Period 1 - Review the project

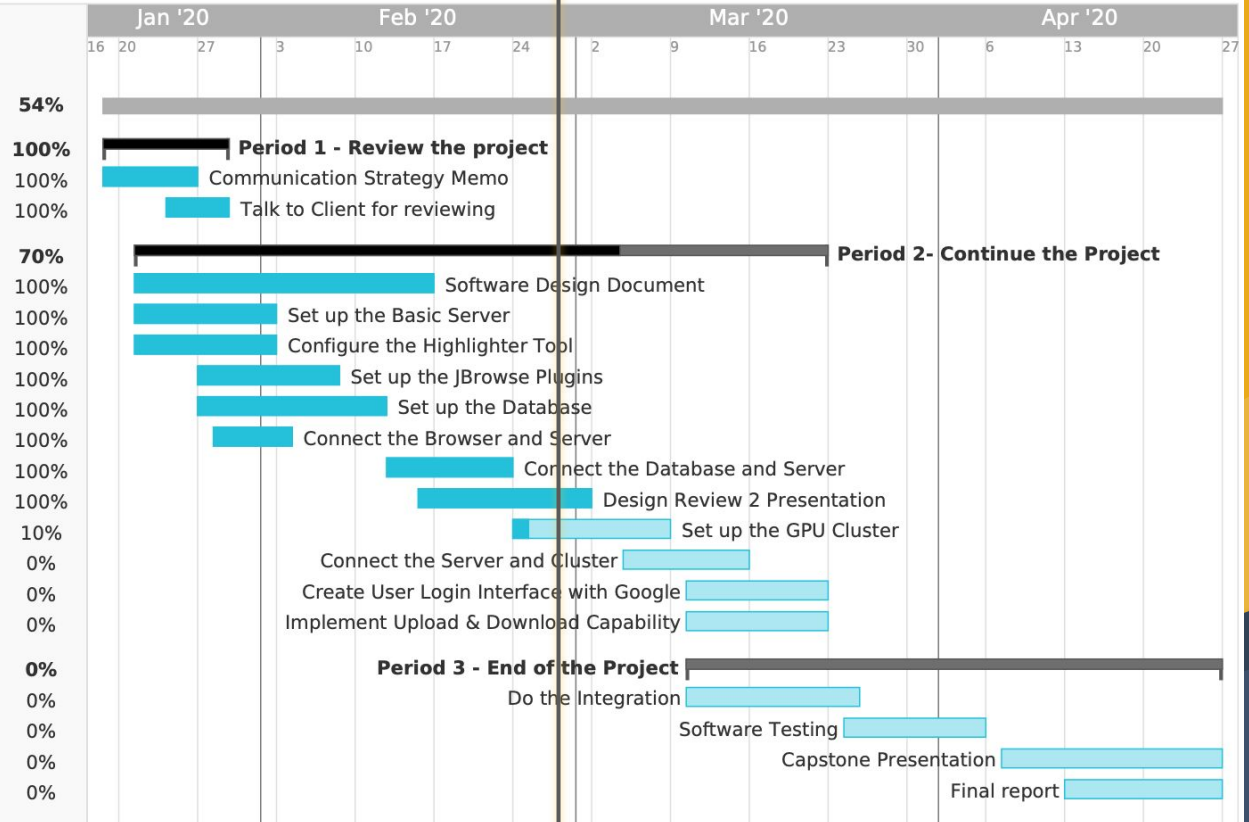
- Communication Strategy Memo
- Talk to Client for reviewing

### Period 2- Continue the Project

- Software Design Document
- Set up the Basic Server
- Configure the Highlighter Tool
- Set up the JBrowse Plugins
- Set up the Database
- Connect the Browser and Server
- Connect the Database and Server
- Design Review 2 Presentation
- Set up the GPU Cluster
- Connect the Server and Cluster
- Create User Login Interface with Google...
- Implement Upload & Download Capab...

### Period 3 - End of the Project

- Do the Integration
- Software Testing
- Capstone Presentation
- Final report



# Conclusion

## Problem

- Genetic diseases
- Current ML is inaccurate

## Solution

- Free web app
- Graphical representation
- Supervised ML based on experts

## Overview

- Will help early detection
- Solutions are viable
- Confident in our ability



**Thank You**