Abstract

[Draw your reader in with an engaging abstract. It is typically a short summary of the document.   
When you’re ready to add your content, just click here and start typing.]

Major Field Project  
Assignment

Data Collection Group - Data Collection

|  |
| --- |
| Fidaulla Sharief, Arman |
| Jacob Mathew, Kevin |
| Jaramillo Mejia, Jose |
| Sivakumar, Sindhu Parkavi |
| Votarikari, Roopesh |
| Yarlagadda, Mukesh |

Princiapal Goal

* This group is responsible for getting data from S3 and preparing it for inclusion in a data warehouse / database.

Goals

* Collect and organize genomic data: The main objective would be to extract the relevant data from the UCSC Genome Browser and organize it appropriately for further analysis and use in the project.
* Develop data manipulation and analysis tools: The project may include the creation of tools and scripts that allow efficient manipulation of the extracted genomic data. These tools may include functions for filtering, transforming, combining, or any other process necessary to prepare the data for inclusion in a data warehouse or database.
* Perform analysis and exploration of genomic data: Once the data is collected and prepared, you can perform exploratory analysis to obtain relevant information. This can include the identification of genetic patterns, the search for specific genetic variants or any other type of analysis necessary for the project.
* Document and share the results: It is important to document your findings, including the methods used, the results obtained, and any relevant conclusions. You may also consider sharing your results and code through a GitHub repository or other means, so that other researchers or collaborators can benefit from your work.
* Keep Data and Tools Up to Date: The UCSC Genome Browser and associated data may be updated periodically. Therefore, one of the objectives can be to establish a process to keep the data updated in your work environment and to adapt the tools and scripts developed to work with the new data updates.

Tasks may include (but are not limited to) the following:



By the mid-way point, there will be an expectation of some progress on the midterm project.

The following groups will have a high expectation of progress at this stage:

Data Collection

Data Storage and Maintenance

Data Research

Instructions

By the mid-way point of the course, it will be expected that you have provided a substantial individual contribution to the overall field project. This can include adding known issues, assigning tasks (to yourself or others), making pull requests, including documentation or using comments and discussion to support others.

<https://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html>

<https://genome.ucsc.edu/goldenPath/help/ftp.html>

Goals

* Collect and organize genomic data: The main objective would be to extract the relevant data from the UCSC Genome Browser and organize it appropriately for further analysis and use in the project.
* Develop data manipulation and analysis tools: The project may include the creation of tools and scripts that allow efficient manipulation of the extracted genomic data. These tools may include functions for filtering, transforming, combining, or any other process necessary to prepare the data for inclusion in a data warehouse or database.
* Perform analysis and exploration of genomic data: Once the data is collected and prepared, you can perform exploratory analysis to obtain relevant information. This can include the identification of genetic patterns, the search for specific genetic variants or any other type of analysis necessary for the project.
* Document and share the results: It is important to document your findings, including the methods used, the results obtained, and any relevant conclusions. You may also consider sharing your results and code through a GitHub repository or other means, so that other researchers or collaborators can benefit from your work.
* Keep Data and Tools Up to Date: The UCSC Genome Browser and associated data may be updated periodically. Therefore, one of the objectives can be to establish a process to keep the data updated in your work environment and to adapt the tools and scripts developed to work with the new data updates.