# Design and Development of an Intuitive Web-Based Genome Browser and Annotation Tool for Visualizing and Analyzing Genetic Data

**CMT403 – MSc Computing**

School of Computer Science and Informatics, Cardiff University


**Author:**

Layth Alhakim (c1832345)

**Supervisor:**

Dr Alexia Zoumpoulaki


**9th September 2023**

Acknowledgments

Firstly, I would like to express my gratefulness to my family for supporting me on my endeavor to study. Especially my parents and younger sister, who provided me with all the motivation, support, and encouragement to pursue and complete my studies.

I would also like to show gratitude towards my supervisor, Dr Alexia Zoumpoulaki, who provided me with guidance and feedback throughout the dissertation process as well as maintaining faith in my capabilities.

Abstract

This project aims to design and develop a web-based visualization tool that enables researchers and users to explore, annotate, and analyse genetic data effectively. The tool will primarily focus on creating an intuitive and user-friendly interface, allowing users to visualize and interpret genomic regions of interest. In addition, the tool will include integrated functional information from external databases to provide functional analysis capabilities within the tool itself. By developing this web-based genome browser and annotation tool, the project aims to provide researchers with an accessible platform to be able to contribute to advancing genomic research through efficient exploration, interpretation, and collaboration among researchers in various fields. This study will pursue an in-depth analysis of combining the complexity of genomic data with usability and accessibility by taking into consideration programming languages utilized, developmental methodologies as well as scalability of the web tool to the advancement of computing and genomics research.

# Contents

# Introduction

## Background and Context

The complexity of genomics research has led to a rise in the need for reliable tools that may assist researchers in making effective use of genetic information. This study attempts to explore and contribute to this domain which includes research on the many online visualization tools that have been developed specifically for the examination of genetic data (Wang et al., 2013), as well as the development of a user-intuitive web-based Genome Browser for visualizing and analyzing Genetic Data. As the field of genomics progresses, it will become more vital to have access to technology that makes it simpler to explore, analyze, and collaborate on data due to the complexity of the data involved. As a result of the rapid advancements in genomic studies, ranging from the utilization of genomic technologies for conservation management (Segelbacher et al., 2021) to research into disease resistance (Ambuj Bhushan Jha et al., 2021), there is a vital requirement in the development of genomic visualization and analytical tools as they are a crucial method that allows researchers to apply their cognitive abilities and experiences to comprehend vast amounts of complex data in order to identify new patterns or generate hypotheses (Nusrat et al., 2019). In addition to the benefits genomic tools provide to the academic community, there are also relevant benefits to the general population. A key motive for this study is to identify a potential method in which genetic visualization tools can provide simplified information for the general population or academics from different fields of study regarding genomics, such as disease-related genes or the purpose of particular genes.

## The Aims and Objectives

The main problem the study aims to tackle is the lack of a user-intuitive platform that provides the necessary information, visualization, and interpretation of complex genomic data in a simple format with accessibility features that users of varying skill levels can utilize. Elaborating on the problem, the project aims to design and develop a web-based visualization tool that enables researchers and users to explore, annotate, and analyze genetic data effectively. The tool will primarily focus on creating an intuitive and user-friendly interface, allowing users to visualize and interpret genomic regions of interest. In addition, the tool will include integrated functional information from external databases to provide functional analysis capabilities within the tool itself. By developing this web-based genome browser and annotation tool, the project aims to provide researchers with an accessible platform to be able to contribute to advancing genomic research through efficient exploration, interpretation, and collaboration among researchers in various fields, as well as providing a simplified tool for the general population, who are less experienced in the field of interest, to obtain information regarding particular genes. Moreover, this research will allow us to evaluate the tool's efficacy in enhancing data interpretation.

## Rationale for the Study

Genomic research is undergoing exponential growth in data generation, encompassing diverse biological processes and encompassing an expanding array of species. In this modern era of increased IT development, practical visualization tools become indispensable for extracting meaningful insights. Existing genomic visualization tools have made substantial contributions; however, they often exhibit limitations when applied to intricate datasets. These limitations include challenges in comprehending multidimensional data, deciphering complex genomic structures, and accommodating diverse user needs. For this research project, the focus is on providing users with a simplified tool that is relevant to their specific needs.

This study seeks to address these critical gaps in the current body of knowledge by introducing an innovative genomics visualization tool. By combining cutting-edge data visualization techniques with insights from the fields of genomics and computer science, the study aims to create a tool that transcends the limitations of current solutions by utilizing features from multiple sources and integrating them into a single tool. This tool will provide a comprehensive view of genomic data. It will enable researchers to navigate intricate datasets effortlessly, find specific genes based on their gene symbol, and extract biologically relevant information more effectively.

## Impact on the field of study

This research can provide multiple potential contributions to the computational development of genomic analysis and search tools. Firstly, it offers a resource to genomics researchers, allowing them to gain insights into complex genomic data. Secondly, it contributes to the advancement of computational biology, fostering the development of user-friendly, high-impact visualization tools with an emphasis on accessibility. Lastly, it paves the way for enhanced collaboration between computer scientists and biologists, synergizing their efforts.

## Scope and Limitations

This research focuses on the development and evaluation of a novel web-based genomics visualization tool, emphasizing its usability, accessibility, and effectiveness for researchers and the broader academic community. The tool aims to facilitate the exploration and analysis of genetic data, aiding in tasks such as identifying genes based on gene symbols and extracting biologically significant information from complex genomic datasets. The study encompasses various functionalities, including user-friendly interfaces, interactive visualization features, and integration with genomic databases. However, it is essential to note that this research does not delve into the development of new genomic algorithms or computational methods. Instead, it leverages existing genomic data and visualization techniques to create an integrated platform (Kim et al., 2015). While the tool's potential contributions are substantial, certain limitations should be acknowledged. Due to resource constraints, the tool's compatibility with specific genomic datasets and its scalability for extensive datasets will be tested within reasonable bounds. Additionally, the study will primarily focus on the usability and accessibility aspects, leaving in-depth biological analyses and interpretations to domain experts.

## The Structure of the Study

- An introduction that provides an introduction to the field of genomics research and outlines the research problem, objectives, and rationale.
- A literature review that delves into existing research related to genomics visualization tools, highlighting their limitations and exploring relevant theories and concepts.
- A methodology that explains the research methods used in developing the Genome Browser and discusses system design and user testing.

- The results, presenting the outcomes of the tool's development, including features, functionality, and user interface.
- The discussion will focus on analyzing and interpreting the results, discussing the implications and contributions of the tool to genomics research, and any limitations encountered during development.
- The conclusion summarizes the key findings, highlights limitations, and suggests avenues for future research.

# Literature Review

## Overview of Genomic Data Visualization

The genome browser is a graphical user interface that allows users to explore the genome as well as the annotation data and sequencing that are linked with it. Web-based genome browsers may be loosely separated into two categories: those that are generic and cover a wide variety of species and those that are species-specific and concentrate on a particular species (Nusrat et al. 2019). Within the large field of genomics research, it is essential to develop effective tools that are able to properly translate such data into coherent visual representations. This will allow researchers from a range of backgrounds to have an easier time understanding complex genetic data. This section provides an in-depth analysis of a variety of genomic data visualization tools, focusing on the important role these tools play in enhancing knowledge of genetic data.

Genome browsers, like the widely used Circos tool developed by Krzywinski and Birol (2011), show comparative genomics in a visually attractive manner. Researchers may use these web browsers to generate circular plots to display information about genes, gene expressions, and variations. This visualization technique aids in the detection and analysis of linkages and patterns within genomes. The interaction between genetic components becomes evident, facilitating the clarification of evolutionary and functional knowledge.

Heatmaps are a commonly used visualization method that allows for the depiction of genetic data with several dimensions. The provided visual representations provide a concise but comprehensive portrayal of gene expression patterns, facilitating the identification of patterns, groupings, and exceptional cases within extensive datasets for researchers. As scholars persist in confronting the intricacy of omics data, the visualization techniques elucidated by Gehlenborg et al. in their publication "Visualization of Omics Data for Systems Biology" (2010) offer significant contributions towards effectively communicating biological findings derived from these extensive datasets.

Network visualizations are very effective methods for revealing the interconnections of proteins, genes, as well as other molecular constituents. The visualizations use nodes and edges as a means to depict constituents and their interconnections inside complex chemical networks. Visual representations have the capacity to reveal hidden associations, hence facilitating researchers in comprehending intricate regulatory networks and signaling pathways contained within the genome.

Although the use of visualization tools offers a method to effectively explore the intricate nature of genomic data, there are still persistent issues in accurately depicting the extensive scale and complexity of these datasets. The considerable quantity of genetic data is a significant obstacle in creating a representation that is both comprehensive and easily understandable. It is crucial to achieve a harmonious blend of presenting intricate particulars and effectively communicating a coherent storyline.

In the present project, which centers on the creation of a web-based genome browser as well as annotation tool, it is crucial to comprehend the capabilities and constraints associated with various display methods. This project intends to create an interface that allows researchers to effectively examine, annotate, and assess genetic data by making use of existing information from academic sources and the benefits of visualization methods.

## Web-Based Genome Browsers

The scientific community has been profoundly affected by the widespread use of web-based genome browsers, which provide interactive platforms for studying and interpreting genomic material. Examples of such browsers that have proved useful for exploring and analyzing complex genomic data are the Integrative Genomics Viewer (IGV), the UCSC Genome Browser, Ensembl genome browser and JBrowse.

### Integrative Genomics Viewer (IGV):

The Integrative Genomics Viewer (IGV), which is well-known for its powerful capabilities and versatility, is able to show a great deal of different forms of genomic data. Researchers are given the ability to rapidly combine information from a wide variety of sources, such as functional annotations, alignment data, and variant calls, thanks to its user-friendly interface. IGV's capacity to simultaneously display many tracks is very useful for studying the interplay between various genetic factors. IGV is a useful tool for both individual and collaborative research projects because of its capacity to detect genomic variations and patterns (Thorvaldsdottir et al., 2013. Its user-friendliness also makes it accessible to researchers with varying levels of expertise, which helps increase representation of underrepresented groups in genomics.

### JBrowse:

JBrowse employs an innovative strategy for genome visualization by using state-of-the-art online technology to provide a dynamic and interactive user interface. The prioritization of expedited data loading and efficient navigation in the system guarantees that researchers may efficiently navigate genomes with diverse levels of complexity. One distinguishing feature of JBrowse is its use of a "linear synteny" perspective, enabling effective cross-species comparisons. This characteristic has significant value in the context of investigating evolutionary links and conserved genetic components. The interface of JBrowse exhibits a contemporary and streamlined design, therefore enhancing the process of investigating genomic areas and creating a captivating user experience. This exemplifies the capacity of online technologies to fundamentally modify the manner in which researchers conduct analysis on genetic data, as shown by Skinner et al. (2009).

## The UCSC Genome Browser:

The UCSC Genome Browser has been around longer than almost any other online genome browser. Its versatility in representing information and the breadth of its data sets have earned it widespread acclaim. DNA sequence, protein-protein interactions, gene annotations, and gene expression data are only few of the many types of genomic information that can be accessed using the UCSC Genome Browser. Genome alignments, Genomic tracks, and comparative genomics are just a few of the visualization tools available in the UCSC Genome Browser. Users may rapidly discover the desired genetic data by using the system's robust search and filtering capabilities. Users and developers alike actively participate in the growth and upkeep of the UCSC Genome Browser (Karolchik 2003).

The UCSC Genome Browser facilitates interactive exploration of genomic areas and data via functionalities like as zooming in and out, panning, and feature selection. The UCSC Genome Browser facilitates the generation of personalized visual representations by integrating several genomic tracks and data categories. Collaboration is facilitated by the provision of a functionality that enables users to disseminate their visual representations and comments to other users. The UCSC Genome Browser has a substantial historical background and a robust community of users and developers, hence ensuring its effective maintenance and support. The repository encompasses a diverse array of genetic data, making it a very useful asset for a multitude of research endeavors. The UCSC Genome Browser provides users with a diverse array of viewing choices, enabling them to investigate genetic material via various means.

## Ensembl Genome Browser:

The Ensembl Genome Browser is an additional popular online genome browser. Its ease of use and ability to combine information from several sources have made it well-known. Researchers in the fields of genetics, genomics, and molecular biology have access to a variety of publicly accessible genomic data through the Ensembl Genome Browser (Newman et al. 2018).

The Ensembl Genome Browser is accessible and simple to use due to its intuitive design. NCBI, UniProt, and the Reactome are just a few of the sources it incorporates. The newest findings in genomics may be found here, and it is continually updated with new data and features (Birney 2004).

Users may zoom in and out, pan around, and pick and choose features as they navigate genomic areas and data in the Ensembl Genome Browser. Combining several genomic tracks and forms of data to create your own unique infographics is also a key feature that differentiates Ensmebl from other genome browsers. It encourages teamwork by facilitating the dissemination of user-generated visualizations and annotations. The browser combines information from several sources to provide consumers a full picture of genetic information.

The value of accessible and user-friendly tools for genomics research is underscored by the introduction of web-based genome browsers. The capabilities of these tools allow researchers to focus their analysis on particular genomic areas, provide annotations for genetic variations, and pinpoint elements with functional significance. These web browsers not only enhance the capabilities of individual researchers, but also promote cooperation by facilitating the exchange of data and facilitating participatory debates around genetic discoveries.

## Usability in Bioinformatics Tools

The usability and user experience of tools play a vital role in achieving effective research results in the field of bioinformatics, where the complex analysis of genetic data is of essential importance. Applying user-centered design concepts is crucial to the success of bioinformatics tool development since it ensures that scientists will be able to take full use of the tools' capabilities.

Beard et al. (2012) conduct a comprehensive investigation of user-centered design as it relates to the creation of bioinformatics software. The research highlights the need of including end-users into the design process from the very beginning. Developers may learn more about the difficulties and needs of actual users if they consult with researchers who use the tools in their work. This iterative method not only produces tools that are tailored to user requirements, but also encourages participation and ownership from the research community as a whole.

Also, Bell et al. (2009) argue that user-centered design in bioinformatics should emphasize on the inclusion of visualization and interaction. Complex genomic data may be intuitively understood with the use of visual representations, and users can explore the data in a way that best matches their research needs with the help of interactive features. dynamic filtering and adjustable parameters and are only two examples of how the tools may be tailored to the requirements of individual researchers, and they're also highlighted in the paper.

It's also crucial to provide users with enough training and ongoing assistance. Comprehensive documentation, tutorials, and responsive user assistance services are highlighted by Erguler et al. (2016) as crucial to improving the usability of bioinformatics applications. Having easily accessible tools that help users through different tasks and fix possible difficulties enhances an already intuitive design.

To guarantee their accessibility and efficacy for the scientists that utilize them, bioinformatics tools must be easy to use. Many bioinformatics tools are difficult to use and need for specialized knowledge from their operators. So, it's crucial that these instruments be created with ease of use in mind. The relevance of user-friendliness in bioinformatics tools for cancer immunotherapy target finding is discussed by Olsen et al. (2014). They suggest that bioinformatics tools should be accessible to researchers without extensive computer science training. Scientists should be able to easily communicate data and findings with one another, thus they suggest that bioinformatics tools should be built with this in mind.

In order to integrate transcriptional, proteomic, and interactome data, Huang and Fraenkel (2010) address the application of user-centered design concepts in the creation of bioinformatics tools. They claim that user-centered design is fundamental to making effective software for people with specific requirements. Another point they make is that user-centered design may help people make fewer mistakes while using these technologies.

Taken as a whole, these studies demonstrate how important it is to prioritize user friendliness while developing bioinformatics software. Taking a user-centric approach to development helps improve tools for working with genomic data, increasing user involvement, and speeding up discovery and analysis. Insights from these studies will guide the design process for the present project, which aims to create a web-based genome browser and annotation tool that prioritizes user requirements and ushers in a new age of accessible and user-friendly genomics research.

## Analysis Tools and Functionalities

The integration of bioinformatics analytic tools into web-based visualization systems creates a robust environment in which scientists may not only view but also conduct in-depth analyses of complicated genetic data. The complexity of the genome may be better understood with the help of the many analytic tools made available by these platforms.

### Variant Calling Analysis:

Variant calling, the process of identifying specific variants in a population or an individual's genome, is a cornerstone of genomics research. Researchers may identify structural changes, insertions, deletions, and single nucleotide polymorphisms (SNPs) by integrating variant calling methods into web-based visualization systems. Using these methods, scientists may better understand genetic variation, illness links, and evolutionary trends. IoIO is an NGS data management system tailored for detecting SNPs in wheat, as shown by research by Simillion et al. (2017) and others. This integration considerably improves the capacity to detect evolutionary trends and identify genetic diversity in complicated genomes.

### Gene Expression Analysis:

Understanding the roles genes play in a variety of biological processes relies on identifying patterns of gene expression. Scientists may now assess differences in gene expression in response to environmental or cellular context using web-based platforms that combine methodologies for analyzing gene expression. These techniques facilitate the identification of genes associated with diseases or bodily processes, from which hypotheses may be formulated.

**Functional Enrichment Analysis:**

The biological significance of gene lists generated from studies can only be understood with the use of functional enrichment analysis software. These programs find biological concepts, pathways, or gene ontologies that are overrepresented in gene lists by comparing the lists to databases of functional annotations. Researchers may better understand the biological significance of their genomic discoveries with the use of web-based technologies that include functional enrichment analysis.

**Pathway Analysis:**

By relating genes to biochemical pathways, pathway analysis software helps scientists make sense of seemingly incomprehensible biological processes. Pathway analysis that is built into web-based platforms provides researchers with a dynamic setting for visualizing the interplay and correlation among proteins, genes, and metabolites. As a result, it's easier to pinpoint crucial participants and regulatory processes in biological circuits.

Web-based visualization solutions that include these analytical tools improve the research process by bridging the gap between data exploration and in-depth analysis. Rapidly discovering functional implications, illness correlations, and mechanistic discoveries follows the visualization of genomic areas of interest. These features serve as a road map for developing a comprehensive platform that allows researchers to not only browse and annotate genetic data, but also conduct in-depth analyses of it, resulting in a more complete understanding of the complexities of the genome.

## Data Security and Privacy in Genomics

While the incorporation of visualization tools for genetic data can lead to revolutionary new possibilities for study, it also raises significant ethical concerns and hurdles in protecting sensitive information. There is an increasing need to safeguard private genetic data as the field of genomics matures, and the volume of collected data grows. In this part, the many aspects of data security and privacy in genomic data visualization tools are considered, looking at research and frameworks that show how to manage these issues effectively. Genomic data may be visualized using these programs, and the results can be rather specific and revealing, which might be used for identification or to divulge private health information.

Genomic data visualization tools may be designed in a variety of ways to protect users' privacy and security. Data encryption before visualization is one method. This may make it harder for hackers and other malicious actors to obtain the information. Another option is to visualize anonymized data. This makes it harder to identify people by stripping out their personal information from the data. There is no silver bullet for the problems of data security and privacy in genomic data visualization programs. The appropriate method to adopt will change based on the data's sensitivity and the requirements of the consumers.

### Addressing Confidentiality and Informed Consent:

Confidentiality is of the utmost significance in the genomic era, when information about an individual's health, family history, and ancestry might be revealed. Rego et al.'s (2019) research examines the complex relationship between privacy and permission in the field of genomic sequencing. The researchers highlight the difficulty of compromising between releasing data for the sake of scientific progress and protecting individuals' right to privacy. They emphasize the need of open dialogue throughout the informed consent process and the necessity of providing participants with thorough information regarding the collection, storage, and dissemination of their data. The ethical obligation to protect participants' privacy while making use of genomics research is highlighted by this method, which also helps researchers build trustworthy relationships with their participants.

## Best Practices for Data Protection:

The work by Brodtkorb et al. (2013) sheds light on the larger issue of data protection using GPU computing in R, even though it does not directly pertain to genomics. This research fits nicely with the changing environment of genomics by using computational ways to improve data security. The genomics industry has made significant strides in protecting sensitive data by adopting secure computer environments, encryption mechanisms, and data anonymization techniques. The emphasis on safe computing procedures in this research highlights the need for advanced safeguards beyond those typically used with stored data. Genomics data visualization tools need to be built with security and privacy in mind. Protecting people' privacy and maintaining the secrecy of genetic data may be aided by adhering to best practices.

Data security and privacy considerations in the context of genetic data visualization tools are complex, as shown by the combined findings of these research. These discoveries inform the creation of robust security measures as the team begins work on a web-based genome viewer and annotation tool. The project's goal is to develop a platform that not only enables researchers to delve into genetic complexities but also protects the privacy and anonymity of individuals whose genetic data contributes to advancing genomics research by implementing secure computational environments, adhering to best practices, and ensuring transparent communication with users. Lessons from these research may be incorporated into the design process to strike a better balance between innovative potential and ethical concerns.

## Web Tool Development and Technology Stacks

A web tool's scalability, architecture, and development process are all heavily influenced by the technological stack on which they are built. Technologies like Flask, Node.js, and Ruby on Rails stand out because of the unique benefits they provide and the way they can be tailored to meet the needs of individual projects. Find below a summary comparison of these many technological approaches:

### Node.js:

The V8 JavaScript engine that Node.js was developed on is well praised for its asynchronous, event-driven design. It's particularly useful for time-sensitive situations that need instantaneous responses. In particular, Node.js shines when dealing with data-intensive tasks like those found in streaming and communication programs. Since it is single-threaded and non-blocking, it is well-suited for applications that must deal with a large number of simultaneous connections. Rapid progress is also enabled by the extensive ecosystem of packages and modules available for Node.js through npm (Node Package Manager). However, its computational limitations may be exacerbated by its single-threaded design (Node.js 2023).

### Flask:

Flask is a lightweight Python web framework that puts an emphasis on code readability and reusability. The framework takes a simple approach, only providing the bare minimum of features necessary for developers to do their work. Flask is a adaptable framework that works well for small to medium-sized applications. Its modular design makes it simple to include other libraries, and its ease of use makes it ideal for fast development. Flask's lack of pre-built functionality makes it very flexible, although it may need further configuration for more sophisticated applications.

### Ruby on Rails:

The Ruby programming language has a web application framework called Ruby on Rails (or just Rails). Rails is built on the notion of "convention over configuration," which simplifies development by mandating standardized practices. Using best practices is pushed for and developer efficiency is emphasized (Małek 2020). Scaffolding and an Object-Relational Mapping (ORM) are two of Rails' inbuilt capabilities that help speed up development by eliminating the need for repeated writing. It's the best option for new ventures and time-sensitive endeavors. But the rules might make it harder to adapt to unusual circumstances.

### Comparison

Many different technologies, each with its own advantages and disadvantages, go into the making of web tools. Node.js, Flask, and Ruby on Rails are three popular alternatives that may be tailored to the needs of a particular project. Because of its asynchronous nature, Node.js is well suited for use in mission-critical programs. While Ruby on Rails streamlines development via the use of convention over configuration, Flask provides simplicity and adaptability. Choosing a technological stack is crucial since it affects the project's development time, scalability, and overall architecture.

Both the Fernandez-Cuesta and Peifer (2017) study as well as the Gonzalez and Uribe (2011) study shed light on several facets of genomics. While these works aren't directly connected to building online tools, they do further our knowledge of genetics and data visualization. Nanoscale proteomics is discussed by Fernandez-Cuesta and Peifer (2017), providing insight into the developing methods in proteomics. To help visualize related datasets, Chen and Boutros (2011) develop VennDiagram, a R program for creating individualized Venn and Euler diagrams. While these studies may not have much to do with web tool development technologies, they do provide important background for the relevance of excellent tools and platforms in promoting research results in the interdisciplinary world of genomics research and data visualization.

## Explanations for Choosing Flask as the Backend Framework:

After considering several available technologies, the tool's development process settled on Flask as the backend framework for the endeavor. The goals and scope of the project provided the foundation for this decision. Flask is lauded for being lightweight and flexible, allowing programmers to create modules that meet the needs of their own projects. Its ease of use and quick development time make it a good choice for meeting the deadline for developing a web-based genome browser and annotation tool. The adaptability of Flask also allows for the easy incorporation of other libraries and tools, expanding the framework's already impressive capabilities. This decision also conforms to the principles of staying within the technological stack, keeping the project's attention where it needs to be: on efficient display and analysis of genetic data.

## Gaps and Challenges:

The prospects and difficulties for creating powerful web-based visualization tools for genetic data are growing as quickly as the discipline of genomics itself. Here, the complexities of building scalable, user-friendly online tools for genomics research and the gaps and limitations in the current literature on web-based display of genetic data are examined.

## Identifying Gaps and Limitations:

The present literature still has significant holes and restrictions, despite the progress made in web-based visualization tools for genomic data. First, there is a lack of tools that can effectively combine both structural genomics (genome structure and order) and functional genomics (gene expression, protein interactions). Another difficulty is dealing with very large datasets, such as whole-genome sequencing. While there are tools that excel at displaying certain genomic areas, there aren't many that can handle the task of studying a full genome. Furthermore, novel methodologies are required to give relevant insights into spatial links within the genome via the display of three-dimensional genomic interactions, which is still a growing topic.

## Challenges in Designing User-Friendly and Scalable Web Tools:

- **Data Complexity:** Genomic data is notoriously difficult to parse since it includes so many different types of biological information (Bauch et al., 2011). It may be difficult to design online tools that adequately convey this complexity to users without overburdening them. Finding a happy medium between detailed data display and intuitive operation is essential.

- **Interactivity:** Maintaining performance while allowing for interaction is difficult. Users need interfaces that are quick to respond so they can zoom, pan, and query genetic data without delay. A smooth user experience relies on fast data loading and rendering technologies.
- **Usability:** Effective online tools prioritize their users in all design phases (Kaur & Sharma, 2018). Researchers come from a wide range of disciplines and experience levels, making it difficult to anticipate their requirements and meet their expectations.

For widespread use, it is crucial to design user interfaces that are accessible to both beginners and specialists.

- **Scalability:** Massive genomic datasets provide difficulties in scalability for storage, retrieval, and display. Growing datasets need web tools that can keep up with them while yet allowing for fast access and analysis. To deal with ever-increasing data quantities without degrading performance, scalability is a must.
- **Data Security and Privacy:** Genomic data is particularly private and should be protected. Maintaining privacy and confidentiality while allowing for data exchange and cooperation is a difficult task. It is critical to find a happy medium between data availability and security.
- **Cross-Disciplinary Collaboration:** Collaboration Across Disciplines: Scientists studying genomics come from many different backgrounds, including medicine (Raman, 2011), computer science, and biology. It is important to take into account the demands and jargon of different fields when designing online solutions that promote teamwork.
- **Sustainability and upkeep:** Web instruments must develop with developing technology and changing research needs. In the fast-evolving field of genomics, it is difficult to ensure long-term sustainability, upgrades, and maintenance.

The complexity of these issues necessitates a collaborative effort across specialists in fields as diverse as genetics, software engineering, user experience design, and data management. In light of these limitations, the project's goal is to provide a web-based genome browser and annotation tool that facilitates the rapid exploration and analysis of genetic material while simultaneously addressing these issues.

## Conclusion

In conclusion, the ecosystem of web-based visualization tools for genetic data has been thoroughly surveyed thanks to the literature study. The papers that were examined provided insight into the many facets of genomics research, such as data visualization, data security, analytic tools, and technology choice. Several major conclusions may be drawn from the available research:

**Advancements in Genomic Visualization**: The emergence of genomic visualization tools has changed the way scientists interact with and understand genetic data. Multiple technologies, such as genome browsers and functional enrichment analysis platforms, have been shown to be helpful in providing deeper understanding of genomic intricacies.

**Ethical Considerations:** The examined papers highlight the significance of openness, informed consent, and data privacy in the genomics age, all of which are important ethical considerations. To guarantee ethical data management, it's crucial to address issues about privacy and secrecy.

**Technology Stacks:** By comparing popular technologies such as Node.js, Flask, and Ruby on Rails, we can see how crucial it is to choose the right backend framework for the task. Each technology has its own set of benefits; picking the right one depends on how well it fits in with the overall scope of the project.

**Usability and Scalability:** Designing online tools that are both accessible and scalable presents a number of challenges. Tools that accommodate a wide range of user requirements and data complexity should prioritize interaction, usability, and data security, as highlighted by the studied literature.

**Gaps and Challenges:** The current body of research shows that there are holes in our ability to visualize whole genomes and combine structural and functional genomics. Complexity of data, interaction, scalability, data security, usability, and working with experts from other fields are all obstacles to overcome.

The results of the literature study strongly support the project's goal of creating a web-based genome browser and annotation tool. The tool's goal is to unite structural and functional genomics visualization by resolving the issues that have been discovered. Researchers should be able to use its user-friendly interface, built-in analytic tools, and secure data with confidence. Despite the ever-changing nature of the genomics environment, this tool represents an important milestone in the progression of genetic research by facilitating effective exploration, interpretation, and cross-disciplinary cooperation. An efficient web-based visualization tool has the potential to radically alter the trajectory of genomics research, and this project highlights that promise by tracing the path from literature review to actual implementation.

# Methodology

## Introduction to Methodology

The Methodology section serves as the guiding framework for the development of the web-based Genome Browser. This essential component of the research outlines the systematic approach taken to design and build the tool, ensuring its effectiveness in addressing the needs of genomics researchers and enthusiasts.

In the rapidly evolving field of genomics, the creation of an intuitive and robust visualization tool requires a structured and iterative process. This segment will describe and explain the steps, processes, and methodologies utilized to reduce the gap between complex genetic data and user-friendly accessibility. It also offers insights into the strategies chosen to navigate challenges and leverage opportunities in the creation of a tool that enhances the genomics research landscape.

## Web Tool Development Approach

In the development of the web tool for genomic data visualization, a strategic approach was adopted that aligned with the unique requirements and objectives of the project. This approach was based on leveraging established web development technologies and frameworks while tailoring them to suit the specific needs of genomics research as well as the development of the web tool. The methods chosen opted for a more tailored solution that addressed the intricacies of genomic data without unnecessary complexities.

### Choice of Technology Stack

The technology stack was carefully selected to ensure efficiency and performance while maintaining a user-friendly interface. The core technologies used included HTML, CSS, and JavaScript. These foundational web technologies provided the building blocks for creating interactive and responsive web pages.

To handle the server-side logic and database management, Python was employed with the Flask framework. Flask is a lightweight and flexible microweb framework (Adam, 2023), making it an ideal choice for the project's scale and objectives. It enabled the web tool to efficiently handle user authentication, database interactions, and routing.

SQLite, a self-contained, serverless, and zero-configuration relational database engine, was utilized for the database management system. This decision was made to simplify deployment and ensure seamless database integration with the web application.

For the database management system, SQLite, a self-contained, serverless, and zero-configuration relational database engine was utilized. This choice was made to simplify deployment and ensure optimum database integration with the web application.

## Development from Scratch:

The web tool was developed from scratch, reducing potential restrictions over the architecture and design and enabling complete control. This approach facilitated the customization of features, ensuring that the tool explicitly catered to the needs of genomic researchers and enthusiasts.

## Justification:

The decision to develop the web tool from scratch and avoid using React.js and D3.js stemmed from carefully considering the project's goals. While React.js and D3.js are potent libraries for data visualization, their extensive learning curves and associated complexities could have introduced unnecessary challenges for users who might need to be better versed in these technologies. By focusing on a more simple, custom-built solution, accessibility and ease of use were prioritized. The goal was to create a web tool that could be intuitively navigated and understood by users from diverse backgrounds, including researchers, educators, students, and the general public, without requiring prior knowledge of specialized libraries.

In conclusion, the web tool development approach highlights a commitment to delivering a user-friendly and effective solution for genomic data visualization, emphasizing simplicity, accessibility, and tailored functionality to meet the specific user's demands.

# System Architecture

## Overall Architecture:

The architecture of the web tool for genomic data visualization is designed to ensure efficiency, scalability, and user-friendliness. It encompasses several vital components and layers, each contributing to the seamless functioning of the system.

## Components:

1. **User Interface (UI):** The User Interface (UI) serves as the primary interaction point for users. The concept of the design is to be as intuitive and user-friendly as possible, enabling users to download and upload genomic data, visualize it through an embedded Genome Browser, and interact with the comment system without any issues.

2. **Web Application Layer**: This component acts as the intermediary between the user and the system's core functionalities. The responsibility of processing user requests, managing user sessions, and routing data is implemented here.

3. **Genome Browser Module**: The Genome Browser module is a vital component of the web tool. The Ensembl Genome Browser is integrated into the web tool, allowing users to fetch up-to-date, simple genomic data. In addition, the integration of UCSC Genome Browser's genome map provides a dynamic and interactive environment for visualizing genomic data and presents it in an accessible format.

4. **Comment System Module**: The Comment System module allows users to annotate and discuss specific genomic regions. It maintains a database of user comments, their associated positions, and relevant user information. Users can view existing comments and add new ones once logged in, encouraging collaboration and knowledge sharing.

5. **Server and Database Layer**: A Server and Database Layer contains the server logic and the database. The server handles HTTP requests, communicates with the Genome Browser and Comment System modules, and manages user sessions. The database, implemented using SQLite, securely stores user data, comments, and other relevant information.

6. **Data Storage and Integration**: The web tool incorporates data storage and integration mechanisms. This includes storing downloaded genomic data locally, storing comments on genome regions, managing user profiles, and ensuring data integrity. Additionally, integration with external resources, such as the UCSC Genome Browser, enables real-time data retrieval and visualization.

## Programming Languages and Technologies

The selection of programming languages, frameworks, libraries, and technologies for the development of the web tool was a critical decision. Each choice focused on achieving key project objectives, ensuring efficiency, and providing a seamless user experience.

### Python:

Python served as the primary programming language for the backend development of the web tool. The versatility, extensive library support, and readability of the language made it an ideal choice. The Flask framework, a lightweight and flexible Python web framework, was utilized to build the web application's core functionalities. Flask provided the necessary tools for handling HTTP requests, routing, and session management, streamlining the development process.

### JavaScript and AJAX:

JavaScript played a pivotal role in enhancing the web tool's interactivity. It enabled dynamic content updates, including real-time comments and the smooth scrolling feature. AJAX (Asynchronous JavaScript and XML) facilitated seamless communication between the front and back end, ensuring a responsive user experience.

### HTML and CSS:

HTML and CSS were fundamental in shaping the web tool's simple, accessible user interface, a prioritized objective in this study. HTML provided the structural foundation for creating web pages, while CSS was employed to style and layout elements, ensuring a visually appealing and intuitive UI.

### SQLite Database and SQL Alchemy (Object-Relational Mapper):

For data storage and management, the SQLite relational database management system was the most accommodating and suitable for the development process. SQLite is renowned for its lightweight design and simplicity, making it a suitable choice for this project. It effectively stores user profiles, comments, and other critical data while ensuring data integrity and security.

## Ensembl Genome Browser Integration:

Integrating the Ensembl Genome Browser was instrumental in providing users with a simple and precise method of obtaining genomic data using the gene symbol. This external resource leveraged RESTful APIs to fetch and display genomic data within the web tool, enriching the user experience.

## UCSC Genome Browser Integration and iframe:

One of the key elements of the web tool's development was the integration of the UCSC Genome Browser genome map, a renowned genomics resource. To achieve this, an iframe was utilized, which are HTML elements that enable the embedding of external web content directly into the platform. This integration allows users to view and analyze genomic data within the interface, facilitating a cohesive research experience. The tool strategically leveraged HTML, JavaScript, and Python to ensure the compatibility and smooth functioning of the UCSC Genome Browser within the web tool while maintaining a relationship with the Ensembl Genome Browser API. This approach enhances the user experience and underscores the commitment to providing researchers with a comprehensive genomics toolkit.

## Summary:

These programming languages and technologies were selected for their ability to synergize and contribute to the project's success while maintaining simplicity at the cost of minimum loss of functionality. Python and Flask streamlined backend development, while JavaScript and AJAX enriched the front end with dynamic features. HTML and CSS ensured a polished user interface, SQLite handled data management efficiently, the Ensembl Genome Browser integration provided actively updated genomic information, and the UCSC Genome

Browser genome map integration was essential in providing the users with the feature to visualize the gene of interest. Additionally, responsive design techniques made the web tool accessible to a broader audience.

In summary, carefully selecting programming languages and technologies played a pivotal role in shaping the web tool. These choices allowed the development of a user-friendly, efficient, and interactive genomic data exploration and annotation platform, aligning with the project's objectives and ensuring a seamless user experience.

## Database Design

The database design constitutes the foundational structure underpinning the web tool's functionality. In this project, SQLite, a lightweight and efficient relational database management system, was employed to manage data (Yang et al., 2018). The project's scale and requirements drove this choice, as SQLite seamlessly integrates with SQL Alchemy, a robust and widely-used Object-Relational Mapping (ORM) library for Python.

The database schema was developed to cater to the specific needs of the web tool. It encompasses the necessary interconnected tables, each serving as a repository for distinct data types. For instance, user information, comments, and references to genomic data are organized within their respective tables and maintain their interactivity features through the use of relationships. This is a critical factor for user interactivity; it was essential that the specific user would be able to post a comment at a particular position of the genome map, which would only be displayed at the position based on the genomic coordinates.

The schema's evolution was an iterative process, refined to meet evolving project requirements. An essential consideration during the design phase was scalability, a significant concern in genomics research due to the vast amount of data involved. However, due to the use of external databases and REST APIs, this concern was addressed effectively, simplifying the process. SQLite, in conjunction with SQL Alchemy, was adept at handling any other storage requirements, such as multitudes of users signing up or posting comments with no issues, allowing for the efficient storage and retrieval of potentially large datasets.

Data integrity and security were also important to consider during the database design. Robust encryption mechanisms were implemented to safeguard sensitive user data and research information, ensuring that the database adheres to best practices in data protection.

While implemented with SQLite and SQL Alchemy, this database design remains scalable and adaptable. It ensures that data is efficiently organized, secure, and readily accessible, supporting the diverse functionalities of the web tool. Its seamless integration with SQL Alchemy simplifies complex database operations, making it an ideal choice for this project's objectives.

## User Interface Design

Practical user interface design is essential to the success of any web-based tool. In developing this project, a thorough approach was undertaken to create an intuitive and user-friendly interface. The design strategy adhered to established principles and methodologies, primarily focusing on user experience (UX) and accessibility.

Figma, a collaborative interface design tool, played a significant role in this phase. It facilitated the creation of wireframes that provided a pictorial blueprint of the web tool's layout and functionality. These wireframes served as a foundation for discussions and iterations, allowing for the refinement of design elements.

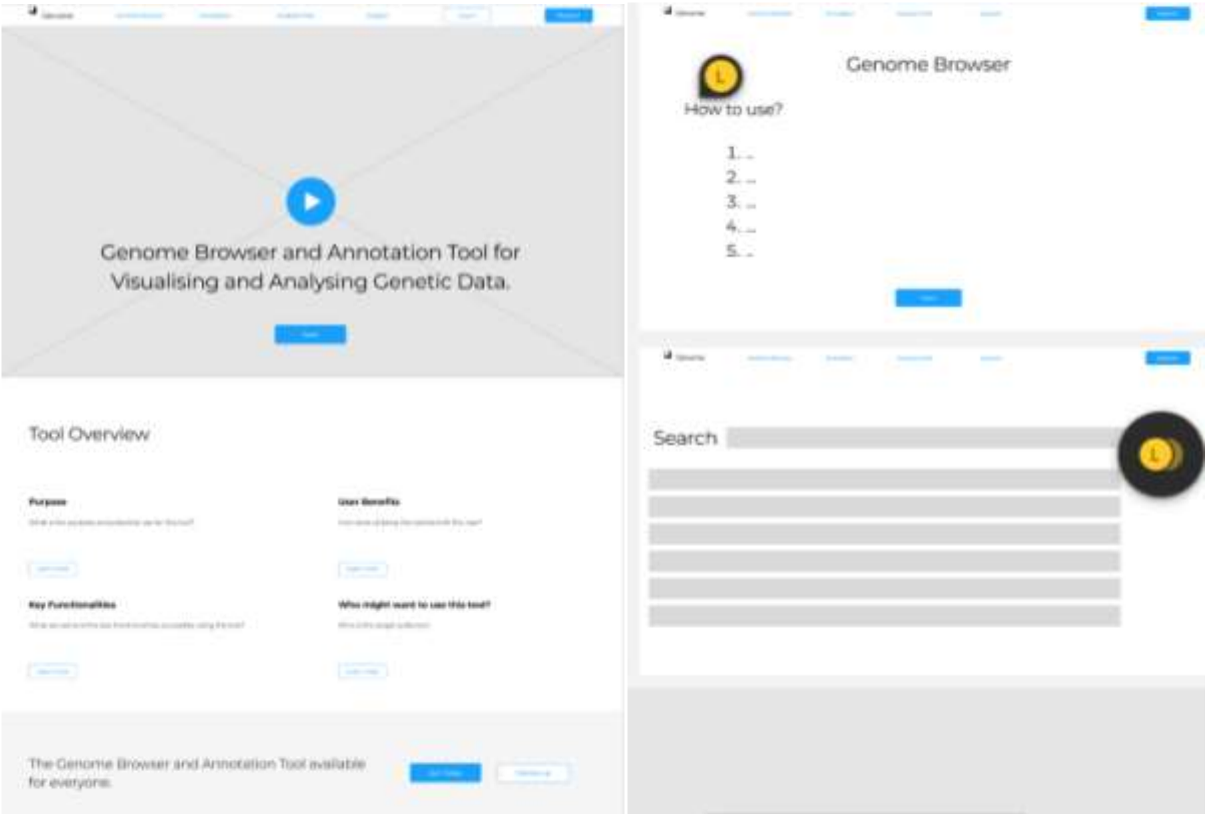


Figure 1: The utilization of Figma to design and create the Home page an Genome Browser pages of the webtool.

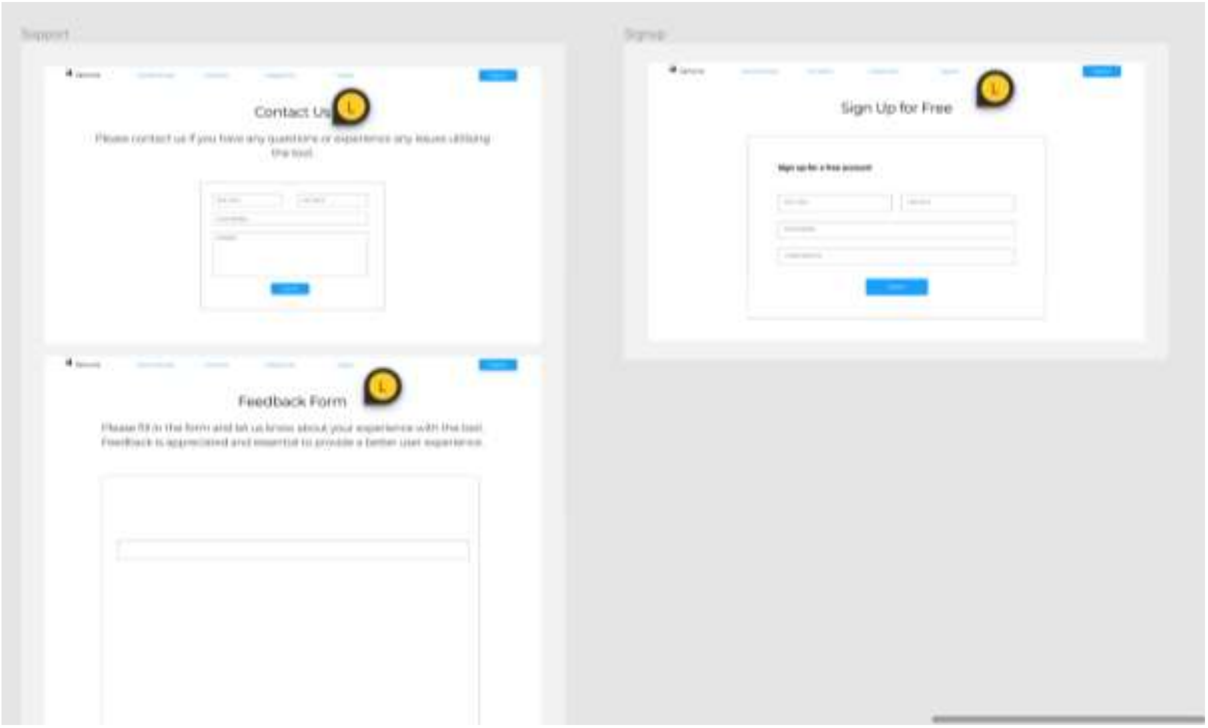


Figure 2: The design of the Support page and the Sign Up page using Figma. In addition to the expected feedback from to be used in gathering user data.

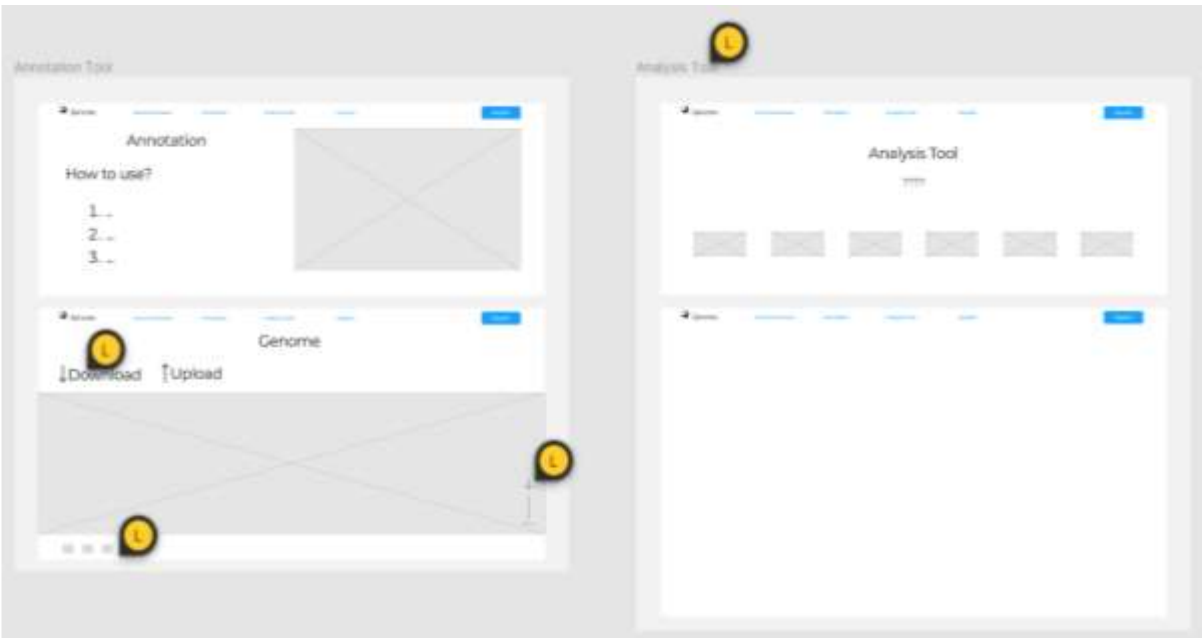


Figure 3: The design of the annotation page and the analysis tool page (discussed in the Discussion section of the project) using Figma.

Accessibility was a core consideration throughout the user interface design process. The web tool was designed to be inclusive, ensuring that users of varying abilities could navigate and interact with it seamlessly. Distinct yet simple color schemes and interactive elements were specifically chosen to ensure readability and usability.

Consistency in design was also prioritized. Recognizable design patterns and conventions were employed to create a sense of familiarity for users, simplifying and enhancing the overall user experience as much as possible.

One part of the study was expected to rely on feedback and improvement through user testing. Feedback sessions should have been conducted to gather insights and identify areas for enhancement. Ideally, these insights were expected to provide the knowledge to produce informed design adjustments, creating a user interface that aligns closely with user expectations and needs. This part was conducted differently than expected due to inadequate timing for ethical approval to gather user data.

In summary, the user interface design for this web tool was a deliberate and user-centric process. Employing Figma for wireframing, the design adhered to principles of accessibility, consistency, and iterative improvement. Although not identical to the expectation, the result is an interface that offers a positive user experience, effectively supporting the tool's objectives.

## Development Process

The development of the web tool followed a systematic and well-organized approach, particularly suited for a small-scale project. The methodology drew upon a modified Waterfall model, which emphasized a linear and sequential progression through project phases. Asana served as the primary platform for managing tasks and project milestones.

The development process can be delineated into several distinct phases:

1. Project Initiation: The project's commencement involved a comprehensive understanding of its objectives. Initial milestones included the finalization of wireframes and the user interface design, which set the foundation for the tool's aesthetics and functionality.
2. Technology Selection: An essential phase was carefully selecting programming languages, frameworks, libraries, and technologies for front and backend development. This milestone played a crucial role in establishing the technical framework for the web tool.
3. Frontend and Backend Development: Parallel development of the frontend and backend components occurred. Frontend development focused on crafting an intuitive and responsive user interface, while backend development handled data processing, implementing integrations, and database management.
4. API Integration: A pivotal achievement was successfully integrating the Ensembl Genome Browser API and the UCSC Genome Browser genome map iframe. These integrations empowered the web tool to access and display genomic data effectively.
5. User Testing and Iteration: User testing sessions were expected to be conducted regularly to gather feedback and identify areas for improvement; however, due to

delays in ethical approval to gather user data. This still needs to be accomplished. The iterative development approach would allow for further future adjustments based on user insights, ensuring that the tool meets user expectations.

6. Performance and Security Optimization: Ensuring optimal performance and robust security measures were paramount considerations. Milestones in this phase included performance tuning and implementing security protocols to safeguard user data, including using security modules to generate and check passwords (werkzeug.security module)
7. Documentation and User Guides: In preparation for user onboarding, comprehensive documentation, and user guides were created to provide essential resources for utilizing the tool effectively.
8. Project Completion: With all milestones achieved, the project concluded with submitting the completed web tool. It was ready to serve its intended purpose, offering a valuable resource for genomics researchers and enthusiasts.

Inspired by the Waterfall model and facilitated by Asana, this development methodology ensured the project's smooth progression, efficient task management, and successful delivery of a robust and user-centric web tool.

## Testing, Quality Assurance and Version Control

### Integration Testing:

Integration testing assessed the interactions of integrated external components ensuring successful functionality. It ensured that the various modules, including the genome browser and iframe tool integrations, seamlessly collaborated to deliver a unified and coherent user experience.

## Exception Handling:

In developing the web tool, exception-handling mechanisms were vital in ensuring reliability and efficient user interactions. The code incorporates comprehensive error-handling routines, as evidenced by implementing try-except blocks. These constructs capture and manage exceptions that may arise during runtime, preventing unexpected crashes and providing users with informative error messages. By systematically identifying and addressing potential issues, the web tool strives to maintain high stability and user-friendliness. The exception-handling mechanisms were evaluated through strict testing, reinforcing the tool's resilience and usability in real-world scenarios.

## Version Control:

The project's source code management relied on robust version control systems, primarily Git. Git proved instrumental in maintaining a structured and organized development process throughout the project's lifecycle. It allowed for systematic tracking of changes and collaboration, ensuring code integrity.

Git enabled the project to:

- Track Code Changes: Every modification, enhancement, or bug fix was systematically recorded, providing a historical timeline of the project's evolution.
- Code Integrity: With Git, code integrity was upheld. In case of issues or regressions, Git allowed easy code rollbacks to previous working states.

- Backup and Redundancy: Git inherently provides redundancy as the codebase is stored locally and remotely, enhancing data resilience.

In summary, Git's use as a version control system was pivotal to the project's success. It ensured a systematic, collaborative, and secure approach to managing the source code, contributing significantly to the development and maintenance of the web tool.

## Deployment and Hosting

The deployment and hosting strategy for the web tool adhered to this project's specific requirements and considerations. Unlike traditional web applications, this web tool is hosted locally, a choice guided by the project's scope and objectives.

Given that the primary users of the web tool are researchers and scientists within a controlled environment, local hosting proved to be the most suitable approach. This decision offers several advantages, including enhanced data security and confidentiality. By deploying the web tool within the organization's infrastructure, it ensures that sensitive genetic data remains on-premises, eliminating the need for external cloud-based hosting services.

# Results

## Sign Up

The web tool incorporates a user registration system that allows individuals to create accounts in order to be able to use certain features. The implementation is conducted through a dedicated "signup.html" file, in addition to associated routes, which users can access to begin the account creation process. The signup form includes fields for essential information input from the user, including username, email address, and password. The data is saved onto the SQL database to create a new user profile.

The application performs validation checks during the signup process to maintain data integrity and security. It verifies whether the entered username and email address are unique, preventing the creation of accounts with duplicate credentials. Moreover, password-hashing methods improve the security of stored user information.

As a user submits the signup form, the web application gathers the provided data and attempts to create a new user profile. Any errors or exceptions are handled, and appropriate messages are displayed to users if issues arise. Successful registration is acknowledged with a confirmation message, and users are directed to the login page to access their accounts.

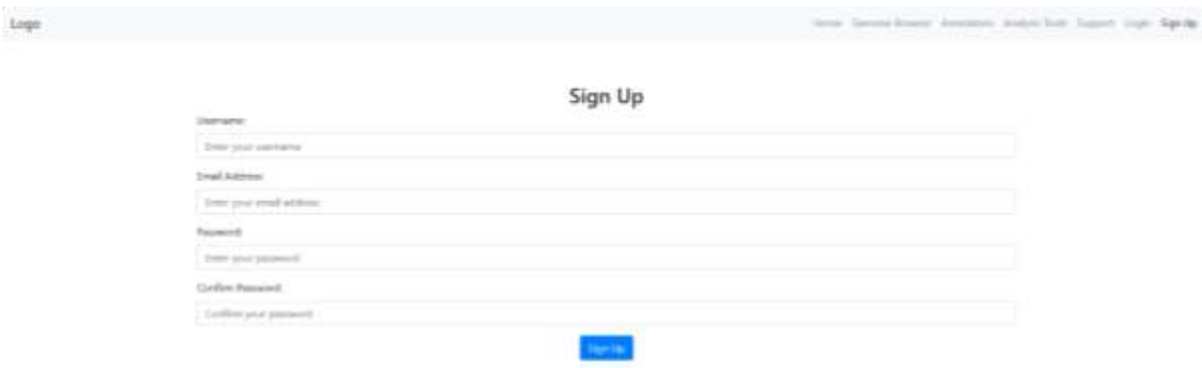


Figure 4: The "signup.html" file/page. As users select the "Sign Up" option on the navigation bar, they will be redirected to this page to complete the form resulting in the creation of their account.

## Log In

The web tool features an uncomplicated login system that permits registered users to access their accounts securely. Users can log in by selecting the login page through the navigation bar and providing their registered email address and password in the appropriate displayed fields. The login form is designed to verify the authenticity of user credentials before granting access to protected features of the tool, such as commenting on the annotation page.

The application validates the provided email and password against the stored user data in the backend. Successfully logging in redirects users to the tool's home page, granting access to all of the tool's current functionalities. In the event of invalid login credentials, the system provides appropriate error messages to users, ensuring transparency in the login process.

Integrating these signup and login features enhances user interaction with the web tool, allowing them to personalize their experience and access personalized features. These components align with best user authentication and data security practices, ensuring that user information remains confidential and secure.


Figure 5: The “login.html” file/page. As users select the “Login” option on the navigation bar, they will be redirected to this page to access the user account based on their saved credentials in the backend.

## Genome Browser Search Feature

The genome browser search feature of the web tool offers users a convenient way to explore genetic information. Users are presented with an intuitive search interface by selecting "Start" on the home page or "Genome Browser" on the navigation bar. They can input the gene symbol of a gene, such as "BRCA2" or "TNF", and by selecting the "Search" button, the tool queries the Ensembl database for the relevant information.

Behind the scenes, this search feature is supported by a Flask route named search_genome. When users initiate a search, the data is processed, and a query is made to the Ensembl REST API. This API call retrieves information about the specified genome from the Ensembl database based on the provided species (e.g., 'homo sapiens' for humans).

If the gene exists in the Ensembl database, the tool displays a concise summary, including its gene symbol, ID, species, biotype, and a brief description. This information aids researchers in quickly identifying the genome of interest.

Moreover, a significant convenience is provided by integrating an "Annotation" button within the search results. When clicked, this button directs users to an annotation page with detailed genomic information for further exploration and analysis.

This search functionality enhances the web tool's usability, making it a valuable resource for genomics researchers and enthusiasts.

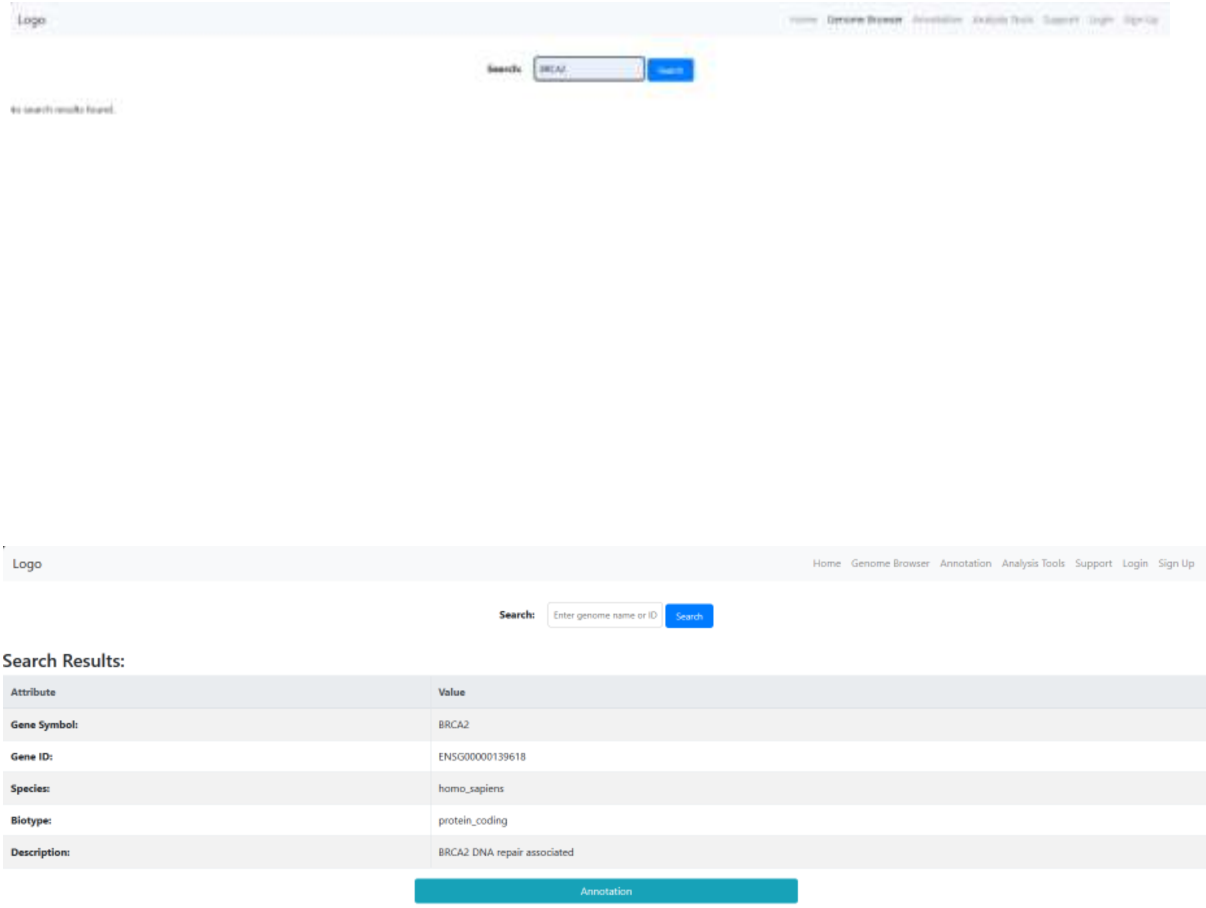


Figure 6: The "genome_browser.html" file/page. Users will have the ability to search for a gene based on the gene symbol, in the example "BRCA2" is used. As a result, users are provided with basic information from the Ensembl database after querying the REST API. The users are presented with an "Annotation" button that will redirect to the annotation page for more information on the genome as well as other features.

## Annotation

The core of the web tool is the annotation space, where users can interact with genomic data. This space offers several critical features:

- Genomic Data Display: Users can view genomic information using an embedded iframe. This iframe pulls data from the UCSC Genome Browser, allowing users to explore genome maps interactively.

- Real-time Comments: Users can post and read comments related to specific genomic regions directly within the annotation space. This collaborative feature promotes discussions and knowledge sharing among researchers.
- Upload and Download: Researchers can upload HTML files containing genomic data. The tool parses these files, identifying whether they contain the necessary iframe. If found, users are redirected to the annotation page for the specified genomic region. Additionally, users can download the iframe content as a separate HTML file for offline analysis or to upload in the future.

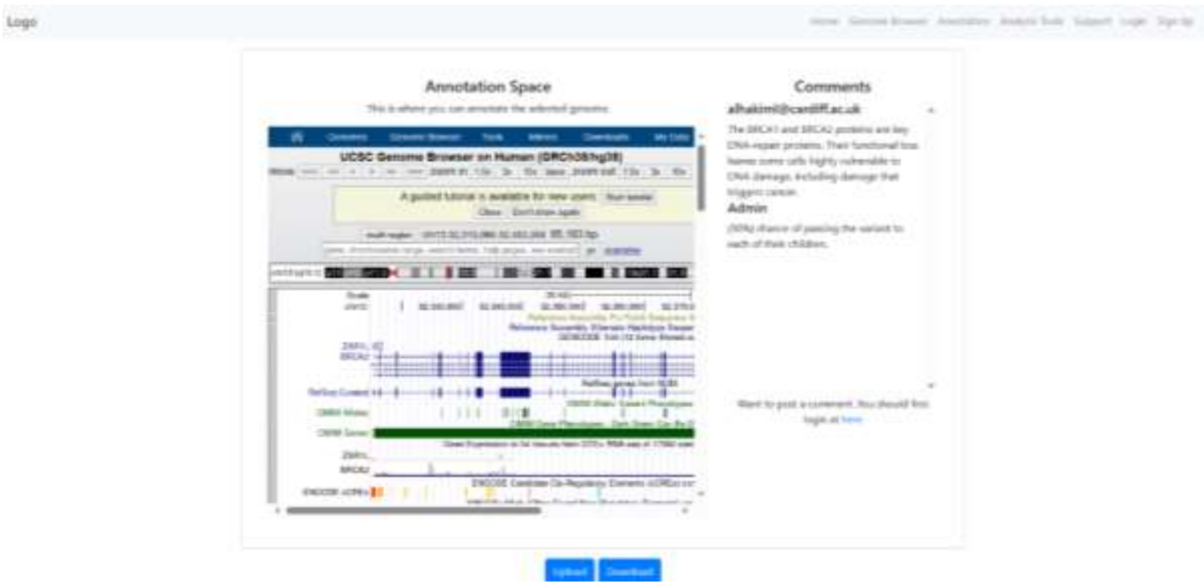

Figure 7: The “annotation.html” file/page. As shown in the image, users will be able to see the area of interest on the genome map on the left side, in this case “BRCA2” is used. On the right side, users will see the commenting section. An account is required for individuals to comment however everyone can view these comments.

It is important to note that to access the genome map, a user has to either utilize the upload/download feature or select the “Annotation” button after searching for a particular gene. The Ensembl Genome Browser API is used to obtain information regarding the genes, as shown previously on the Genome Browser page. However, the UCSC Genome Browser is used to display the region of the gene on the genome map. In other words, searching

“BRCA2” and selecting “Annotation” will redirect the user to the BRCA2 gene region on the genome map. Equally, if the user searches for “TNF” and selects the annotation option, they will be redirected to the TNF gene region. In addition, each region will have their respective comments displayed by users. This is achieved by utilizing the genome coordinates of the genes from the two genome browsers as they use identical genome assemblies.

# Discussion

## Introduction to the Discussion

The Discussion section of this dissertation holds crucial significance as it serves as the academic section for the comprehensive evaluation and contemplation of the web tool developed in the context of genomics. As part of the analysis, an in-depth investigation of the tool's design, functionality, and ramifications is necessary to heighten our understanding of its contribution to the field of genomics and computational biology.

The ensuing research encapsulates several key themes that deserve a thorough examination. Firstly, the usability limitations of the web tool's features will be examined, including the Genome Browser, Annotation Space, and User Authentication system. Identifying how these elements can provide researchers with the ability to facilitate their interaction with complex genomic datasets, aiding in annotation tasks, and fostering collaborative discussions is the first stage. This discussion offers critical insights into the practical utility of these features in enhancing the comprehension of complex biological processes such as the genome.

Furthermore, an analysis of the technical underpinnings of the web tool is required, scrutinizing the architecture and technology stack choices. It is important to consider how selecting specific programming languages, frameworks, and database design principles harmonize with the tool's performance and scalability requirements; through this, understanding the tool's structural foundations and how they align with the unique demands of genomics research will be significantly easier.

Lastly, extrapolating the broader implications of this web tool within the genomics community is required to understand the research's impact on future development and progress. This will allow an analysis of how the web tool contributes to the broader field, aiding genomics research, promoting accessibility, and facilitating collaboration between computer scientists and biologists. This discussion underscores the transformative potential of this innovation and concept.

In summary, the Discussion section assumes the role of an academic crucible wherein the overall research project is studied. It is a testament to the confluence of technological innovation and biological exploration.

## Web Tool Development Overview

The web tool's start is outlined as a comprehensive assessment of genomics researchers' challenges. This includes recognizing the burgeoning volume and complexity of genomic data while maintaining accessibility to users. Therefore, the development process prioritized the creation of an intuitive, user-centric interface. By adopting modern web technologies, the tool

integrates complex datasets from external databases, presenting them to researchers and other less experienced individuals in an accessible manner. This amalgamation of user-friendliness and data complexity renders the web tool unique.

An innovative feature of the web tool lies in its Genome Browser, which utilizes the power of the Ensembl REST API. This integration provides users with a simple and interactive platform for identifying basic information regarding genes of interest without including complex, unnecessary information unless further research and associated annotations are desired. This feature transcends existing visualization tools by enabling real-time navigation of genomic data and offering extensive visual customization options and analysis tools through the UCSC Genome Browser. Users can research the complexities of genomic structures and effectively identify biologically relevant insights.

This feature on the annotation space fosters collaboration and discussion by enabling users to comment on specific genomic regions, thereby enhancing the collective understanding of genomic data. It acts as a bridge between the complexity of genomic research and the collaborative nature of scientific exploration.

The development process also showcases a commitment to reliability. Strict testing and quality assurance procedures were implemented to promptly highlight and resolve integration issues. The choice of a modified Waterfall development methodology and project management tools like Asana ensured efficient progress monitoring and milestone attainment throughout the project.

## User Interface and Experience

- Design Principles: The UI's design principles revolved around clarity and functionality. A clean and uncluttered layout is displayed to users, making finding essential functions and tools easy. The color scheme, composed of plain tones, was chosen to minimize visual distractions while simultaneously providing contrast for crucial buttons. This keeps the focus on the genomic data and tools. Buttons were intuitively labelled, minimizing ambiguity and reducing the learning curve.

- User Experience (UX): The UX was developed to meet expected user anticipations. The Genome Browser, for instance, offers simple, readable information for users without unnecessary complex data, allowing users to identify their desired information effectively. The genome map allows the users to zoom, pan, and customize their view of genomic regions. This familiar interaction paradigm ensures that users can engage with genomic data comfortably.
- User Feedback and Usability Testing: Initially, the web tool's development process was expected to revolve around user feedback and usability testing. By releasing a prototype version to a select group of individuals with varying skill levels, valuable insights were expected to be gathered. Their input would refine the layout, feature placement, and functionality. Usability testing sessions would have focused on simple user tasks such as searching for specific genes, annotations, or commenting. Observations from these sessions would have been used to fine-tune the platform's user interactions. Unfortunately, this was not possible due to delays in ethical approval for participant research.
- Accessibility: Recognizing the diversity of users, including those with disabilities, the web tool adheres to web accessibility standards. This commitment extends to providing the entire web tool with simple, straightforward functionality, leaving no ambiguity and ensuring compatibility with screen readers. The aim is to make genomics research accessible to a broader audience.
- Responsiveness: The modern web is multi-device, and the web tool acknowledges this reality with a responsive design. Researchers can access and use the platform on various devices, including smartphones and tablets, without compromising

functionality. This adaptability caters to the flexible research environments of today's scientists.

- Future User-Centric Development: User feedback is integral to the web tool's ongoing and future potential development. The evolving needs and preferences of the user base would drive regular updates and new features where necessary, enhancing the web tool's capabilities. Utilizing this approach ensures that the web tool remains a valuable resource in the dynamic field of genomics research.

In conclusion, the web tool's user interface and experience were developed with the intention of maximizing accessibility and ease of use to larger audiences with a human-centred design approach. The emphasis on clarity, accessibility, and potential future user feedback has resulted in a platform empowering genomics users to delve into complex data confidently and efficiently. This commitment to user satisfaction and ongoing improvement underscores the tool's dedication to advancing genomics research.

## Functionality and Features

The web tool offers a comprehensive list of functionalities and features designed to address the diverse needs of genomics researchers and other users, ranging from data exploration to annotation and collaboration. These features were developed to overcome challenges specific to genomics research while enhancing usability.

**1. Genome Browser Integration:** At the core of the web tool lies its Genome Browser, allowing users to explore genomic data. This feature addresses the need for an intuitive platform to search for genomic regions, genes, and associated information. Researchers can obtain basic information efficiently on a specific gene while still being provided with the option for a further in-depth look into the gene.

**2. Annotation Space:** The Annotation Space feature enables users to add comments and annotations to specific genomic regions. In addition, it allows users to visualize genome regions based on their genomic coordinates using the genome map provided by the UCSC Genome Browser. It facilitates collaborative research by allowing scientists to share insights and hypotheses, fostering a sense of community. This addresses the need for a collaborative environment in genomics, where insights can be collectively discussed and refined on a central platform.

**3. Search Functionality:** The search functionality simplifies locating genes and genomic regions of interest. This feature tackles the challenge of navigating vast genomic datasets efficiently. Users can search by gene symbol, facilitating the identification of relevant genomic elements.

**4. User Authentication:** User accounts with authentication ensure data security and personalized experiences. This feature responds to the need for secure access control in genomics research. Researchers can securely log in, preserving the confidentiality of their work.

**5. Upload and Download Tools:** Researchers can upload custom genomic data files for analysis within the web tool given that they are of suitable format (iframe), simplifying data integration. Conversely, the download tool enables users to save annotated regions for future reference. These features address the need for flexibility in data management.

**6. Integration with External Genomic Resources:** The web tool integrates with external genomic resources and APIs, such as Ensembl and UCSC Genome Browser, expanding its functionality. Researchers can harness the abundance of data in these databases to enhance their analyses.

## Technical Challenges:

Implementing these features came with its share of technical challenges. Ensuring the integration of external genomic resources required careful API integration and data synchronization. Especially considering two different genome browsers were utilized. By identifying the genome assembly used by these genome browsers, the web tool takes the genomic coordinates as a reference point, linking the data between the search result from the search function to the genome map visualization. Furthermore, real-time annotation updates presented concurrency challenges that had to be addressed to maintain data integrity. AJAX and JavaScript functions were required to ensure a user-friendly commenting system.

In conclusion, the web tool's functionalities and features offer a variety of solutions for problems in genomics research. They not only address the identified problem of efficient data exploration and collaborative annotation but also tackle technical challenges associated with integrating complex genomic data sources. The platform's user-friendly interface, data security measures, and flexibility make it a valuable asset in the genomics research landscape, allowing users to gain as much insight into the genome's complexities as they wish.

# Technological Aspects

Careful choices in programming languages, frameworks, and technologies underpinned the development of the web tool. These choices significantly influenced the performance and scalability of the tool.

## Programming Languages and Frameworks:

The backend of the web tool was primarily built using Python, a versatile and widely adopted programming language. Flask, a lightweight web framework, was employed to facilitate rapid development. This combination efficiently handled HTTP requests, routing, and database interactions.

For the front end, HTML, CSS, and JavaScript were utilized. HTML provided the structure, CSS contributed to styling, and JavaScript enabled dynamic client-side interactions. Notably, the jQuery library was employed to simplify DOM (Document Object Model) manipulation and AJAX requests, enhancing the responsiveness of the web tool.

## Database Technology:

SQLite was chosen as the database system for its simplicity and suitability for small to medium-sized applications. It integrated with SQL Alchemy, an Object-Relational Mapping (ORM) tool. This combination facilitated the creation and management of database models in Python code, facilitating the handling of user data and comments.

## Impact on Performance and Scalability:

The choice of technologies played a pivotal role in shaping the performance and scalability of the web tool. Python's efficiency in handling requests allows for a responsive user experience and is expected to handle large volumes of user activity. Flask's lightweight nature also contributed to rapid development without introducing unnecessary overhead.

While suitable for smaller-scale applications, SQLite may need to be improved as the dataset grows significantly. However, SQLite provided adequate performance and data integrity for this project, which involved a limited user base.

## Scalability and Performance Challenges:

While the chosen technologies were well-suited for the initial scope of the web tool, it is essential to acknowledge potential scalability challenges. As the user base and dataset expand, transferring to a more powerful relational database system like MySQL may become necessary to maintain optimal performance and data management.

Additionally, while Python is known for its readability and ease of use, it can be less performant than lower-level languages like C++ for compute-intensive tasks. However, these concerns were mitigated in the web tool's context, as the primary focus was on data retrieval, display, and interaction with genomic information.

In conclusion, selecting programming languages, frameworks, and technologies was crucial to appropriate web tool development, given the required objectives to meet. These choices, namely Python, Flask, SQLite, and JavaScript, contributed to a responsive and user-friendly tool. However, as the web tool continues to evolve and potentially experiences increased user demand, scalability and performance optimization considerations will be necessary.

## Database Design and Management

The database design and management were fundamental to the functionality and data integrity of the web tool. This section evaluates the effectiveness of the database design, outlines data management challenges encountered, and offers insights into potential future challenges.

### Effectiveness of Database Design:

The database design was structured to record and retrieve user data, comments, and genomic information. SQLite and SQL Alchemy facilitated the creation of database models that accurately represented the application's data schema. The design encompassed two main tables, 'User' and 'Comment', linked by foreign key relationships.

The 'User' table stored user information, including usernames, emails, and hashed passwords. This design ensured data security and user authentication. Meanwhile, the 'Comment' table was responsible for storing comments posted by users, associating each comment with a specific genomic position.

The database design effectively supported the core functionality of the web tool, enabling users to register, log in, post comments, and associate comments with specific genomic regions.

## Data Management Challenges:

One primary data management challenge was ensuring data consistency and integrity, particularly concerning comments linked to genomic positions. This required implementing mechanisms to handle potential race conditions when multiple users simultaneously posted comments for the same genomic position. The web tool prevented conflicts through careful database transaction management and maintained data integrity.

## Future Potential Challenges:

While the current database design was sufficient for the project's initial scope, potential future challenges should be considered. As the user base grows and the volume of comments and genomic data increases, scalability becomes a concern. While suitable for smaller-scale applications, SQLite may face limitations in handling larger datasets.

To address future scalability challenges, the migration to a more powerful relational database system, such as PostgreSQL or MySQL, might be necessary. These database management systems offer enhanced performance and can better accommodate growing datasets and user traffic.

Moreover, data privacy and security will remain a major concern. In order to safeguard user information, compliance with data protection regulations will be crucial and can be achieved by implementing stern security measures, such as encryption and access controls.

In conclusion, the database design effectively supported the web tool's functionality. It enabled secure user registration, authentication, and comment storage. While data management challenges were addressed through careful transaction management, potential challenges related to scalability and data security must be proactively managed as the web tool evolves and its user base expands.

## Security and Privacy

Security and privacy are foremost concerns for any web application, especially when handling user data and genomic information. This section evaluates the security measures, data privacy considerations, and the approach to mitigate potential security vulnerabilities.

### Security Measures:

The web tool implemented several security measures to protect user data:

- **Password Hashing**: User passwords were securely hashed using the SHA-256 algorithm. Storing only the hashed passwords rather than plaintext ensured that sensitive user information remained protected even during a data breach.
- **Authentication**: The application employed the Flask-Login extension for user authentication. This provided an additional layer of security while simplifying the authentication process as necessary.
- **Input Validation**: User inputs were rigidly validated before interacting with the database to prevent SQL injection and other attacks.
- **Access Controls**: Access to certain functionalities, such as posting comments, was restricted to authenticated users. This prevented unauthorized access and actions within the web tool.

- **HTTPS Encryption**: Secure communication was ensured by using HTTPS to encrypt data transmitted between the user's browser and the web server. This safeguarded against eavesdropping and man-in-the-middle attacks.

## Data Privacy Considerations:

Data privacy was a central concern in the web tool's development:

- **User Consent**: In future development, users may be required to explicitly consent to the web tool's terms and privacy policy during registration.
- **Data Minimization**: Only essential user data, such as usernames and email addresses, were collected and stored. Unnecessary data collection was avoided to minimize privacy risks.
- **Genomic Data Handling**: Genomic data used for annotation purposes was fetched from trusted sources (Ensembl and UCSC Genome Browser) via APIs. This data was not stored within the application's database, reducing potential privacy implications.

## Mitigation of Security Vulnerabilities:

Despite implementing strong security measures, remaining vigilant about potential vulnerabilities is essential. Regular security checks and vulnerability assessments should be conducted. Some potential vulnerabilities and their mitigations include:

- **Cross-Site Scripting (XSS)**: Input validation and output encoding can be used to prevent XSS attacks. Regular security scans can help identify and address new vulnerabilities (KirstenS, 2020).
- **Cross-Site Request Forgery (CSRF)**: Flask-WTF's CSRF protection can defend against CSRF attacks by generating and validating unique tokens for each form submission (CSRF Protection — Flask-WTF Documentation (0.15.x), n.d.).

- **Session Management**: Session management and token-based authentication can be used to ensure that user sessions remain secure and unauthorized access is prevented.
- **Data Backup and Recovery**: Regular database backups can be used to maintain and mitigate data loss in case of unforeseen events or attacks. Additionally, disaster recovery plans should be developed as precautionary measures.

In conclusion, the web tool prioritized security and privacy throughout its development. Strong measures, such as password hashing, input validation, and HTTPS encryption, were implemented to protect user data. Data privacy considerations were also considered, with data minimization practices and user consent requirements. However, further security measures can be conducted through regular security screenings, and potential vulnerability assessments remain integral to ensuring the ongoing security of the web tool, especially as it continues to evolve and expand its user base.

## Usability

The usability of the web tool was a central consideration during its development. The aim was to create an intuitive and user-friendly interface despite needing more user testing and feedback. There was a firm reliance on industry-standard best practices in user interface design to achieve this.

- **User-Centred Design Principles**: Adherence to fundamental user-centred design principles focuses on creating a layout and navigation structure that logically organizes information and functions appropriately.
- **Responsive Design**: The design ensured the web tool was responsive, adapting to various screen sizes and devices to cater to a broader audience and providing responsive integration with the genome visualization map.

- **Accessibility**: The web tool was built to align with web content accessibility guidelines (WCAG) to ensure that it was as inclusive as possible, even though it could not directly be tested with users with disabilities.
- **Clear and Concise Interface**: The interface was designed to be straightforward, with a clean layout, well-labelled buttons, and minimal distractions.

## Scalability and Performance

Scalability and performance are aspects to be considered of any web tool, even those developed on a smaller scale. While the web tool was designed for a specific purpose and audience, scalability and performance were still considered in its development.

### Strategies Implemented for Scalability:

1. **Efficient Codebase:** Maintain a clean and efficient codebase, optimizing algorithms and minimizing resource-intensive operations. Focusing on these factors helped ensure the tool could handle increased user loads without significantly decreasing performance.
2. **Modular Design:** The web tool was designed with modularity in mind. Each component was developed as an independent module, making it easier to scale and enhance specific features without impacting the entire system.
3. **Downloading/Uploading Data:** This allows the web tool to provide the ability to save data without the utilization of cloud-based servers to store large amounts of data.

## Future Scalability Enhancements:

For future developments, several methods can be explored to further enhance the scalability of the web tool:

1. **Load Balancing:** Utilizing this method, incoming traffic can be distributed across several servers. This ensures that no single server becomes a bottleneck, thus improving response times during increased activity— a potentially helpful feature to consider as the user base expands.
2. **Database Optimization:** Continued optimization of the database design and queries can enhance the tool's ability to handle larger datasets efficiently. As mentioned previously, transitioning from SQLite to MySQL may be ideal.
3. **Scalable Infrastructure:** If necessary, consideration of cloud-based solutions or scalable infrastructure services can provide the flexibility needed to adapt to growing user bases.
4. **Performance Monitoring:** Analytical tools centred around performance monitoring can assist in identifying performance issues and areas requiring real-time optimization.
5. **Content Delivery Networks (CDNs):** Utilizing CDNs can help reduce the load on the primary server by providing static content from geographically distributed edge servers.
6. **Horizontal Scaling:** Exploring options for horizontal scaling, where additional server instances can be added to the system as demand grows, should be considered.

In conclusion, while the web tool was designed to meet its intended audience's requirements, proactively considering scalability strategies is helpful. As part of a future development process, there is potential for further scalability enhancements, including load balancing, database optimization, scalable infrastructure, performance monitoring, CDN utilization, and horizontal scaling. These strategies will ensure the web tool can adapt and perform optimally as its user base expands.

## Comparison with Existing Solutions

The web tool was developed specifically to provide an efficient and user-friendly solution for genome annotation and analysis. To evaluate its effectiveness and uniqueness, it is essential to compare it with existing solutions in the field.

### Advantages of The Web Tool:

1. **User-Friendly Interface:** The web tool offers an intuitive and user-friendly interface, making it accessible to a broad audience, including researchers and biologists with varying technical backgrounds. This ease of use is a significant advantage over many existing solutions that may require a steep learning curve.
2. **Integration of Multiple Tools:** Unlike existing tools that provide only one aspect of genome analysis, this web tool integrates multiple functionalities, including genome browsing, annotation, and commenting. This integration facilitates the research process and reduces the need for users to alternate between various applications.

3. **Local Hosting:** By allowing users to host the web tool locally, data privacy and security concerns are addressed. This local hosting option is not always available in existing solutions that rely on cloud-based platforms.

## Comparison to Existing Solutions:

1. **UCSC Genome Browser:** The UCSC Genome Browser is a widely recognized tool in genomics. While it offers powerful genome visualization features, it lacks the developed web tool's integrated annotation and commenting capabilities. Researchers often need to use additional software to annotate and collaborate on findings.
2. **Ensembl Genome Browser:** Ensembl is another popular genome browser with rich genome data resources. However, it primarily focuses on data presentation and needs the developed web tool's collaborative features and user-friendly interface.
3. **Apollo:** Apollo is a genome annotation curation tool that allows collaboration among researchers. While it excels in genome annotation, it may be more complex for non-specialists to use compared to the developed tool.

## Disadvantages of the Web Tool:

1. **Limited User Testing:** One of the tool's limitations is the need for more extensive user testing. Unlike some existing solutions that have undergone rigorous testing with a diverse user base, the tool's usability and user experience have yet to be as comprehensively evaluated.
2. **Scalability Challenges:** Although scalability considerations were part of the design, the tool may face challenges when handling large datasets or accommodating a high volume of concurrent users. Some existing solutions with powerful infrastructure may have an advantage in this area.

In summary, the web tool offers unique advantages in terms of user-friendliness, integration of multiple tools, and flexibility in hosting options. It stands out as a solution tailored to meet the needs of a wide range of users in genomics research. However, it is important to acknowledge the need for further user testing and potential scalability enhancements for future development. Compared to existing solutions like the UCSC Genome Browser, Ensembl Genome Browser, and Apollo, the tool provides a compelling alternative, particularly for those seeking an all-in-one genome annotation and analysis solution.

## Future Development and Enhancements

There are several potential opportunities for the future development and enhancement of the web tool. These possible improvements consider user feedback, technological advancements, and emerging needs in genomics research.

1. **Advanced Analysis Tools:** One of the foremost enhancements envisioned is incorporating advanced analysis tools. Currently, the tool provides genome browsing, annotation, and commenting features. However, there is a growing demand for in-depth analysis capabilities. The plan is to integrate tools that allow users to perform comparative analyses of genes, visualize gene expression data through heatmaps, and employ other sophisticated analytical methods. These additions will allow researchers to attain more in-depth insights from genomic data directly within the platform.
2. **Enhanced Collaboration Features:** Building on the existing commenting feature, bolstering collaboration capabilities would be ideal as it is a strong advantage of the web tool. This includes real-time collaboration on annotations, sharing of analysis

results, and collaborative data curation. Collaboration is a crucial pillar of scientific research, and the tool will continue to facilitate it effectively.

3. **Improved User Testing:** Recognizing the importance of user feedback, conducting comprehensive user testing to refine the user experience would provide a more refined tool. This testing will encompass researchers with diverse backgrounds and research goals. By identifying issues and preferences, the tool can be iteratively improved.
4. **Integration of Additional Genomic Databases:** Genomic data continually expands, and different databases offer unique resources. Integrating more genomic databases ensures users have access to the latest and most comprehensive data sources for their particular needs. This expansion will further enrich the research possibilities within the tool.
5. **Enhanced Scalability:** A strong focus on enhancing the tool's scalability to accommodate larger datasets and growing user bases is appropriate based on the complexity that the tool could potentially achieve. This may involve data storage optimizations, faster genome information retrieval, and improved handling of concurrent users. Ensuring the tool remains effective even as genomics datasets grow is a priority.
6. **User Customization:** Introduction of features that allow users to customize their workspace within the tool. Including personalizing the layout through icons, saving frequently used settings, and creating custom annotation tracks.
7. **Educational Resources:** Recognizing the tool's potential as an educational resource, developing guides, tutorials, and documentation to assist users in maximizing its utility may prove useful. These resources would cater to both novice and experienced researchers.

8. **Data Visualization Enhancements:** Investing in data visualization enhancements to aid researchers in interpreting genomic data effectively. This includes improved charting options, interactive data exploration, and data visualization in 3D space where applicable.
9. **Data Export and Integration APIs:** Introduction of data export capabilities and APIs that allow users to integrate the tool with other bioinformatics platforms and workflows, fostering interoperability, similar to how external APIs were integrated into the web tool.

In conclusion, the future development of the web tool is driven by a commitment to meeting the evolving needs of genomics researchers. By incorporating advanced analysis tools, enhancing collaboration features, improving scalability, and remaining responsive to user feedback, the aim is to ensure that the tool continues to be a valuable asset in the genomics research community. These developments will allow users to explore genomic data more comprehensively and efficiently, advancing the understanding of the intricacies of the genome.

## Conclusion of the Discussion

This comprehensive discussion explored the web tool's development, features, and future prospects tailored for genomics research. The web tool represents a potential milestone in genomics research, addressing the need for an integrated platform combining genome browsing, annotation, and user collaboration. While acknowledging that no user feedback or testing was conducted, the objective was to create a platform that offers valuable functionalities and the potential for future growth.

Throughout this discussion, several critical aspects have been highlighted:

- **Development Success:** The tool's development process displays a commitment to innovation and adaptability. By leveraging a combination of programming languages, frameworks, and technologies, the tool establishes a method to effectively incorporate the extensivity of genomics data with an intuitive user interface.
- **User-Centric Approach:** Although formal user testing was not conducted, the user-centric approach prioritized creating a tool that aligns with the needs and expectations of genomics researchers. Features such as genome browsing, annotation, and collaboration were designed to enhance research efficiency.
- **Security and Privacy:** Strict security measures have been implemented to safeguard user data, ensuring the privacy of sensitive genomic information. Although a formal security assessment was not conducted, the commitment to data protection remains.
- **Scalability and Performance:** The groundwork for accommodating larger datasets and concurrent users is established by understanding the importance of scalability in genomics research. Future plans would include further optimization in this regard.
- **Comparison with Existing Solutions:** In a comparative analysis, the advantages and disadvantages of the tool have been highlighted when juxtaposed with existing solutions. The flexibility and potential for future development sets the tool apart as a promising resource.
- **Future Prospects:** Integrating advanced analysis tools, enhanced collaboration features, and a mobile-friendly interface are relevant and user-friendly features that will enhance the effectiveness of the web tool for a wider

audience — a commitment to addressing user feedback, improving scalability, and expanding the database resources is essential for development.

In conclusion, the web tool represents a step forward in genomics research and web development incorporation within other fields. While recognizing the absence of formal user testing, the tool's potential impact is clear. With a strong foundation and a clear roadmap for the future, it is possible to assist researchers in exploring the complexities of the genomic landscape. This research can be marked not by an endpoint but by an ongoing commitment to advancing computational and genomic knowledge, resulting in encouraging discoveries.

# Conclusion

Throughout this dissertation, The primary aim has been to design and develop an intuitive web-based genome browser and annotation tool for visualizing and analyzing genetic data. Specific objectives, such as the pressing need for an accessible and efficient platform for genome information, annotation, and analysis, were addressed throughout the research.

**Summary of Key Findings:**

The development of the web tool has led to significant achievements. Through intensive work and commitment, a user-friendly web tool was developed that integrates the UCSC Genome Browser via iframes and integrates the Ensembl Genome Browser utilizing REST APIs. It provides researchers with an innovative means of exploring genomic data while maintaining simplicity where necessary. Through a well-thought-out technological framework, a successful web tool was developed that highlights combining external databases into one platform, providing users with multiple features for their desired purposes.

## Discussion of Research Objectives:

Reflecting on the research objectives, it is clear that substantial progress was made. However, as with any complex project, several deviations were encountered, and as a result, the project had to be adapted due to unforeseen challenges. The absence of direct user testing is a notable deviation, emphasizing the importance of comprehensive feedback in future endeavors significantly more.

## Contributions to the Field:

The contributions to the field are characterized by innovation and practicality. By creating an association between genomics and web development, a novel tool encompassing a multitude of features from various platforms into one tool can be considered for future development. This has the potential to catalyze genomics research and make it more accessible to a broader audience.

## Practical Implications:

The practical implications of the web tool extend beyond academia. It can be utilized as a research tool for the general public in discussing genetic conditions or diseases. With continuous development, it also has the potential to revolutionize genomics research, accelerating discoveries and applications in fields as diverse as medicine, agriculture, and conservation.

## Limitations:

The research limitations are clear, particularly the absence of direct user testing, which may have significantly impacted the tool's usability. This limitation may have severely influenced the design and developmental process of the tool due to a lack of user feedback to enhance accessibility and integrate particular desired features; this acts as a reminder of the importance of the iterative nature of software development.

## Future Research Directions:

This project provides the ideal stepping stone for continuous development as there are multiple opportunities to expand the tool to include additional features, such as gene comparisons through heatmaps, customizability amongst users, integration of other tools from other APIs, and more. Moreover, the tool has a stable structure for continued refinement and scalability enhancements to improve its size.

In conclusion, this dissertation represents a significant step in genomics research and web development with continued research. It emphasizes the importance of technological innovation in advancing science. While challenges were encountered, the process has been gratifying, and the potential impact of the web tool is both exciting and far-reaching.

# Personal Reflection:

On a personal note, this dissertation project has been a transformative experience. At the beginning of my pursuit of an MSc in Computing, I was keen on applying my scientific biological knowledge to my newly obtained knowledge in computing. Through this research project, I have achieved my goal and have deepened my appreciation for the convergence of science and technology, underlining the importance of adaptability and continuous learning.

As part of this project, I significantly enriched my web development skills through research on API integration, the use of AJAX, and iframes implemented into the web tool. Additionally, research into other web frameworks and programming languages highlighted the advantages and disadvantages of each based on requirements for the particular project.

On the other hand, I also encountered significant challenges when attempting to implement technical ability through the use of JavaScript to execute certain functions. Moreover, I underestimated the work required to achieve the expected product. As a result, my original aim for the web tool was significantly more advanced than the current tool. Despite this, I am content and satisfied with the progress and features implemented into the current web tool.

The result displays my project management weaknesses and inability to identify appropriate work amounts given a set deadline. However, it also portrays the advancement in my web development and problem-solving skills, as I was able to solve most technical issues independently.

I am enthusiastic about the impact this web tool can offer the future of genomics research. The potential for discovery and positive impact is immeasurable, and I anticipate the achievements that will be developed in genomics will be driven by innovative tools like this.

# References


Adam. (2023, March 19). Python Frameworks. W3schools | JavaTpoint | Tutorialspoint.

Ambuj Bhushan Jha, Gali KK, Alam Z, Reddy B, Warkentin TD. Potential Application of Genomic Technologies in Breeding for Fungal and Oomycete Disease Resistance in Pea. 2021 Jun 21;11(6):1260–0

Bauch, A., Adamczyk, I., Buczek, P., Elmer, F.-J., Kaloyan Enimanev, Pawel Glyzewski, Kohler, M., Tomasz Pylak, Quandt, A., Ramakrishnan, C., Beisel, C., Lars Malmström, Aebersold, R., & Rinn, B. (2011). openBIS: a flexible framework for managing and analyzing complex data in biology research. BMC Bioinformatics, 12(1). Available from: https://doi.org/10.1186/1471-2105-12-468

Beard L, Schein R, Morra D, Wilson K, Keelan J. The challenges in making electronic health records accessible to patients: Table 1. Journal of the American Medical Informatics Association [Internet]. 2012 Jan;19(1):116–20. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3240757/

Bell G, Hey T, Szalay A. COMPUTER SCIENCE: Beyond the Data Deluge. Science [Internet]. 2009 Mar 6;323(5919):1297–8. Available from: https://science.sciencemag.org/content/323/5919/1297/tab-article-info

Birney E. An Overview of Ensembl. Genome Research. 2004 May 1;14(5):925–8.

Brodtkorb AR, Hagen TR, Schulz C, Hasle G. GPU computing in discrete optimization. Part I: Introduction to the GPU. EURO Journal on Transportation and Logistics. 2013 May;2(1-2):129–57.

Carrere S, Gouzy J. myGenomeBrowser: building and sharing your own genome browser. Bioinformatics. 2016 Dec 23;btw800.

Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011 Jan 26;12(1).

CSRF Protection — Flask-WTF Documentation (0.15.x). (n.d.). Flask-Wtf.readthedocs.io. Retrieved May 16, 2023. Available from: https://flask-wtf.readthedocs.io/en/0.15.x/csrf/

Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, et al. Visualization of omics data for systems biology. Nature Methods. 2010 Mar;7(S3):S56–68.

Fernandez-Cuesta, L., & Peifer, M. (2017). Nanoscale proteomics. Trends in Genetics, 33(12), 904-905.

Huang S . C, Fraenkel E. Integrating Proteomic, Transcriptional, and Interactome Data Reveals Hidden Components of Signaling and Regulatory Networks. Science Signaling. 2009 Jul 28;2(81):ra40–0.

Karolchik D. The UCSC Genome Browser Database. Nucleic Acids Research. 2003 Jan 1;31(1):51–4.

Kaur, R., & Sharma, B. (2018, August 1). Comparative Study for Evaluating the Usability of Web Based Applications. IEEE Xplore. Available from: https://doi.org/10.1109/ICCS.2018.00023

Kim, M. J., Wang, X., Han, S., & Wang, Y. (2015). Implementing an augmented reality-enabled wayfinding system through studying user experience and requirements in complex environments. Visualization in Engineering, 3(1). Available from: https://doi.org/10.1186/s40327-015-0026-2

KirstenS. (2020). Cross Site Scripting (XSS) | OWASP. Owasp.org; Owasp. Available from: https://owasp.org/www-community/attacks/xss/

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome research [Internet]. 2009 [cited 2019 Jan 27];19(9):1639–45. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19541911

Małek P. Everything You Need to Know about Ruby on Rails Web Application Framework [Internet]. Blog by Railsware. 2020. Available from: https://railsware.com/blog/ruby-on-rails-guide/

Newman V, Moore B, Sparrow H, Perry E. The Ensembl Genome Browser: Strategies for Accessing Eukaryotic Genome Data. Methods in Molecular Biology. 2018;115–39.

Node.js. Introduction to Node.js [Internet]. Introduction to Node.js. 2023. Available from: https://nodejs.dev/en/learn/

Nusrat S, Harbig T, Gehlenborg N. Tasks, Techniques, and Tools for Genomic Data Visualization. Computer Graphics Forum. 2019 Jun;38(3):781–805.

Olsen LR, Campos B, Barnkob MS, Winther O, Brusic V, Andersen MH. Bioinformatics for cancer immunotherapy target discovery. Cancer Immunology, Immunotherapy. 2014 Oct 26;63(12):1235–49.

Segelbacher, G., Bosse, M., Burger, P., Galbusera, P., Godoy, J. A., Helsen, P., Hvilsom, C., Iacolina, L., Kahric, A., Manfrin, C., Nonic, M., Thizy, D., Tsvetkov, I., Veličković, N., Vilà, C., Wisely, S. M., & Buzan, E. (2021). New developments in the field of genomic technologies and their relevance to conservation management. *Conservation Genetics*. https://doi.org/10.1007/s10592-021-01415-5

Raman, R. (2011). Evidence-based Medicine and Patient-centered Care: Cross-Disciplinary Challenges and Healthcare Information Technology-enabled Solutions. The International Journal of Person Centered Medicine, 1(2).

Rego S, Grove ME, Cho MK, Ormond KE. Informed Consent in the Genomics Era. Cold Spring Harbor Perspectives in Medicine. 2019 Sep 30;10(8):a036582.

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. JBrowse: A next-generation genome browser. Genome Research. 2009 Jul 1;19(9):1630–8.

Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics. 2012 Apr 19;14(2):178–92.

Wang J, Kong L, Gao G, Luo J. A brief introduction to web-based genome browsers. Briefings in Bioinformatics [Internet]. 2013 Mar 1;14(2):131–43. Available from: https://academic.oup.com/bib/article/14/2/131/208726

Yang, H., Han, H.-J., Heo, J.-M., Jeong, J., Lee, T., Hu, W., & Kwak, S. (2018). Ocean Color Algorithm Development Environment for High-Speed Data Processing of GOCI-II. Available from: https://doi.org/10.1109/igarss.2018.8517983