جامعة الملك سعود
King Saud University
كلية علوم الحاسب والمعلومات
قسم تقنية المعلومات

GRN VTOOL

IT 497: Graduation Project Report
Product Release-2

Prepared by
<Rawan Alhabeeb, 441203926>
<Nora Alanzi, 441201069>
<Ream AlQuraishi, 441201233>
<Norah Albarrak, 439201100>

Supervised by
< Dr.Nuha Saud Bin Tayyash >

First Semester 1445
Fall 2023

Table of Contents

GRN VTOOL

جامعة الملك سعود
King Saud University

كلية علوم الحاسب والمعلومات
قسم تقنية المعلومات

Nora Alanzi [1], Rawan Alhabeeb [2], Norah Albarrak [3] and Ream AlQuraishi [4].

[1]Information Technology Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia; 441201069@student.ksi.edu.sa
[2]Information Technology Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia; 441203926@student.ksi.edu.sa
[3]Information Technology Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia; 439201100@student.ksi.edu.sa
[4]Information Technology Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia; 441201233@student.ksi.edu.sa

The field of Gene Regulatory Network (GRN) construction plays a pivotal role in unraveling cellular processes and gene pathways. However, conducting GRN analytical studies has posed significant challenges for biologists. Technical expertise is often required for tool installation and configuration, while the reliance on programming command lines hampers flexibility in adjusting tool parameters. Moreover, the absence of a standardized format for visualizing GRNs across different tools complicates comparative analysis. GRN VTOOLS is a web-based platform that offers pre-installed GRN visualization tools, including DIANE and SeqNet, which enable researchers to overcome technical hurdles. We chose to implement DIANE and SeqNet as they use GENIE3 and the Gaussian graphical model (GGM) for GRN inference. GENIE3 is a random forest ensemble machine learning procedure that was among the best performers of the DREAM challenges. GRN VTOOLS presents results from the inference methods in an interactive dashboard with a unified interface. A notable feature of GRN VTOOL is its capability to facilitate the comparison of two GRNs in one session. These GRNs could result from different inference methods or resulted from using one inference tool on different datasets. This is helpful for biologists, particularly in patient scenarios, where they can conveniently analyze GRNs before and after treatments or interventions thus eliminating the need for manual comparison. GRN VTOOL empowers biologists to efficiently conduct comparative analyses, gain deeper insights into regulatory dynamics, and accelerate their research, particularly in patient-focused studies.

:Abstract (Arabic)

يلعب مجال شبكات التحكم الجيني (Gene Regulatory Network)، دورا حيويا في فك شفرة العمليات الخلوية ومسار الجينات. ويشكل إجراء الدراسات التحليلية على شبكات التحكم الجيني تحديا كبيرا للأحيائيين تحديا كبيرا لما يتطلبه من خبرة تقنية لتثبيت الأدوات وضبطها. حيث يعتمد كثير منها على الأوامر البرمجية في تعديل متغيرات الأدوات. علاوة على ذلك لا يوجد نموذج موحد لعرض شبكات التحكم الجيني الناتجة من الأدوات المختلفة مما يعصب المقارنة التحليلية بين النتائج المختلفة. GRN VTOOL هي منصة عبر الويب تقدم أدوات عرض شبكات التحكم الجيني (DIANE وSeqNet) مسبقة التنصيب والتي تتيح للباحثين التغلب على الصعوبات الناتجة عن تنصيب الأدوات. اخترنا أداتي DIANE وSeqNet لاحتوائهما على أداة GENIE3 بالإضافة إلى أدوات أخرى لاستنباط شبكات التحكم الجيني. GENIE3 هي أجراء تعلم آلة باستخدام طاقم الأشجار العشوائية لاستنباط شبكات التحكم الجيني الفائزة بتحدي DREAM. تعرض GRN VTOOL نتائج أدوات عرض شبكات التحكم الجيني لوحة تحكم تفاعلية بواجهة موحدة. الميزة بارزة لـ GRN VTOOL هي قدرتها على تيسير المقارنة بين شبكتي تحكم الجيني في جلسة واحدة. هذه الشبكات إما أن تكون ناتجة عن طرق استنباط مختلفة او عن طريق استخدام أداة واحده على مجموعتين من مختلفتين من البيانات. هذه الميزة تساعد الأحيائيين على تحليل الشبكات بسهولة خاصة في حالة المرضى ومقارنة الشبكات قبل وبعد العلاج او التدخل مما يقلل الحاجة المقارنة اليدوية. تمكن GRN VTOOL الأحيائيين من إجراء تحليلات مقارنة بكفاءة، والحصول على رؤى أعمق حول التنظيمات الحيوية للجينات وتسريع بحوثهم، خاصة في الدراسات المركزة على المرضى.

Keywords: GRN, visualization, dashboard, interactive, and inference tool.

1 Introduction

In bioinformatics applications, Gene Regulatory Network (GRN) shows biological interactions between genes and provides a systemic comprehension of cellular signaling and regulatory processes [1]. Many computation methods have been developed for inferring network structure ranging from probabilistic network-based approaches, correlation-based approaches; partial correlation-based approaches; and information theory-based approaches[2],[3]. Comparative evaluation among different methods for constructing large scale GRNs revealed the strengths and weaknesses of each method concerning different scenarios, with no single method outperforming others universally.

For biologists conducting a successful analytical GRN study is erratic, where the problems range from the required technical background for installation and configuration to the stand-alone results and manual comparisons. Using pre-installed, up to date tools, where the biologist runs a comprehensive analysis and compares the results of different tools in one-place could be a notable aid for biologist that help them to conduct more efficient studies.

1.1 The Problem

Numerous tools have been developed for Gene Regulatory Networks construction as they represent a key factor to reveal the cellular processes and gene pathways [4]. However, conducting a GRN analytical study is challenging for biologists for many reasons. It starts with the technical knowledge required for tools installation and configuration. Programming command lines are usually the methods used to run tools functions which make changing functions parameters difficult.

Usually, each GRN tool uses a single inference method and as we mention there is no single method that outperforms all methods. Thus, to compare the resulted GRN of different tools the biologist needs to run each tool independently and compare these results manually. In addition, the tools visualize GRN graphs in different formats, which made a comparative analysis not a standardized process as figures 1.1 and 1.2 show. In figure 1.1, we can see a GRN visualization produced by DIANE [5] tool where squares are two types of gene. The light green are regulators, the dark green is grouped regulators, circles are normal genes, where arrows with different colors have different meaning. Figure 1.2, show a GRN visualization resulted from SeqNet tool [6] where the circles are of different sizes depending on the number of the connected links and there are links between genes not arrows. So, it is hard to understand which gene initiates the biological process and impacts the other genes.

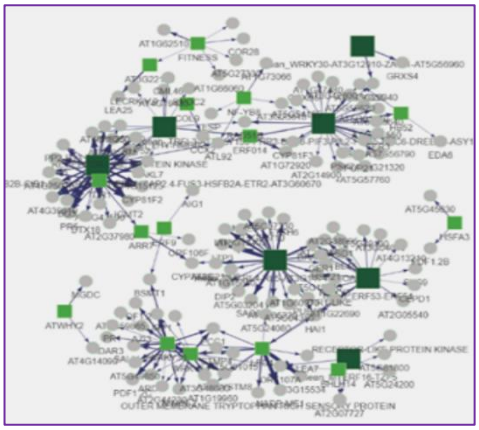


Figure 1.1 GRN generated by DIANE

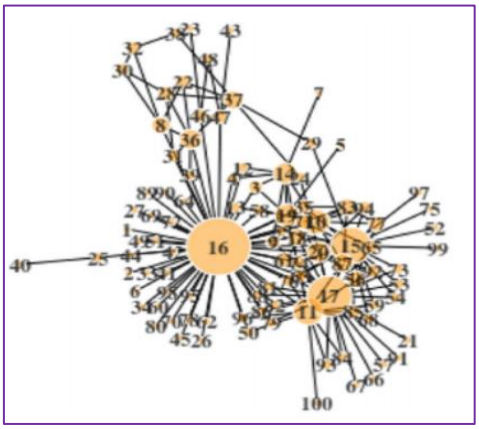


Figure 1.2 GRN generated by SeqNet

In addition, the results are usually images of large number of connected genes, so it is hard to spot the similarity and differences between different GRN visualizations. For example, figures 1.3 and 1.4 are different. However, it is hard to notice the difference because of that, SeqNet developers implement a method to highlight the difference between the two networks as figure 1.5 shows.

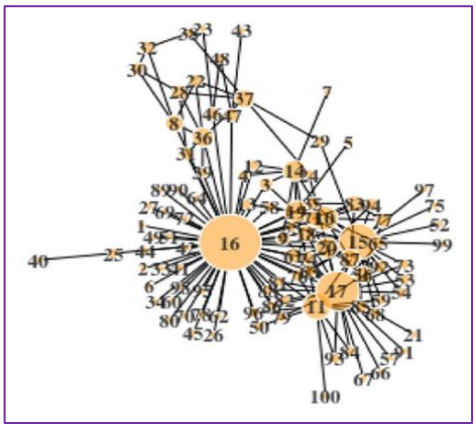


Figure 1.3 generating network_1 using SeqNet

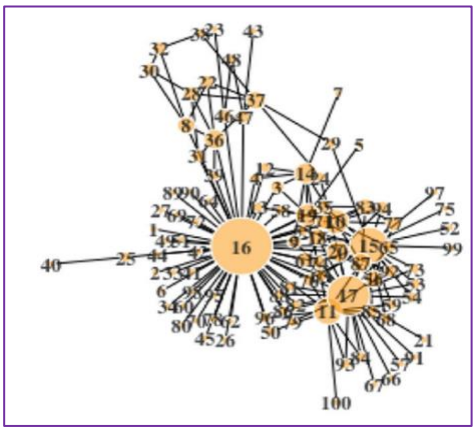


Figure 1.4 generating network_2 using SeqNet

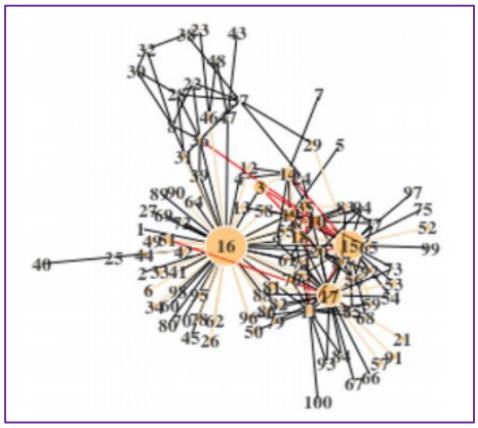


Figure 1.5 Highlights the differences using gen_partial_correlations () function.

In GRN analytical studies, biologists usually focus on hub genes or partial networks. However, extracting a partial network to do in-depth analysis is not direct and the current tools usually allow the user to download the whole network resulted from the inference method not a selected part. Moreover, biologists usually use these tools to compare the GRN resulted on two different datasets, e.g., compare the GRN for the patient and non-patient samples, and to do that biologists need to run the GRN inference tool on the first dataset and download the result then,

start a new session to get the result on the second dataset to compare the two results independently.

In addition to the challenge of comparing gene regulatory networks from different datasets, biologists often encounter difficulties when studying patient data. In the case of patient scenarios, biologists need to analyze gene regulatory networks before and after a treatment or intervention. However, they face a significant hurdle in the absence of a centralized platform that allows them to conveniently view and compare both network representations side by side. The lack of such a solution hampers their ability to comprehensively assess the impact of the treatment on the regulatory dynamics and hinders their ability to identify crucial regulatory changes associated with the patient's condition.

○ The Solution:

GRN VTOOLS is a website that allows biologists to run many pre-installed GRN tools and save them from technical difficulties. It presents the results of different inference methods in one place which help the biologists to conduct efficient comparative analysis. GRN VTOOLS is an interactive network where the user can click on gene to get extra information and that will help the biologist to get further insight into the results. In addition, GRN VTOOLS has a search feature that allows the user to navigate in large networks to help them to focus their analysis. Furthermore, GRN VTOOLS allows biologists to run one tool on two datasets simultaneously and visualize the results in one place, accelerating the comparative analysis. GRN VTOOLS offers two states of the art tools, DIANE and SeqNet [6] [5].

The main feature for GRN VTOOLS are:

- Help the user to run GRNs without worrying about requiring technical knowledge for installation.
- Compare the results of different GRNs in one session.
- Interactive visualization for better results interpretation.
- Compare the results of the GRN tool on two datasets in one session.
- Search for a particular gene in the network.
- Download the results.

GRN VTOOLS is an interactive website to help biologists to view Genes network better and find connections and comparisons between gene networks using different tools. It will be a website only and uses English as its main language, GRN VTOOLS won't be available as a mobile app and won't cover other languages, it will cover functions such as searching, zooming, marking, and comparing the genes, GRN VTOOLS will only cover 2 tools to be used, GRN VTOOLS will not cover the inference of the genes network, nor send the results to the user as the user is only able to view it on the website using their personal account.

○ Product Vision

For biologists,
Who needs to analyze gene regulation networks.
The GRN VTOOLS is a website that helps biologists to provide various types of tools to analyze more than one GRN at a time.
Unlike other tools and websites that only provide one tool and require installation and perform single analysis at a time.
Our product provides GRNs simultaneous visualization with interactive features.

○ Project Approach

To achieve the necessary learning and understanding of the GRN VTOOLS platform, extensive research and exploration have been undertaken. This involved gathering information from a variety of reliable sources, including scientific literature, technical documentation, online resources, and user feedback. The process began by studying the functionality and features of the platform's dashboard, which serves as the central interface for users to access and analyze gene regulatory networks.

To gain a deeper understanding of the analysis tools provided by GRN VTOOLS, namely DIANE and SeqNet, an in-depth review of their methodologies, algorithms, and applications was conducted. This involved studying research papers, documentation, and tutorials related to these tools. By analyzing their strengths, limitations, and use cases, we gained insights into how they contribute to gene regulatory network inference and analysis.

Furthermore, the database utilized within GRN VTOOLS was thoroughly explored. This involved investigating the data collection methods, data preprocessing steps, and the scope of gene expression profiles available. Understanding the characteristics of the database was crucial for comprehending the type and quality of input data used in network analysis.

Additionally, research was conducted on network analysis concepts, including network visualization, network inference algorithms, and the interpretation of network structures and dynamics. By studying the underlying principles and techniques, we gained a deeper understanding of how gene regulatory networks are generated and analyzed within the GRN VTOOLS platform.

○ Project contribution:

GRN VTOOLS addresses the challenges faced by biologists, including the patient scenario, by providing a comprehensive solution for gene regulatory network analysis. In the case of patient data, the platform allows biologists to conveniently compare gene regulatory networks before and after a treatment or intervention. By running the SeqNet or DIANE tools simultaneously

on two datasets, biologists can visualize and analyze the results in a single interface, eliminating the need for manual comparison. The platform's interactive network visualization enables biologists to click on genes and access additional information, facilitating deeper insights and aiding in the interpretation of results. Additionally, the search feature assists biologists in navigating through large networks and extracting relevant subsets for focused analysis. With GRN VTOOLS, biologists can efficiently conduct comparative analysis, gain valuable insights into regulatory dynamics, and accelerate their research in patient scenarios

1.1.1 Hardware/Software Tools and Cost

Hardware Tools	
Personal computers	available
Software Tools	
Ubantu AWS virtual server	$0.093/hr
MySQL RDS database	Free
GitHub	Free
Dropbox	Free
Zotero	Free
R	Free
DIANE	Free
SeqNet	Free

In summary, the steps taken to achieve the necessary learning involved extensive research on the dashboard, DIANE and SeqNet tools, the database, server, and network analysis concepts. By leveraging information from diverse and reliable sources, we acquired knowledge about the functionalities, methodologies, and applications of these components within the GRN VTOOLS platform. However, it is important to verify the information with the latest sources and consult experts for specialized or cutting-edge knowledge.

In this document, we begin with an introduction to the challenges faced by biologists in analyzing Gene Regulatory Networks (GRN), followed by our proposed solution and the development of our product, GRN VTOOL. We delve into a detailed overview of Gene Regulatory Networks in the 'Background' section and explore their significance in biology.

Our 'Literature Review' then examines current technical advancements and their limitations in facilitating efficient GRN analysis. The core of our work is detailed in the 'System Design and Development' chapter, where we discuss our methodology, system requirements, including user interactions and the development roadmap, followed by a comprehensive system design with architectural and class diagrams, and component-level design considerations. The 'Data Design' section describes our data models and data preparation strategies, along with 'Interface

Design'. The 'Implementation' section presents the practical application of our designs. Our 'System Evaluation' includes experimental results, user acceptance testing with participant demographics and feedback, along with an assessment of quality attributes. The 'Discussion' section critically analyzes the findings.

We conclude with insights into our future work in the 'Conclusions and Future Work' section, express our gratitude in 'Acknowledgements', and provide our references and supplementary materials in the 'References' and 'Appendix' sections respectively.

2 Background

We provide in this part of the document a general discussion and perspective of gene regulatory networks. Cells are intricate systems consisting of several interconnected parts with various roles and properties, including the capacity to interact with one another through a range of chemical as well as mechanical signals [9]. Long DNA molecules are found in almost every cell and are crucial for the development and upkeep of the organism. The fundamental structural and operational component of a cell is called a gene, which is a segment of the DNA. Only a portion of a cell's genes are expressed or turned on. Figure 2.1 [10], depicts how a cell differentiates into a certain type as a result of the phenomena known as gene regulation [11].

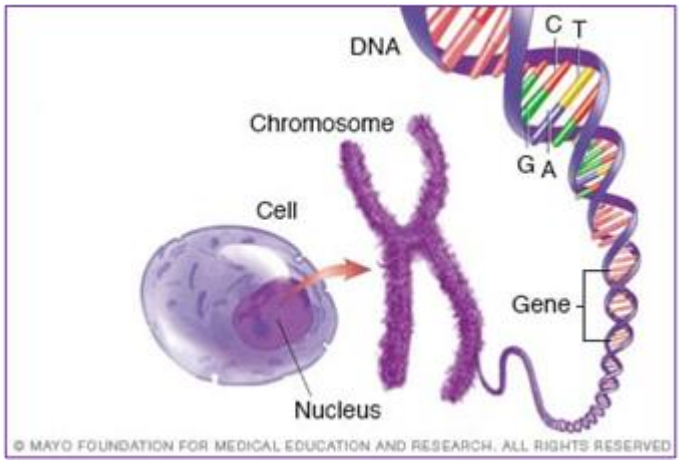


Figure 2.1 Long DNA

A Gene Regulatory Network (GRN) is a network inferred from gene expression data, offering insights into regulatory interactions among regulators and their potential targets, including gene-gene and protein-protein interactions [12]. Such networks play a crucial role in addressing various biological and biomedical challenges [13]. As the DNA sequence of the genome significantly shapes the architecture of a GRN, and DNA modifications enable direct testing of GRN concepts, these networks are inherently genome-oriented. It is essential to focus on the expected DNA inputs, serving as the foundational elements for the model.

In the realm of cancer research and therapeutic decision-making, molecular tumor characteristics play an increasingly pivotal role. Individualized gene regulatory networks are inferred for patients, contributing to personalized treatment strategies. This capability has been facilitated by advancements in single-cell RNA sequencing methods, allowing the generation of thousands of transcriptomic samples from a single patient [14]. Consequently, these approaches enhance our understanding of major oncogenic mechanisms within a patient's cancer, influencing therapeutic decisions in the ongoing battle against cancer.

Each regulatory module within the genome processes a diverse array of inputs, which can be represented mathematically as combinations of logic functions (e.g., "and," "switch," and "or" functions) [15]. A gene regulatory network (GRN) essentially functions as a network of analog computational devices, and its operations are contingent upon the inputs it receives. At the system level, it is comprised of assemblies of these information-processing units.

As outlined in [16], GRNs often involve complex interactions among dozens of genes, influencing each other through various mechanisms. Understanding the intricate mechanisms of these networks poses significant challenges. Referring to reviews in [17] and[16], uncovering the individual aspects of regulatory interactions necessitates employing multiple experimental approaches. This includes the analysis of spatial and temporal gene expression data, requiring assessments at various time points and developmental phases.

Identification of functional interactions among genes is crucial, involving two approaches: trans-perturbation, which perturbs the transcription factor, and cis-perturbation, which mutates the binding sites for specific genes. Additionally, discerning physical interactions between transcription factors and binding sites is an essential aspect of comprehending gene regulatory networks [16].

Gene regulatory networks are heterogeneous assemblies of various subcircuit types, each of which serves a particular purpose[15].This idea is crucial because it unlocks the fundamentals of network architecture. Similar to how specific electronic subcircuit devices are used in many types of processors, certain biological subcircuits, such as the majority of signal transduction subcircuits, are employed in a wide variety of biological situations. Others are more sophisticated and devoted to the same biological processes wherever they are found. These latter types of subcircuits are "wired" in a way that makes it difficult to reorder them. Some kinds of network linkages, on the other hand, are far more adaptable and changeable, according to evolutionary comparison. Morphological variety across related animal species can be created by even minor changes in these latter links [15].

It will be possible to statistically compare these networks as more gene regulatory networks from other physiological and pathological states become accessible[18]. This will help us understand how interactions between different physiological or disease situations might alter, and it will also improve our understanding of these phenotypes from a biological and biomedical perspective [19] [20]. Topological indices could be used in this analysis in addition to more traditional comparative measures like the graph edit distance and the Zelinka distance. It will be important to create databases that offer open access to the inferred gene regulatory networks from various physiological and pathological conditions in order for this strategy to be successful, comparable to sequence or protein structure databases. Due to the estimated effort required to develop such a database and interactive access interfaces being more than

that of sequence databases, it could be essential to create an international coalition in order to achieve this.

Before generating the gene regulatory network, data preprocessing procedures are performed to ensure the quality and reliability of the data. Data normalization is a crucial step, as it removes technical biases and variations that can arise during the RNA-seq experiments. Normalization methods such as TPM (Transcripts Per Million) or FPKM (Fragments Per Kilobase Million) are commonly used to account for differences in sequencing depth and gene length [35].

In this project we chose to use DIANE [refrence to DIANE] as the first tool to visualize GRN from gene expression data. DIANE (Dashboard for the Inference and Analysis of Networks from Expression data) is an R package, that allows the user to observe the normalized expression levels of several genes of interest. DIANE enables gene expression, profiles clustering using the statistical framework for inferring mixture models through an Expectation-Maximization (EM) algorithm [21], [22]. In DIANE, the package chosen for GRN reconstruction is GENIE3, a machine learning procedure that was among the best performers of the DREAM challenges. GENIE3 uses Random Forests[23], which is a machine learning method based on the inference of a collection of regression trees. It has the advantage of being a non-parametric procedure, requiring very few modelling or biological priors while being able to capture interactions and high-order combinatorics between regulators. Figure 2.2 shows DIANE workflow with the analytical results.

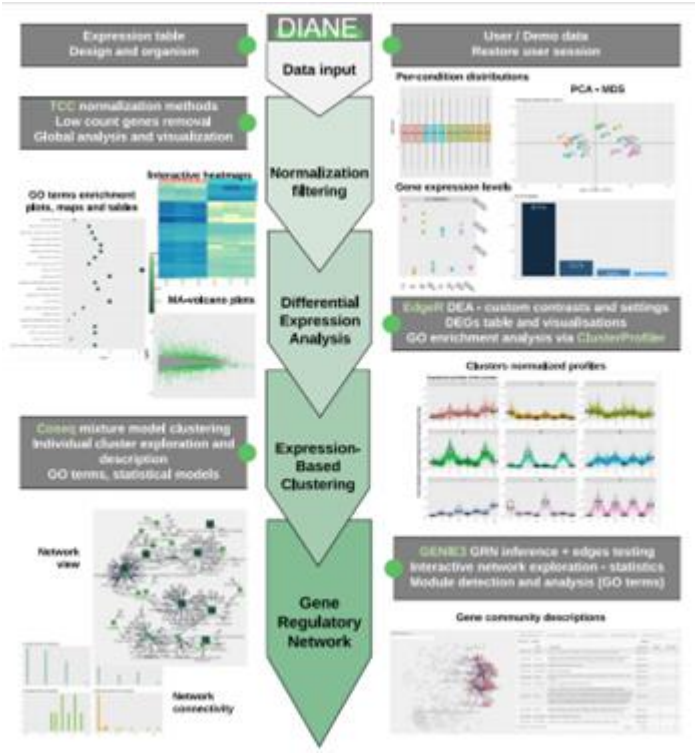


Figure 2.2, DIANE's workflow. The main steps of the pipeline available in the application -data import, normalization, exploration, differential expression analyses, clustering, and network inference- alongside with some chosen visual outputs

SeqNet [6] is the second method we chose for GRN visualization. It provides users with tools for (1) creating random networks that are similar to real transcription networks and (2) generating RNA-seq data. SeqNet has three main components: the network generator, the Gaussian graphical model (GGM), and a converter from GGM values to RNA-seq expression data. Interested readers could refer to SeqNet [6] for model details about the package methodology. Table 2.1 shows the main functions for SeqNet package.

Table 2.1 Summary of the main SeqNet functions

Function	Description
Random_network()	Generate a random network containing p nodes (genes). Additional tuning parameters for the network generating algorithm can be used here, including nu, prob rewire, prob remove, and alpha.
As_single_module()	Used to coerce a network into a single, large module rather than a collection of overlapping modules. This affects how data are generated when passed into gen raseq0,The decision to use this function should be based on the user's model of the data generating process.
Perturb_network()	Rewires a network to obtain a differential network. By default, a hub node is randomly sampled from the network, and each of its connections are removed with probability 0.5. Note, this function can be used iteratively to turn off many hub nodes.
Gen_partial_correlations()	Generates weights for the connections in a network. When multiple network objects are provided, connection that are common across networks are given the same weight.
Print()	Printing a network displays the number of nodes, edges, and modules, along with the

	average degree, clustering coefficient, and average path length in the network.
Plot()	Generates a plot of the network. See also plot network.
Plot_network_diff()	Generates a plot of the differential network between two network objects.
Gen_rnaseq()	Generate an RNA-seq dataset containing n samples (rows) with a dependence structure defined by a given network. A reference dataset can be provided,
sample_reference_data()	Helper function for sampling p genes (columns) from a reference dataset. The argument percent ZI is useful when the reference contains zero-inflated genes to control the proportion of the p sampled genes that are zero-inflated.
get_adj acency_matrix()	Creates an adiacency matrix from a network. If genes i and j are connected in the network (in any module), then the ith entry of the adjacency matrix is 1, and otherwise is 0.
Create_network_from_adjacency_matrix	Creates a network from an adjacency matrix. This is useful when the user has a prespecified network structure, i.e. doesn't want to generate a random one, but does want to generate RNA-seq data.

3 Literature Review

In the next subsections, we started by showing the related work in Gene Network visualization, then we presented the automated solutions for GRN.

3.1 Gene Regulatory Network Visualization

This section presents some of the previous work on gene network visualization. In a survey[11] made it clear that discovering how a gene influences other genes usually requires performing a large number of lab experiments in which cells are placed in different environments or in which their genes are artificially turned on or off to observe how these changes the activation of other genes (i.e. the synthesis of the corresponding RNA and proteins).

In[36], during development, the nuclei of specific spatial components of the organism express various regulatory gene sets, which ultimately result in structural gene sets. As a result, we must model, visualize, and comprehend GRN activity within a single cell and at a certain developmental stage in the context of the cell's past states as well as those of its parents, neighbors, and other cells. Two complementary viewpoints can be used to compare the genetic status of several groupings of cells across time. Modeling genetic interactions and their effects for a certain cell lineage is one strategy. The complementary view, or the sum of all interactions in all the cells, provides the genomic perspective.

As [37] mentioned, many network inference methods have been introduced in recent years, and numerous comparisons have been made. Akutsu and colleagues [38];[39];[40];[41]; [25] shows that the type of the data (simulated, real), the network size, the number of samples, the amount of noise, the experimental design (observational, experimental, interventional), the type of the underlying interaction structure (scale-free, random, small-world), the error measure (global, local), and other factors all appear to have a significant impact on the outcomes of such technical comparisons between GRNs. Because of this, it is doubtful that there is a single “correct” approach that works optimally in all biological, technological, and experimental design circumstances.

However, there is evidence that the differences might not be as significant as recently thought if one asks less technical and more biological questions about the meaning of the inferred networks, i.e., by evaluating the biological consistency of inferred networks resulting from different network inference methods. Table 3.1 shows a variety of GRNs inference methods that could be used to compare the resulted GRNs.

Table 3.1 Brief overview of statistical network inference methods that have been introduced in recent years and the key methods (second column) on which the inference algorithms are based on to estimate interactions.

References	Method	Name
[42]	Mutual information	RN
[24]	Mutual information, DPI	Aracne
[27]	Mutual information with background	CLR
[26]	Maximal mutual information	C3Net
[25]	Bagging C3Net	BC3Net
[29]	Regression	GENIE3
[30]	Full partial correlation	GGM
[34]	Conditional mutual information	MRNet
[43]	Three-way mutual information	MI3

C3Net, BC3Net, and Aracne [12] are examples. So, it is doubtful that a single technique will perform better than all others under all circumstances; rather, a variety of techniques will produce an overlapping spectrum with the ability to infer biological information.

3.2 Competitive Product Analysis

This section presents a comparative analysis of existing recent software tools and web applications for Gene Network Visualization.

DeepSEM [1], is a deep generative model that can jointly embed the gene expression data and simultaneously construct a GRN that reflects the inner structure of gene interactions in single cells without relying on any additional information. DeepSEM develops a neural network version of the structural equation model (SEM) to explicitly model the regulatory relationships among genes. The overall framework of DeepSEM is a beta-variational autoencoder (beta-VAE)23 in which the weights of both the encoder and decoder functions represent the adjacency matrix of the GRN. However, DeepSEM does not provide user with GRN visualization; it provides the user with a low-dimensional representation using UMAP.

GeNeCK [46], is a web server with a user-friendly graphical interface that allows users to upload their own gene expression data and choose their preferred method to infer and visualize the network, as well as integrate different methods to obtain a more confident result. Genes with a high degree of connection (hub genes) will be plotted with different colors. Users can interactively explore the constructed network. Clicking on a specific gene will highlight the gene itself along with its connected neighbors, and the corresponding information will be displayed at the bottom. However, the current version of GeNeCK does not provide a function

for users to download the figure or to compare the results of different tools in one session. Moreover, it makes the combination of several probabilistic inference strategies easily available, but there is no possibility of selecting a subset of genes to be considered as regulators during inference. Figure 3.1 shows the GRN visualization produced by GeNeCK website.

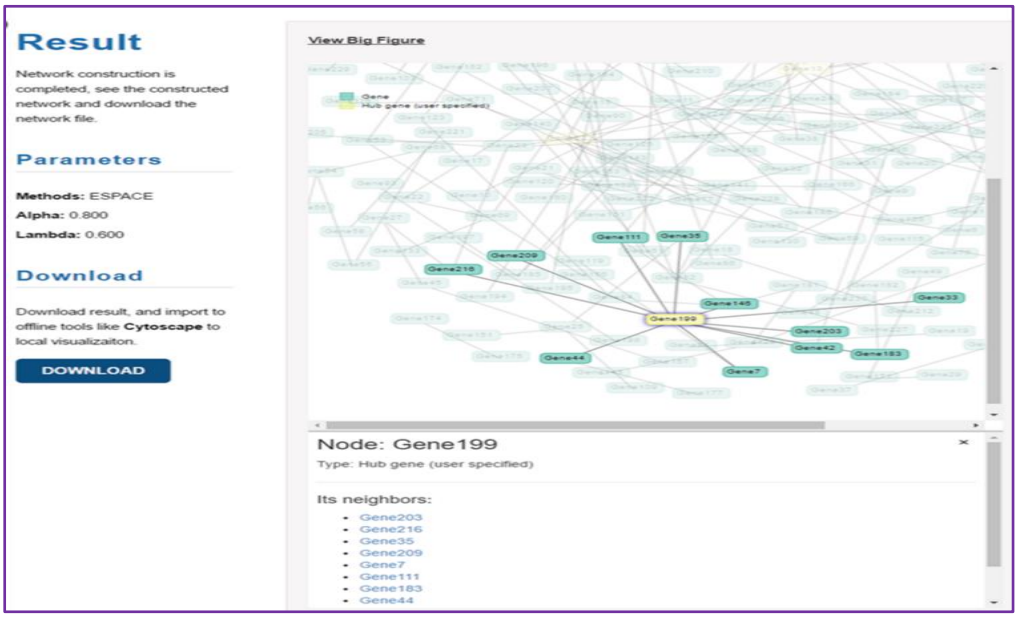


Figure 3.1Visualization of the constructed network in the GeNeCK result page.

NetworkAnalyst [44], is similar to GeNeCK but it provides more tools and features to biologists. Users can perform gene expression profiling for 17 different species. In addition to generic PPI networks -Proton pump inhibitors represent a class of medications used to treat pathological related to the stomach's acid production- [45]. Users can create cell-type or tissue-specific PPI networks, gene regulatory networks, gene co-expression networks as well as networks for toxicogenomics and pharmacogenomics studies. The resulting networks can be customized and explored in 2D, 3D as well as Virtual Reality (VR) space. For the meta-analysis, users can visually compare multiple gene lists through interactive heatmaps, enrichment networks, Venn diagrams, or chord diagrams. In addition, users have the option to create their own data analysis projects, which can be saved and resumed later. Also, they have visual analytics methods to address several key challenges in biological big data analysis. NetworkAnalyst accepts five types of data inputs - one or multiple gene lists, a single gene expression data table, multiple gene expression data tables, raw RNAseq reads as well as common network files. NetworkAnalyst aims to provide a web-based tool for creating and visualizing biological networks to complement the widely used stand-alone tools such as Cytoscape.

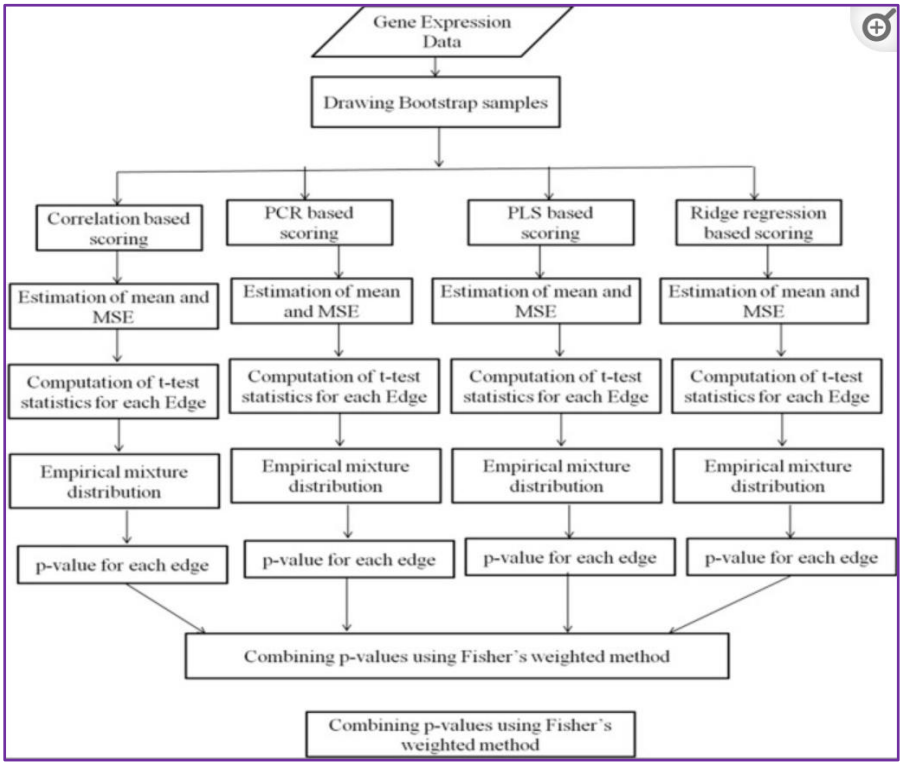


Figure 3.2 GRN construction pipeline

GRN Construction[47] is an online tool with a user-friendly and interactive web page for merging the results from four different methods—correlation, principal component regression, partial least squares, and ridge regression by using Fisher's weighted method to create a consensus GRN. Users can upload gene expression data for constructing consensus GRN. The output obtained from analysis will be available in downloadable form in the result window of the web tool and it can be visualized using network visualization tools like Cystoscope. Figure 3.2 show the main pipeline for GRN Construction tool.

4 System Design and Development

In this chapter, we start by presenting the methodology we followed to achieve our goal in the project, then show how we collected the system requirements. Next, we showed GRN VTOOLS system design, data design and interface design. Finally, we explain our implementation.

Methodology

In developing the GRN VTOOLS website, our team embraced the Agile methodology, valuing its flexibility and focus on continuous iteration. This approach allowed us to effectively adapt to changing requirements and continuously improve our product through each phase of development. Agile's emphasis on client satisfaction and iterative progress ensured that we could test and refine our product at every stage, keeping the end-user's needs at the forefront.

Integral to our methodology was the Scrum framework, a specific Agile methodology that emphasizes iterative development management. The framework is structured around three roles (Product Owner, Scrum Master, and Development Team), five events (Sprint Planning, Daily Scrum, Sprint Review, Sprint Retrospective, and The Sprint), and three artifacts (Product Backlog, Sprint Backlog, and Increment), provided a clear and organized approach to project management. Our team, consisting of these roles, engaged in regular Sprint Planning to determine the scope of each sprint, followed by daily stand-up meetings to assess progress and tackle any issues. The Sprint Reviews allowed us to showcase our progress and gather feedback, while Retrospectives offered a chance to reflect and improve upon our processes.

Utilizing tools like GitHub [1] for code sharing and version control, we enhanced our collaboration and efficiently tracked changes. Jira[2] was also a crucial tool, aiding in visualizing our product backlog, monitoring sprint progress, and maintaining transparency across the team. This combination of Agile and Scrum, supported by these tools, ensured that our development process was not only effective but also aligned with delivering a user-friendly and high-quality website.

1 Github: https://github.com/4n8x/GRN-Vtool/tree/main

2 Jira: https://2023-3rd-gp1.atlassian.net/jira/software/projects/GP/boards/1/timeline

4.1.1 Scrum Team

Skill Set Requirements

Technical Skill Required	**What is the current level of the team (*beginner- intermediate- advanced*) for each skill? How will the gap be bridged? *(if necessary)* Learning plan**
Front-back-end devolvement	intermediate
R programming language	Beginner, Will be bridged by self-learning and taking online courses on w3shcools[8].
Bioinformatics GRN tool (DIANE)	Beginner, full documentation, and examples [5].
Bioinformatics GRN tool (SeqNet)	Beginner, full documentation, and examples [6].

4.1.2 Roles and Responsibilities

Scrum Team	
Product Owner:	Dr.Nuha Saud Bin Tayyash
Developers:	Rawan Alhabeeb, Nora Alanzi , Norah Albarrak, Ream alquraishi
Scrum Master (SM):	Dr. Maha Al-Yahya
Stakeholders:	Dr. Balsam Alsugair, Dr. Alhanoof Althnian, and Dr. Alia

GRN VTool source code: https://github.com/4n8x/GRN VTOOLS.git
Jira:https://2023-3rd-gp1.atlassian.net/wiki/home?cloudId=1d967434-c1c2-4476-b08a-15b09f7194ea

4.2 System Requirements

4.2.1 System Users

GRN VTOOLS users are biologists who want to visualize gene regulatory network in more convenient approach with a variation of features which makes their work easier. Basic technical experience is required to use this tool.

4.2.2 Requirements Elicitation and Analysis

For our requirements elicitation, we used the interviews method, we got one interview with a biologist, Dr. Soaad Aldahian, and we got the following results:

Dr. Soaad used a website (GeneMANIA) to visualize GRN. Because she uses a website she did not face any installation issues. GeneMANIA features are: enrichment analysis, gene function prediction and network generation. GeneMANIA has interactive graphs and analyze partial networks but it uses a single GRN inference method. Dr.Soaad mention that she is satisfied with GeneMANIA features.

4.2.3 Similar system comparison

Table 4.1 summarizes the features provided by the software tools and web-based described in the literature review chapter 3.

We found that biologists need lots of procedures to help them in getting the appropriate result. Conducting this analysis by using software tools often require important prior knowledge and skills in statistics and computer programming. In addition, tools dedicated to the analysis, exploration, and visualization of RNA-Seq data are very often dispersed.
From the literature review, we found that the current version of GeNeCK does not provide a function for the users to download figures and save the analysis results. For NetworkAnalyst, they do not provide their users with the ability to select a subset of genes to get a partial network which is an extremely useful feature when dealing with large datasets. For DeepSEM, it is hard to visualize the GRN, however, the user gets a low dimensional representation as an image that summarizes the GRN.

Unlike these previous examples, we propose a website with a graphical user-friendly interface that allows biologists to run many GRN tools and saves biologists from many technical difficulties when using software tools that ask them for a lot of programming commands. Also, we allow the biologist to choose from a variety of tools in one-place. Our website provides biologists with an interactive network where they can click on genes to get extra information by representing the gene name or id and highlighting the gene itself along with its connected neighbors. GRN VTOOLS contains a search feature to help them navigate in the large network, partial network extraction for more in-depth analysis are essential functionalities especially when dealing with a large network that GeNeCK and NetworkAnalyst don't provide. In addition, we will allow the biologists to upload two different datasets and run one tool they

choose and display the two GRN simultaneously. That could help them in comparing the generated networks in a convenient where there is no other tool or website provides this feature.

Table 4.1 Copmpatitive analysis

Features	Software tools			Web based			
	SeqNet	DIANE	DeepSEM	GeNeCK	NetworkAnalyst	GRN Construction	GRN VTOOLS
Interactive gene network	Yes	Yes	No	Yes	Yes	No	Yes
Multiple tools	No	Yes	No	Yes	Yes	Yes	Yes
Search bar	No	No	No	No	Yes	No	Yes
Compare two networks	Yes	No	No	No	Yes	No	Yes
Similarities and differences highlighting	Yes	No	No	No	No	No	No
partial network extraction	No	Yes	No	No	No	No	No
Run one tool on two different datasets	No	No	No	No	Yes	No	Yes
Visualize multiple GRNs in one session	No	No	No	No	No	No	Yes

4.2.4 User Interactions

We use a high-level use case diagram as illustrated (see figure 4.1) to show the key interactions between the system user (Biologist) and (Admin) with our system.

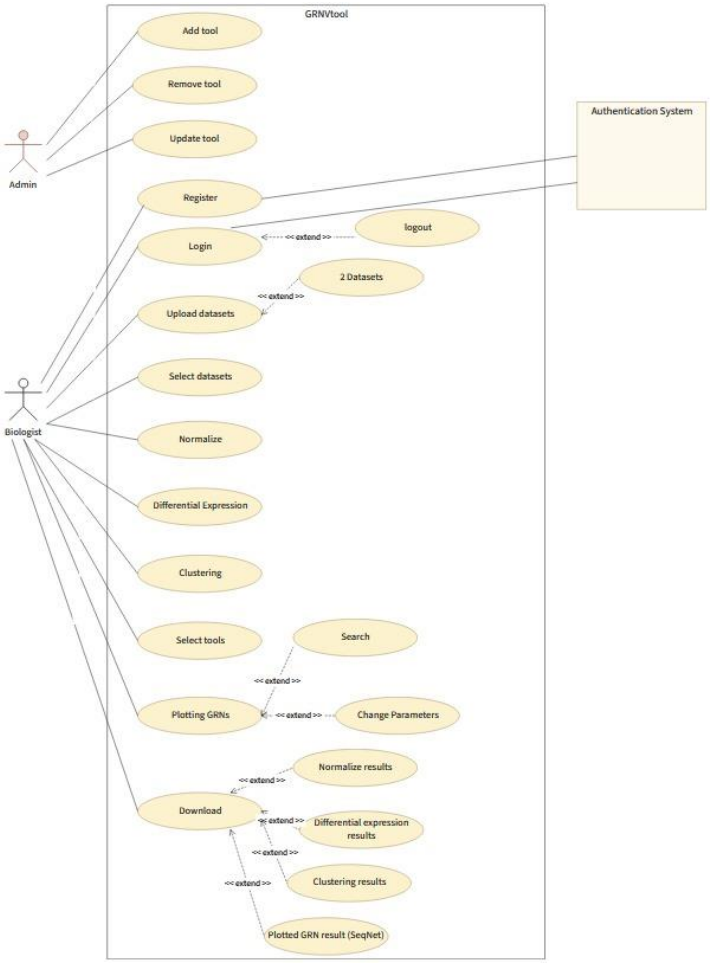


Figure 4.1: Use case diagram

Figure 4.1 shows the GRN VTOOL use case diagram, which has two main actors which are the biologist and admin, and one external actor which is the authentication system. The biologist can log in to their accounts and register if they are new in the system, the admin can add and remove tools from the system. To visualize the GRN The biologist must upload a valid .csv dataset file for the gene expression. Then, he has the option to visualize the GRN for the uploaded dataset immediately or run a preprocessing step such as normalization, differential expression, and clustering. User will be able to import another dataset, the user will have to select tool to generate the GRN network using the selected tools from DIANE, SeqNet. The biologist can also change parameters for the Plotted GRN. Once the result visualization is displayed in the screen the biologist can search for gene in the network, compare plotted GRNs, and highlight connections on the plotted GRN. After the analysis process, the biologist can download the processes he made, and save the results as csv file, for SeqNet tool, user will be able to save the GRN plotted results.

○ Roadmap and Product Backlog:

Here we showed the GRN VTOOL roadmap and product backlog.

▪ Product Roadmap

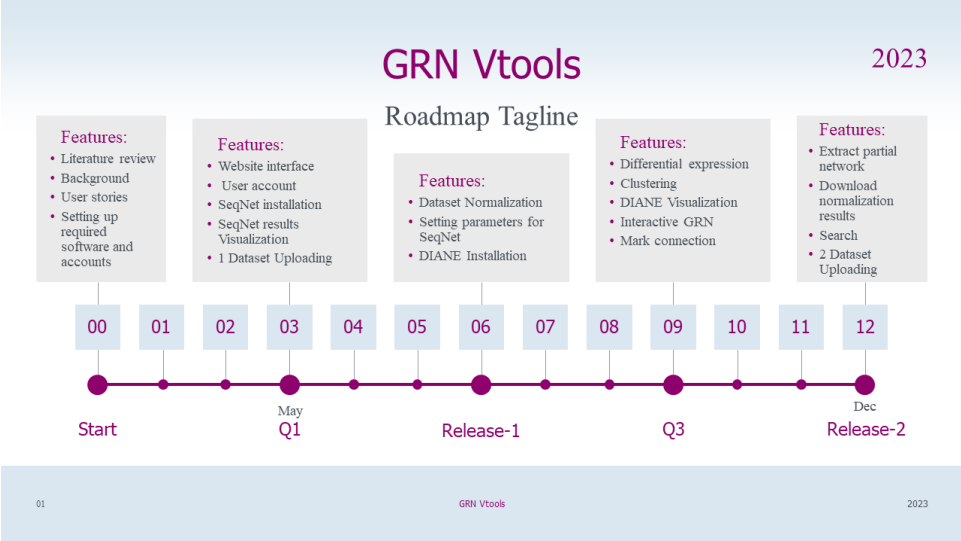


Figure 4.2: project Roadmap

Figure 4.2 shows our roadmap for this project (GRN Vtools). It has 5 sprints and 2 releases over two semesters. The first semester has 3 sprints and 1 release, While the second will have 2 sprints and Final release.

In sprint 0, we presented a GRN background followed by a literature review to understand the existing GRN tools. Then, showed GRN Vtools system description through background and user stories. Finally, we set all the accounts for the tools that we needed in our implementations.

In sprint 1, we made the website interface, set up user account through login and register. Also, we installed the tool SeqNet in shiny server, visualized its results, while enabling user to upload 1 of their own datasets for the analysis.

In sprint 2, we installed DIANE. Then, we used it to normalize the datasets for the other tool Seqnet. We also set up parameters editing for SeqNet.

In sprint 3, we implemented the main functionalities of the GRN Vtools, which was running the main pipeline for RNA-seq analysis using: differential expression, clustering and visualizing the GRN inferred by DIANE. In addition, we converted the results of GRN

to interactive graph to let the user select a gene and present its information and view connected nodes to this gene and marked the connections.

4.3 Product Backlog

Table 4.2 Copmpatitive analysis

ID	PBI (user story)	Size (Story points)	Type (Feature, defect, technical work, knowledge acquisition)	Status (To do, in progress, or Done)	Acceptance Criteria The conditions of satisfaction that must be met for that item to be accepted.
1	As a user, I want to be able to register a new account, so that I can maintain my account	2	Feature	Done	The user should be asked to enter his username and password
2	As a user I want to be able to log in to my account, so that I can access the website.	2	Feature	Done	As a user, if I didn't enter the correct username or password, I should receive an error message.
3	As a user, I want to be able to upload datasets on the website, so that I can view the gene network.	3	Feature	Done	As a user, if I uploaded the dataset, I should see "upload completed" on the screen.
4	As a user, I want to be able to choose a tool on the website, so that I can choose the visualization tool.	3	Feature	Done	As a user, if I choose a tool, I should see a GRN inferred by the selected tool.
5	As a user, I want to be able to edit the parameter, so that the function works efficiently.	4	Feature	Done	As a user, if I changed the parameter, I should recognize the changes.
6	As a user I want to be able to view gene network so that I can analyze it.	4	Feature	Done	As a user, if I clicked generate button, I should be able to view gene network.
7	As a user, I want to be able to click on genes so that, I can get extra information.	5	Feature	Done	As a user, if I clicked on the gene, I should see extra information displayed.

8	As a user, I want to have a search feature, so that I can navigate in a large network.	5	Feature	Done	As a user, if I selected a gene from the list, I should be directed to the gene of interest in the network.
9	As a user, I want to have automated differences highlighting between two GRNs so that, I can get faster analysis	3	Feature	To do	As a user, if two GRNs generated, I should see the differences highlighted.
10	As a user, I want to be able to extract a partial network so that, I can have a more in-depth analysis.	4	Feature	To do	As a user, if I selected a partial network, I should see it in a new window.
11	As a user I want to be able to upload two datasets simultaneously, so that I can compare them.	3	Feature	Done	As a user, if I uploaded two datasets, I should be able to view them both at the same time.
12	As a user I want be able to view the connections for the gene of interest, so that I can get comprehensive knowledge about the gene.	4	Feature	Done	As a user, if I searched for a gene, I should be able to see the connection.
13	As an admin, I want to add/ remove tools from the server so that I can add a variety of tools for the users.	3	Feature	Done	As an admin, If I add/remove a tool, it should be added or removed.
Non-Functional requirements					
14	As a user, I want the website to be available 99% of the time when I try to access it so that I don't get frustrated and find another site to use.	2	Feature	Done	The website being down for 20 minutes max.
15	As a user, I want the website to be easy to use and very clear so that I don’t make more than 5 mistakes while working.	3	Feature	Done	The user not making more than 5 mistakes while working.
16	As a user, I want to generate a genes network easily with the process not taking more than 20 minutes to do.	3	Feature	Done	The network generation process takes 20 minutes max.
17	As a user, I want the website to be able to handle datasets with hundreds of genes and not crash.	3	Feature	Done	The website not crashing with 1000 genes.

18	As a user, I want the website to be reliable with genes network results every time using the same tool and same dataset.	3	Feature	Done	The website has an accuracy of 90% or higher.

4.4 System Design

4.4.1 Architectural Diagram

For the system's architecture, GRN VTOOLS website will use client-server architecture, the reason behind that is to be able to serve a lot of users and their requests at the same time. Figure 4.3 illustrate the model of client-server architecture [48] :

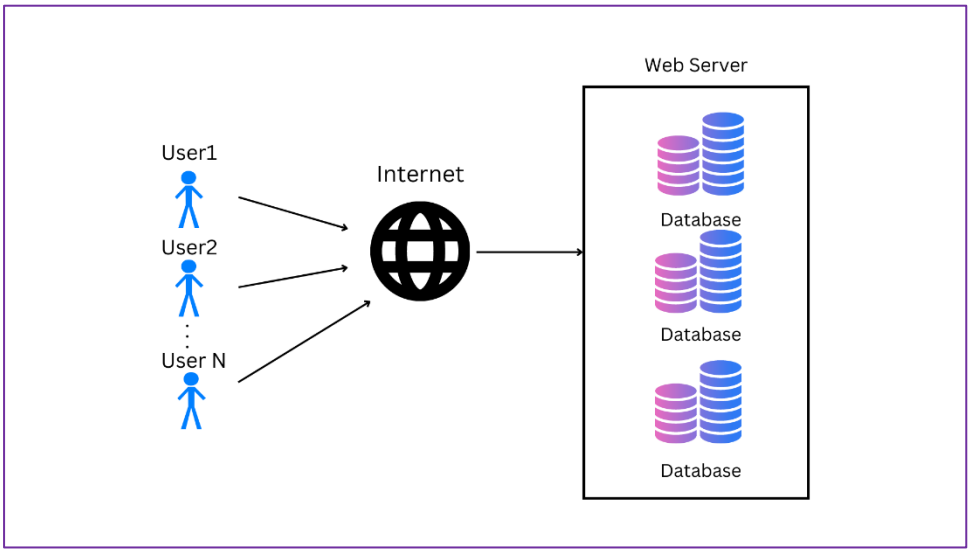


Figure 4.3: client-server architecture

The GRN VTOOL system composes of a user, admin and AWS to save users database where they interact with shiny server (GRN VTOOL) as illustrated below (Figure 4.4):

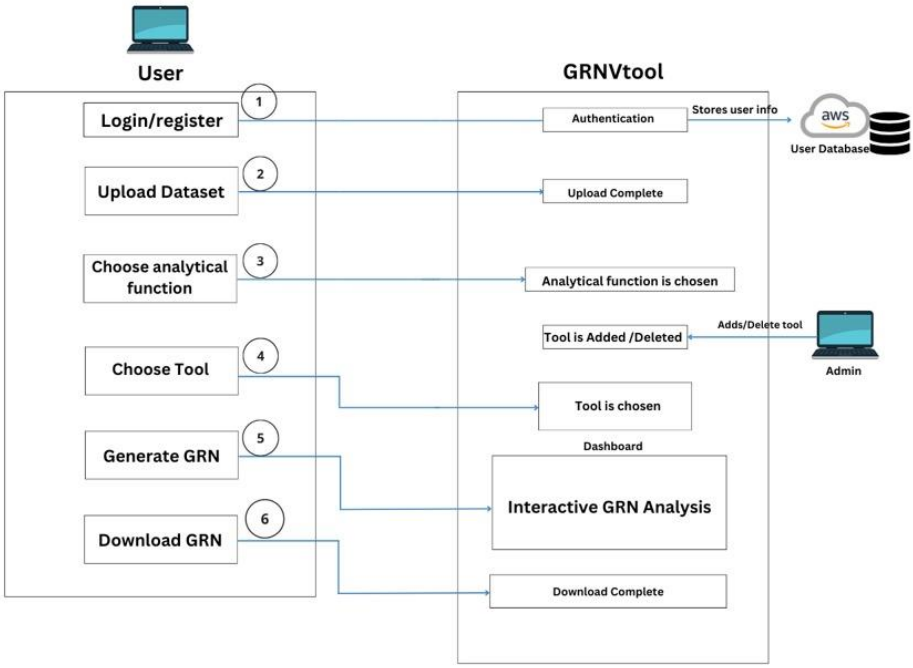


Figure 4.4: System architecture diagram

The admin role is to install the GRN tools for the visualization and he is responsible for maintaining the tools and performing any required update or maintenance. In this project, we chose to install DIANE and SeqNet tools.

The users must create a profile to access GRN Vtools using a username and a password, then, the system will create a profile and save it in the user database where an authentication is required each time the user login through AWS database as shown in step 1. Then, the user can upload one or two datasets, to run the analysis as shown in step 2, for the second dataset, we assume it has the pipeline done for it already, and the user should go through the whole pipeline for the first dataset .

The GRN analysis starts with the user selecting one or two GRN tools either DIANE or SeqNet. The user can use two datasets and select a single GRN tool to be run on the datasets, the system will record the user choice of tool to be used as visualization method as shown in step 3.

The user will follow the pipeline analytical functions steps (normalization, differential expression and clustering) to be performed on the uploaded dataset. The system will run DIANE selected fucntion as shown in step 4. Note that DIANE and Seqnet require these analytical functions as prerequisites to do GRN inference. When the user click generate network (step 5) In addition, the system will check if the user has selected two tools or uploaded two datasets to split the window into two part to display two GRNs. User will be able to download the normalization results from GRN analysis as a csv file as shown in step 6.

The user can click on any part of the network to get extra information where the system will get these gene information from the datasets provided. The user can type a gene name in the search bar and the system will highlight this gene and focus the visualization on the area where the gene is located.

4.4.2 Class Diagram

Below we show the class diagram (figure 4.5) of GRN Vtools. This class diagram illustrates the various components and interactions within our system, including datasets, users, administrators, tools, and gene regulatory networks

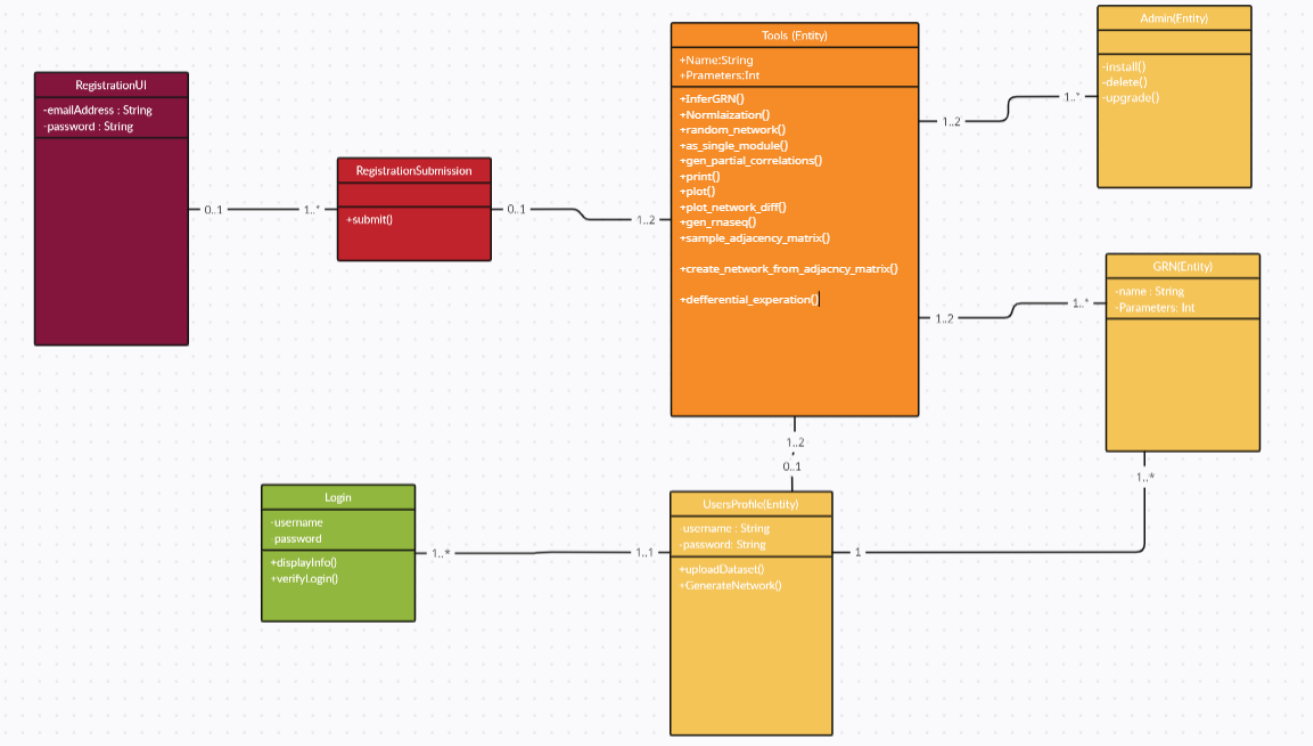


Figure 4.5: Class diagram

User: Represents a user in the system. It has the following attributes:

Name: The name of the user.

Password: The password of the user.

It also has the following operations:

Register(): Registers a new user in the system.

Login(): Allows a user to log in to the system.

Logout(): Logs out the currently logged-in user.

Admin: Represents an administrator user in the system. It has the following operations:

Install(): Installs a tool in the system.

Delete(): Deletes a tool from the system.

Upgrade(): Upgrades a tool in the system.

Tools: Represents a tool in the system. It has the following attributes:

Name: The name of the tool.

Parameters: The parameters of the tool.

It also has the following operations:

inferGRN(): Infers a Gene Regulatory Network using the tool.

Normalize(): Normalizes the data using the tool.

random_network(): Generates a random network using the tool.

as_single_module(): Converts the network into a single module using the tool.

gen_partial_correlations(): Generates partial correlations using the tool.

print(): Prints the results using the tool.

plot(): Plots the results using the tool.

plot_network_diff(): Plots the difference between two networks using the tool.

gen_rnaseq(): Generates RNA-Seq data using the tool.

sample_reference_data(): Samples reference data using the tool.

get_adjacency_matrix(): Retrieves the adjacency matrix using the tool.

create_network_from_adjacency_matrix(): Creates a network from the adjacency matrix using the tool.

GRN: Represents a Gene Regulatory Network. It has the following attributes:

Name: The name of the GRN.

Parameters: The parameters of the GRN.

This class diagram illustrates the various components and interactions within our system, including datasets, users, administrators, tools, and gene regulatory networks.

4.4.3 Component Level Design

Here we draw a component diagrams for three main functions, data upload, data normalization, and network visualization.

Data Upload Component:

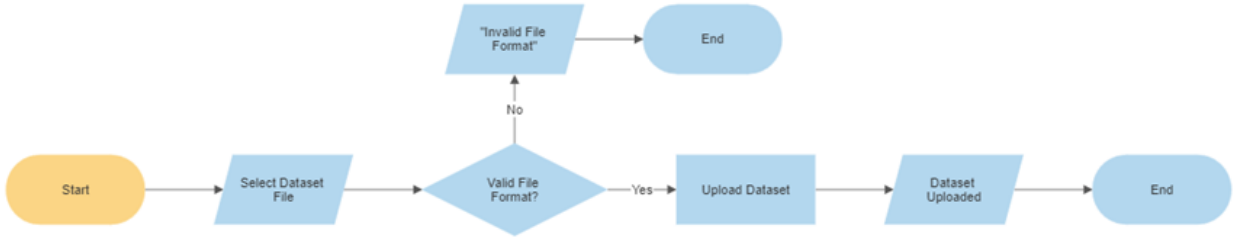


Figure 4.6: Data Upload Component

Start

Input datasetFile

If datasetFile is a valid file format:

Upload datasetFile

Output "Dataset Uploaded"

Else:

Output "Invalid File Format"

End

Dataset Normalization Component:

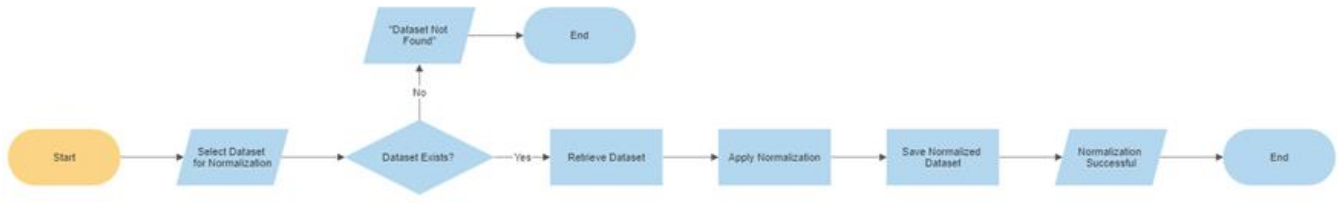


Figure 4.7 Dataset Normalization Component

Start

Input datasetForNormalization

If datasetForNormalization exists:

Retrieve datasetForNormalization

Apply normalization to the dataset

Save the normalized dataset

Output "Normalization Successful"

Else:

Output "Dataset Not Found"

End

Network Visualization Component:

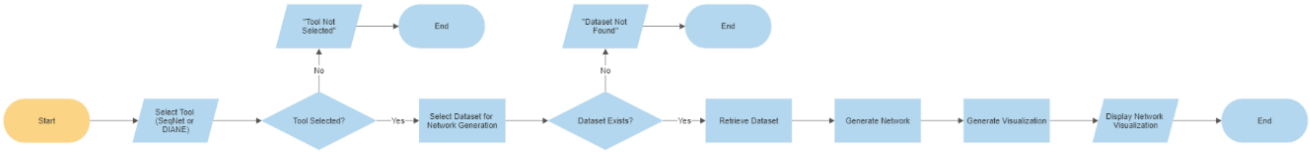


Figure 4.8 Network Visualization Component

Start

Input selectedTool

If selectedTool is "SeqNet" or "DIANE":

Input datasetForNetworkGeneration

If datasetForNetworkGeneration exists:

Retrieve datasetForNetworkGeneration

Generate network from the dataset using the selected tool

Generate visualization for the network

Output "Display Network Visualization"

Else:

Output "Dataset Not Found"

Else:

Output "Tool Not Selected"

End

5 Data Design

5.1.1 Data Models

The ER (Entity-Relationship) diagram for the GRN VTOOLS system consists of four main entities and their relationships. The main entities include "Users" , "GRN" , "Admin" and "Tools" as figure 4.9 shows. The "Users" entity stores information about the users of the system, including their names and passwords. Each user can visualize one or more GRN, represented by the "GRNs" entity, which contains an ID, normalized dataset, differential expression genes list and the results of clustering genes and the visualized GRN as lists of genes and their connection as pair.

The relationship between "Users" and "GRNs" is a one-to-many relationship, indicating that a user can visualize multiple GRNs.

The "Tools" entity captures the various analytical tools utilized within the platform. Each tool is defined by a name and associated parameters that dictate its operation and influence on the GRN generation process. The "Using" relationship underscores that multiple tools can be applied to one or two data, offering users a versatile approach to their analyses.

Additionally, the system includes an "Admin" entity, which represents the administrator of the system. The "Admin" entity stores the admin's ID, name, and password. Similar to users, the admin can also upload one or more datasets.

This ER diagram illustrates the relationships between users, diagrams, and the admin, providing a clear overview of the structure and connections within the system. It showcases how users and admin interact with datasets, leading to the generation of diagrams representing gene networks.

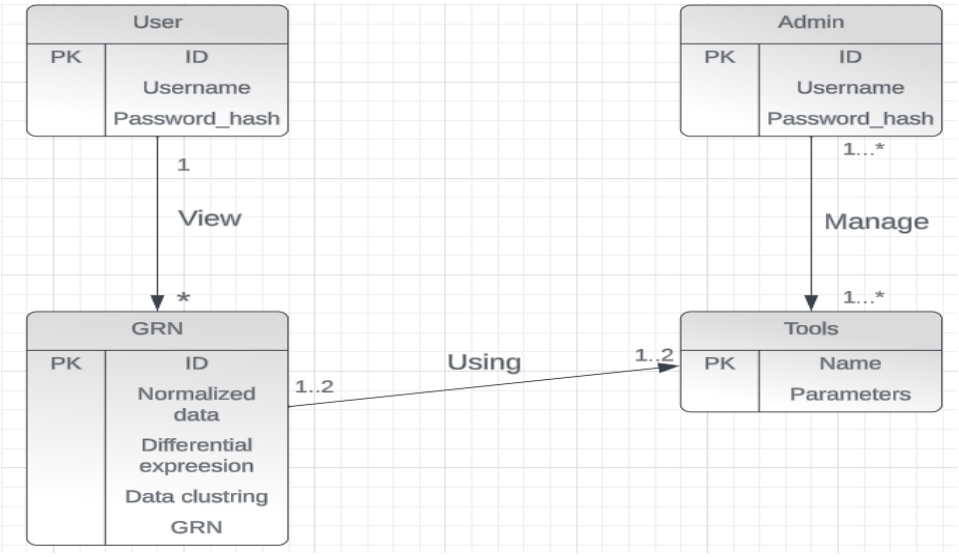


Figure 4.9 ER data model

5.1.2 Data Collection and Preparation

In the GRN VTOOLS platform, the system collects data from an external source, specifically RNA-seq data, which is commonly used in gene regulatory network analysis. The data creation process involves conducting RNA-seq experiments, where gene expression levels are measured for different samples or conditions [49].

Once the raw RNA-seq data is generated, it undergoes a data collection process where it is gathered and organized for analysis. This may involve obtaining the data from public repositories, such as ARCHS4, or directly from user-uploaded files. The data is typically organized in a tabular format, where rows represent genes and columns represent samples or conditions [7].

5.2 Interface Design

The Interface Design of the GRN VTOOLS platform is carefully crafted to provide users with a seamless and intuitive experience. With a focus on user-friendly navigation and visually appealing elements, the interface is designed to enhance usability and streamline workflow.

Each page within the platform offers a well-organized layout, ensuring that users can easily access the desired functionalities. Whether it's uploading and normalizing datasets, performing exploratory analysis, or generating network models, the interface provides clear instructions and intuitive controls. Additionally, the use of modal dialogue boxes for important notifications and feedback ensures that users stay informed about their actions and any errors encountered. The thoughtfully designed interface empowers users to effortlessly navigate through the platform, making their journey efficient and enjoyable.

These UX guidelines helped shape the design of GRN VTOOLS, providing a user-friendly and efficient interface for biologists to access and utilize gene regulatory network tools:

1. Simplicity: The interface design prioritized simplicity by keeping the layout clean and organized. It aimed to present the essential elements without unnecessary clutter, allowing users to focus on the core functionalities and tasks.

2. Intuitiveness: The interface was designed to be intuitive, ensuring that users can easily navigate and access the available options. Clear labels, familiar icons, and logical placement of elements contribute to a user-friendly experience.

3. Visual Clarity: The use of visual elements, such as buttons and dialogue boxes, was implemented to provide clear and concise information to users. The visual hierarchy and prominence of important elements help users quickly understand and interact with the interface.

4. Feedback and Validation: Prompt feedback and validation were incorporated into the interface design. For example, error messages in modal dialogue boxes inform users about incorrect login credentials, helping them identify and resolve the issue. This real-time feedback enhances the user experience and reduces frustration.

5. Personalization: The inclusion of personalized welcome messages in modal dialogue boxes upon successful registration or login contributes to a sense of familiarity and value for users. This personalization creates a positive user experience and establishes a connection with the platform.

6. Consistency: The interface design aimed to maintain consistency throughout the platform. Consistent placement of elements, color schemes, and visual styles across different pages and features contribute to a cohesive user experience, allowing users to easily navigate and understand the interface.

7. Usability Testing and Iteration: The development process incorporated usability testing and iterative design. User feedback was actively sought and incorporated to identify and address any usability issues or areas of improvement. This user-centric approach ensures that the interface design aligns with the needs and preferences of the target users.

Sitemap

The sitemap of the GRN VTool platform outlines a well-connected flow of pages, ensuring a seamless user experience. The sitemap (figure 4.10) illustrates the logical and interconnected structure of the platform, enabling users to efficiently move between different functionalities and make the most out of their GRN VTOOLS experience.

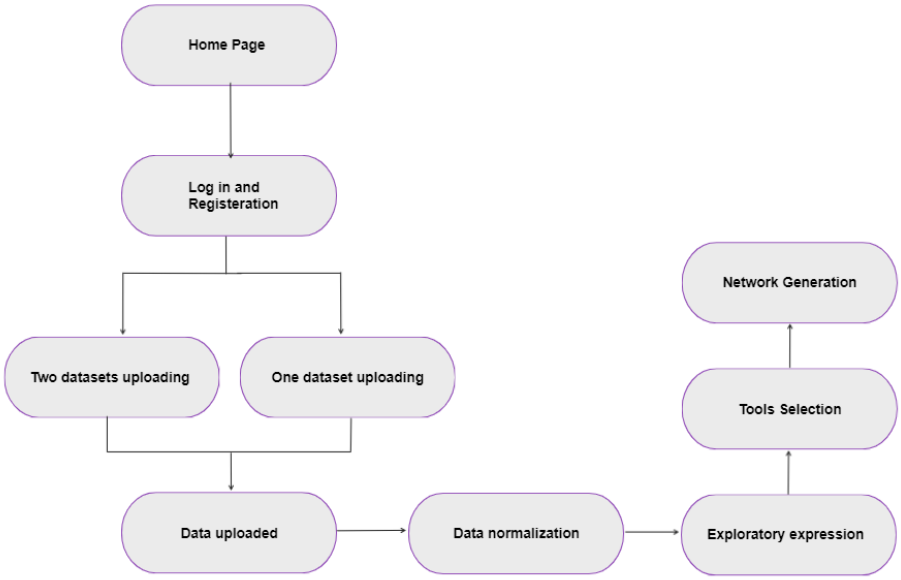


Figure 4.10

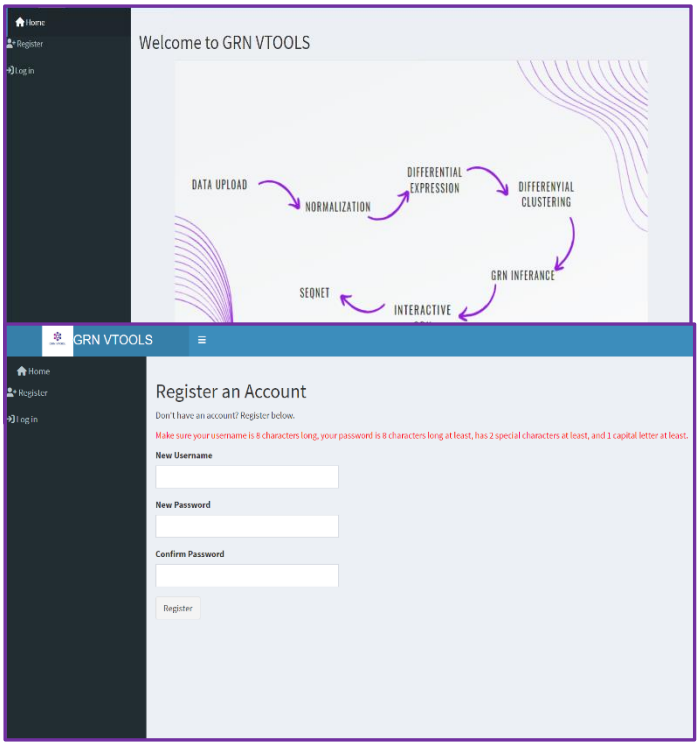


Figure 3 Registration Page

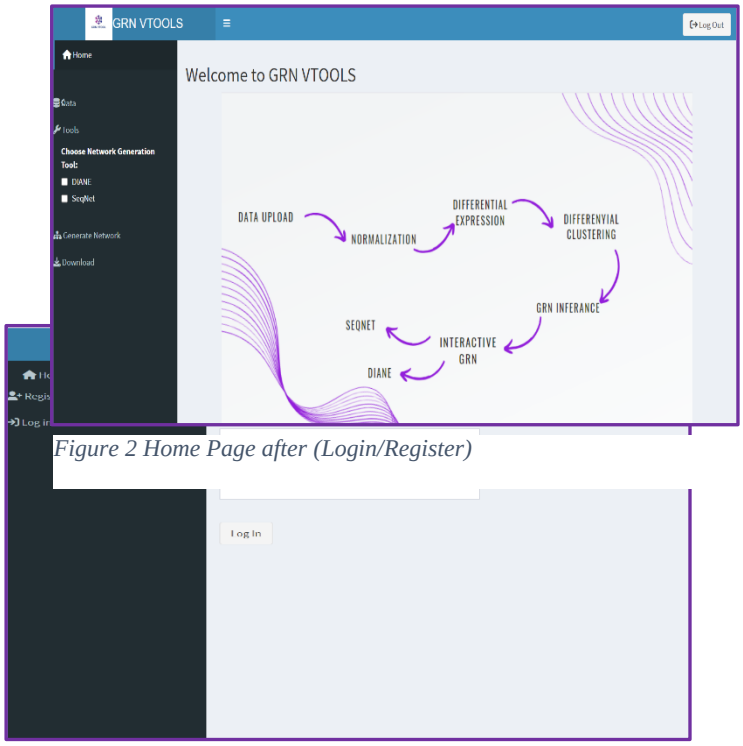


Figure 4 Log-in page

Figure 2 Home Page after (Login/Register)

Before registration figure 1 or login figure 2, the home page offers only a preview of the GRN VTOOLS capabilities. Full access to the site's features is restricted to ensure the privacy and security of your scientific exploration. Registration is straightforward, and upon logging in, a comprehensive suite of tools becomes available. These tools include data upload, data normalization, differential expression analysis, differential clustering, and interactive GRN (Gene Regulatory Network) inference and visualization, all facilitated by the use robust DIANE and SeqNet systems.

In the registration process, users are guided to create secure login credentials by adhering to specific requirements highlighted in red. A username and password must be crafted with a minimum of 8 characters, incorporating at least 2 special characters and 1 uppercase letter, enhancing the security of user accounts. Once registered, accessing GRN VTOOLS is straightforward users need only enter their username and password on the login page, ensuring a swift and secure entry point to the platform's extensive features.

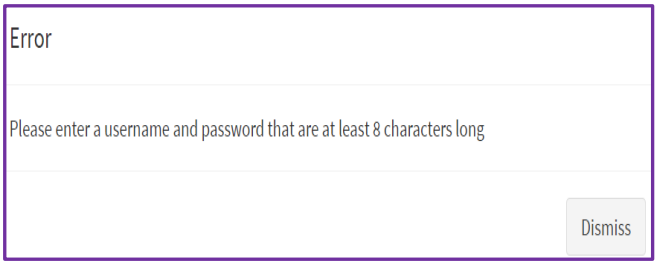


Figure 5 Warning

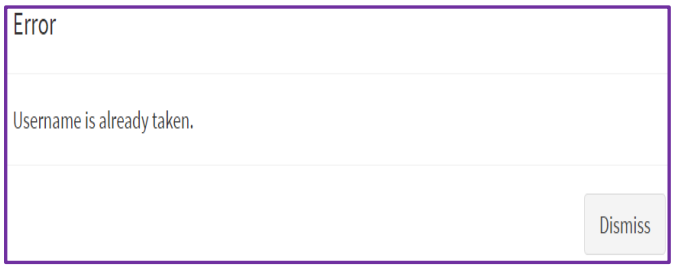


Figure 6 Warning

During the registration process, users may encounter critical prompts ensuring the integrity of their account setup. One such warning advises that both username and password need to fulfill a length requirement of at least 8 characters (figure 5), reinforcing the need for robust access credentials. Another alert (figure 6) notifies the user if the chosen username is already in use, thereby maintaining unique identifiers for each user. These measures are in place to ensure that user accounts are secure and distinctly recognizable within the GRN VTOOLS platform.

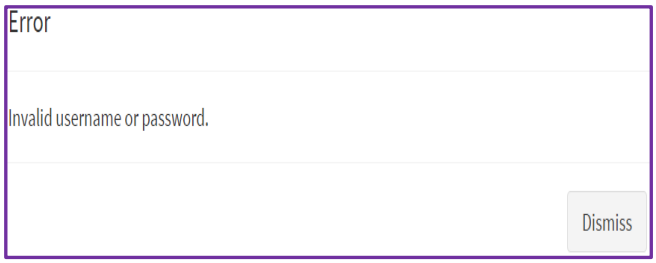


Figure 7 Log-in warning

In the login sequence, a critical safeguard is in place to prevent unauthorized access: a warning message (figure 7) that appears when there is a discrepancy in the entered credentials. This message, indicating "Invalid username or password," serves as a prompt for users to re-examine and correct their login details. This alert contributes to the security protocols of GRN VTOOLS, ensuring that access is granted only to users with verified credentials, thereby protecting user data and the integrity of the platform.

Figure 8 Welcoming after (Log-in /Register)

Upon successful registration or login, users are greeted with a welcoming message (figure 8) that not only acknowledges their identity but also provides immediate direction on where to begin their journey within GRN VTOOLS. This personalized greeting reinforces the user's connection to the platform and smoothly transitions them towards the first actionable step—navigating to the Data tab.

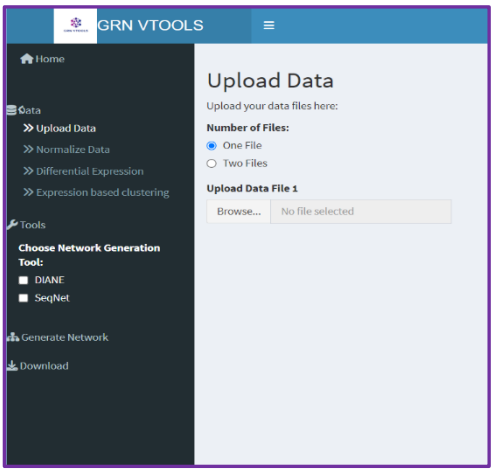


Figure 9 Upload one data file.

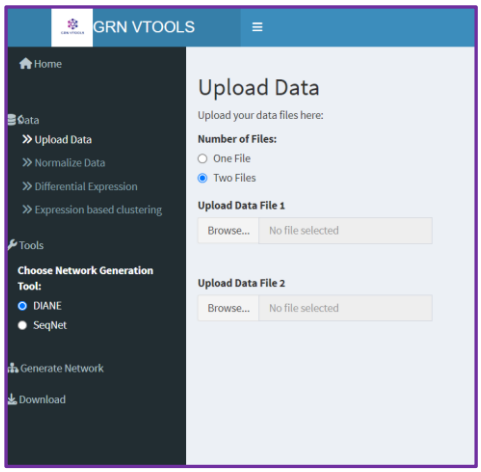


Figure 10 Upload Two data files

The initial step in utilizing GRN VTOOLS involves selecting the number of data files to upload. Users can choose to upload either 'One File' or 'Two Files'. Selecting 'One File' provides the flexibility to choose either one or both available network generation tools — DIANE and SeqNet. However, if 'Two Files' are chosen, the user is limited to selecting only one network generation tool to proceed with the data analysis. This choice dictates the subsequent options and tools that will be available for data processing and analysis in GRN VTOOLS.

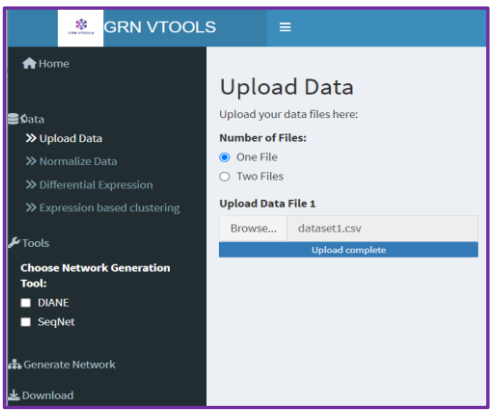


Figure 11 Uploading completion.

After the data is uploaded provides users with a comprehensive and user-friendly environment to explore and analyze their datasets. Once the data is uploaded, the interface dynamically adjusts and displays upload complete and displays the file's name that has been uploaded.

Overall, the interface design of GRN VTOOLS after data upload prioritizes usability, interactivity, and accessibility. It empowers users to explore, analyze, and visualize their datasets efficiently, making complex gene regulatory network analysis more accessible and enabling researchers to derive meaningful insights from their data (figure 11).

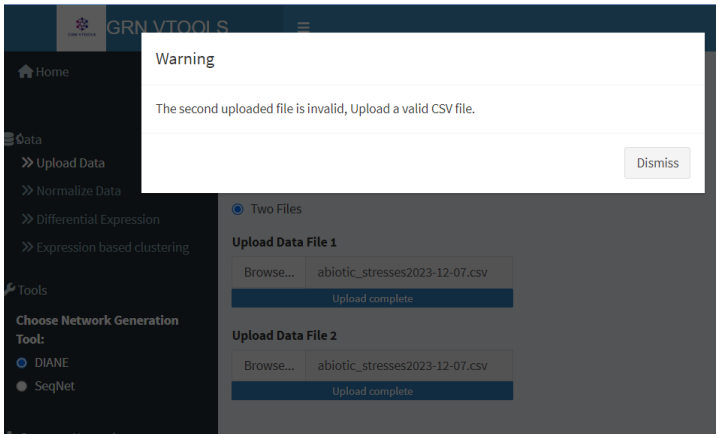


Figure 12 warning data uploading

The interface design of the GRN VTOOLS platform features a user-friendly alert system that notifies users of issues with their data uploads. If an error is detected in any uploaded file, the system promptly displays a specific warning message indicating which file is invalid. This ensures that users can quickly identify and rectify the problem, enhancing the efficiency of the data upload process.

Figure 13 Data Normalization

Following data upload, the next phase in the GRN VTOOLS pipeline is data normalization figure 13, which becomes accessible when the user presses the "Normalize" button. While the interface suggests a sequential pipeline for ease of understanding the logical order of data

preparation steps, adherence to this sequence is not mandatory. Users with pre-processed data can proceed directly to network generation, ensuring the platform's versatility for different stages of data analysis and user expertise.

Figure 14 Differential Expression

Following the normalization step, users are presented with a clear and informative Differential Expression page figure 14. This page displays a graphical representation of the M-A plot, which is crucial for identifying differentially expressed genes (DEGs). Allowing users to visually distinguish between significant (highlighted in green) and non-significant genes (in grey), aiding in the interpretation and analysis of gene expression data. This intuitive design ensures that even those with limited bioinformatics expertise can effectively utilize the tool for their research needs.

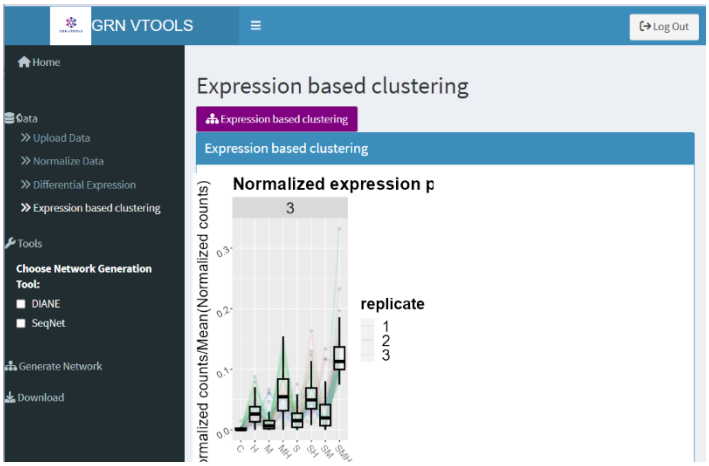


Figure 15 Expression Based Clustering

Following the differential expression analysis, the GRN VTOOLS platform guides users to the Expression Based Clustering page figure 15. This page features a box plot visualizing the normalized expression levels across multiple replicates, providing insights into the variability and distribution of gene expression within the dataset. The clear categorization of replicates and the visual differentiation of expression levels underscore the platform's commitment to delivering complex analyses in an accessible format, allowing researchers to quickly interpret clustering results and proceed with their gene regulatory network investigations.

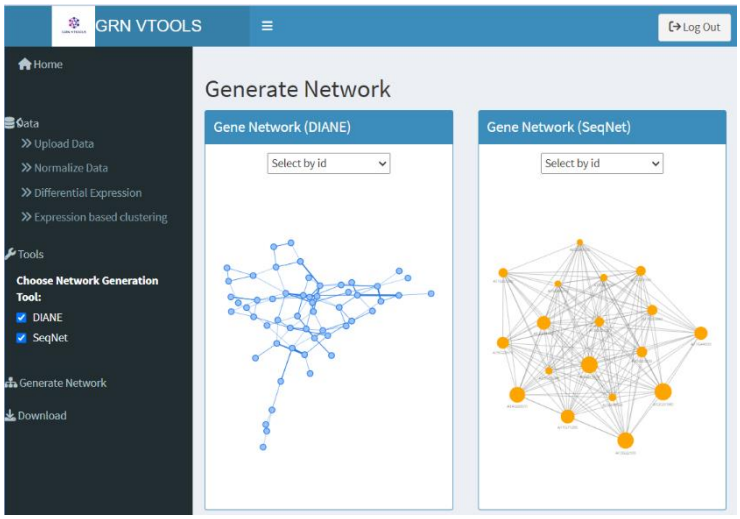


Figure 16 Network generation using two tools.

The GRN VTOOLS platform streamlines gene network generation figure 16, presenting networks created using DIANE and SeqNet tools for a single dataset. The comparative visualization allows users to examine the varied network structures and interactions revealed by the two methodologies. This feature underlines the platform's commitment to providing comprehensive analytical tools and highlights its capacity to deliver multiple analysis angles on the same dataset, thus enriching the user's research experience.

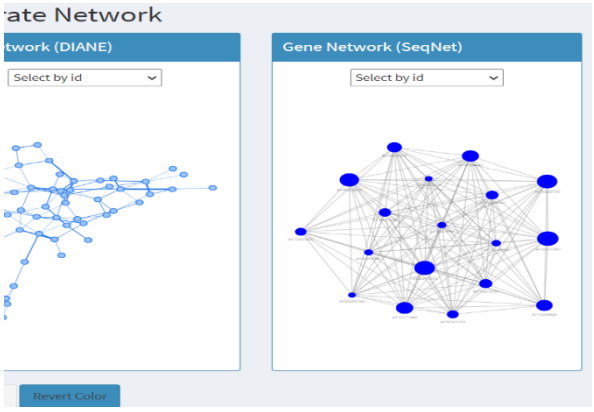


Figure 17 change color parameter

Upon using the color change functionality in the GRN VTOOLS platform, users can instantly visualize the difference in their gene networks. The 'Change Color' feature figure 17 offers an interactive way to differentiate and analyze the network dynamics by altering the node color, thus enhancing the visual distinction between the two tools—DIANE and SeqNet. This feature not only aids in visual analysis but also adds a layer of user engagement with the network representation, making the exploration of gene interactions more intuitive.

The GRN VTOOLS platform is designed with advanced analytical features such as network

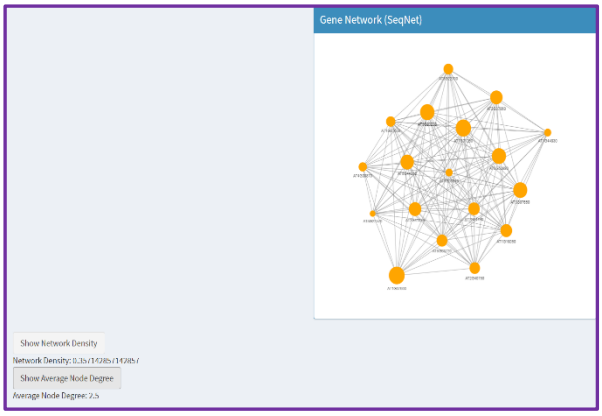


Figure 18 SeQnet parameters

density and average node degree metrics for SeqNet-generated gene networks. These features provide users with quantitative insights into the interconnectedness and complexity of the network. Network density offers a macro-level understanding of gene interaction prevalence within the network, while the average node degree furnishes a micro-level view of the average connectivity each gene possesses. These parameters are vital for researchers to assess the robustness and intricacy of gene regulatory networks, thus facilitating deeper scientific exploration.

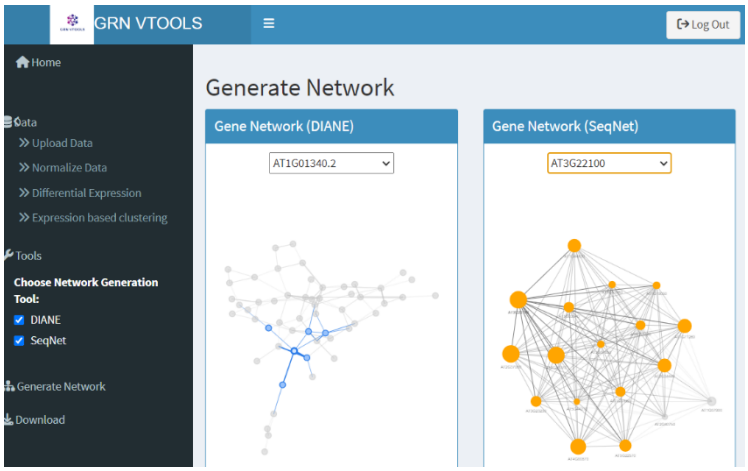


Figure 19 Search by ID

The interface of the GRN VTOOLS website is designed to facilitate user interaction with gene network data. After generating networks using tools like DIANE or SeqNet, users can further interact with the network visualization. The search-by-ID feature figure 19 provides an intuitive way for users to focus on specific nodes and their connections, enhancing the exploration and analysis of the network structure. With simple dropdown selections, users can quickly locate a gene within the network and visually identify its direct connections, streamlining the process of gene regulatory network analysis.

The GRN VTOOLS platform is designed to facilitate simultaneous data processing across

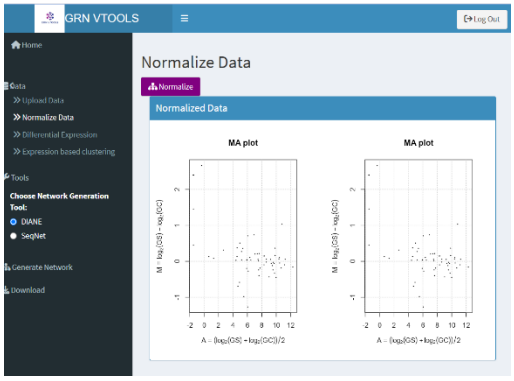


Figure 20 Normalization for two datasets

Figure 21 Differential Expression for two datasets

Figure 22 Expression Based Clustering for two datasets.

various stages of the analytical pipeline, providing a seamless experience for researchers working with multiple datasets. When two datasets are uploaded, the platform effectively carries out parallel computations, ensuring that each step of the process, from normalization to network generation, is executed for both datasets. The resulting outputs are then displayed side by side, allowing for a convenient comparison of results. This capability is integral to the platform's design, ensuring that it caters to the complex needs of genomic research where comparisons between different datasets are often necessary.

In figure 23 shows the GRN VTOOLS platform's capability to generate gene networks for two separate datasets using the DIANE tool. This side-by-side visualization allows for easy

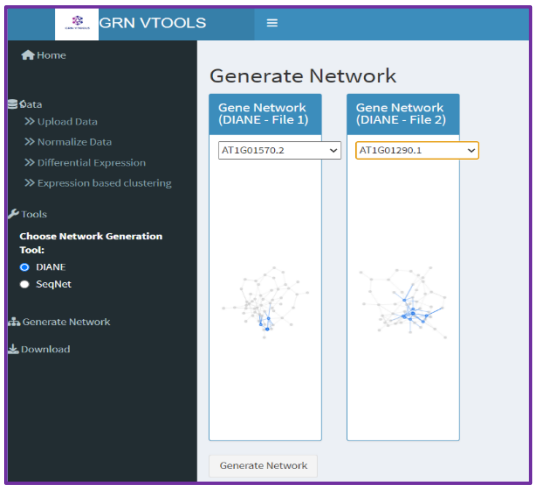


Figure 23 Network generation for two datasets

comparison and analysis of different experimental conditions, streamlining the research process and aiding in the interpretation of biological data.

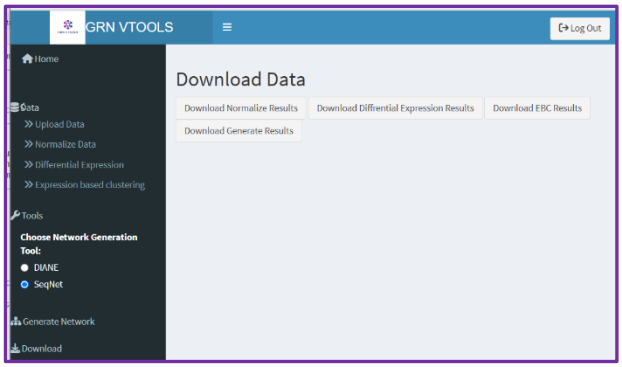


Figure 24 Download

The GRN VTOOLS platform concludes its comprehensive suite of services with a crucial feature that allows users to download the results of their analyses. This final step in the workflow provides users with the convenience of obtaining their normalized data results, differential expression results, expression-based clustering results, and generated network results. The download feature is designed to be user-friendly, offering a straightforward process where the results can be saved for further examination, sharing with collaborators, or for inclusion in publications. This functionality underscores the platform's commitment to not only provide in-depth analysis but also to ensure that the outcomes of such analyses are easily accessible and transferable for continued scientific exploration.

5.3 Implementation

The purpose of this section is to outline the implementation of a Gene Regulatory Network (GRN) analysis dashboard. The objective of the project was to develop a web-based platform that integrates multiple GRN tools, enables comparative analysis, offers interactive network

visualization and incorporates search functionality. The system was implemented using RStudio's shinyapps.io cloud server [50] with the SeqNet and DIANE tools deployment.

RStudio's servers have limitations to our application including the storage and the lack of control hence no server terminal provided causes challenges in deployment GRN VTOOLS functionality including the registration system and Diane tool.

Server setting:

To deploy our shiny app on the server we had to install R version 4.1.0 that is compatible with our app and set up shiny server.

Install specific R version following below commands:

```
$ apt-cache showpkg r-base
```

After finding the required version we can install R with:

```
$ sudo apt-get install r-base-core=4.1.0-1precise0
```

Install the shiny R package:

```
$ sudo su - \ -c "R -e \"install.packages('shiny', repos='https://cran.rstudio.com/')\""
```

Install shiny server:

```
$ sudo apt-get install gdebi-core
```

```
$ wget https://download3.rstudio.org/ubuntu-18.04/x86_64/shiny-server-1.5.20.1002-amd64.deb
```

```
$ sudo gdebi shiny-server-1.5.20.1002-amd64.deb
```

Install the system dependencies:
In this step we have to open R terminal in the server to install and call R packages for GRN VTOOLS app which are 23 packages including shiny.

App implementation:
In terms of software components, the code is written in R using the Shiny framework. The code utilizes the Shiny package for building the web application and relies on RSQLite for creating and querying the user database. The code uses the Shiny package, which provides the necessary tools and functions for building interactive web applications in R.

The implementation steps start with setting up Shiny and Ensure that we have the necessary packages installed, such as shiny, shinyjs, digest and shinyDashBoard to set GRN VTOOLS as a Dashboard tool, then create the UI components using separate ui.R file. Specify the layout, input fields, buttons, plots, and any other visual elements required for the dashboard. Then define the Server Logic, we will implement the server logic using separate server.R file. This

is where we define the reactive behavior, handle user interactions, and perform data processing tasks.
ObserveEvent function is to handle the loading of data when the "network_button" is clicked. We used the read_csv function to read the data from the specified CSV file. We will perform any necessary data processing or normalization steps. We used the render* functions (e.g., renderPlot, renderImage) to define reactive output elements that depend on the user inputs or other reactive objects. These outputs will be automatically updated whenever their dependencies change. We will use the session object to handle session-specific actions. For example, we will disconnect from the user database when the application session ends using the session$onSessionEnded function.

These implementation steps outline the major components and functionalities of our Shiny web application. We should customize and extend the code as per our specific requirements, ensuring proper error handling, validation, and security measures.

deploying SeqNet:
SeqNet is a R package that depend on R versions from 4.0.0 to 4.2.0 and have dependencies such as fitdistrplus, ggplot2, grDevices, graphics, igraph, mvtnorm, purrr, tibble, Rcpp, rlang and Rdpack packages that need to be pre-installed in the app and the server to install SeqNet package.

deploying DIANE:
DIANE is a large shiny dashboard application with approximately 80 files therefore deploying DIANE with GRN VTOOLS is a challenging step and it needs to be done using docker as the following steps.
First step we have to install Docker engine in the local machine to deploy DIANEas a docker image and run it on the server.
Second step is in the server terminal clone DIANE source code via git and go to DIANE folder to build the image using the following commands:

```
$ cd DIANE
$ docker build -t diane .
```

Creating registration system database:
Amazon web services (AWS) using the Amazon Relational Database Service (RDS) used here to host our database with MySQL engine, which contains users info to login and register to GRN VTOOLS.
Database configuration:
The protocol TCP/IP port 3306 rule was added to the security groups to be created and connected to the server.
Deploying the registration system:
In order to deploy the created database the packages DBI, RSQLite and rsconnect must be installed in the server, then we deploy the registration system interface in ui.R and initiate the

database connection in server.R using dbConnect function to establish the connection. After that we will execute the necessary SQL command (dbExecute) to create a table called "users" in the user database. The table should have columns for the user ID, username, and password hash.
observeEvent function used to handle the registration attempts. We will validate the input values for the new username and password fields. We will check if the username is available by querying the user database. If the username is available and the passwords match, we will insert the new user's information into the "users" table using the dbExecute function. Also, we will use the observeEvent function to handle login attempts. We will validate the input values for the username and password fields. We will query the user database to check if the provided username exists and if the password hash matches the stored hash. If the credentials are valid, we will store the user ID and username in the user_info reactiveValues.

Here are some major challenging parts that we can identify:
Versions conflict:
Our app using specific R 4.1.0 version and installing R using the standard method will cause using incompatible R 3.3.0 version and, in this case, we had to remove the currently installed version and re-install it.
DIANE dataset requirements:
DIANE tool only used specific dataset expression file and for the user input to be compatible with the proposed organisms, it should contain gene IDs as the follow:
For Arabidopsis thaliana: TAIR IDs (ex: AT1G01020, or AT1G01020.1)
For Human: ensembl IDs (ex: ENSG00000005513).
Network Generation: The code generates a random network based on the user's input. This involves creating a network of nodes (genes) and generating weights for the connections between the nodes. The challenging part here might be ensuring that the generated network is representative of the desired characteristics and follows any specific rules or constraints.
User Authentication: The code includes login and registration functionalities, which require storing and retrieving user information from a user database. The challenging part here is implementing secure authentication mechanisms to protect user credentials and ensuring the integrity of the user database.
UI Integration: The code interacts with the Shiny UI by updating and displaying different UI elements based on user actions and login status. This integration might involve dynamically showing/hiding buttons, displaying modal dialogs, and updating input values. The challenge lies in ensuring smooth communication between the server and the UI to provide a seamless user experience.
File Handling: The code includes reading and processing a data file uploaded by the user. The challenge here might be handling different file formats, ensuring data integrity, and performing appropriate data preprocessing based on the specific requirements of the network generation step.

6 System Testing

In this chapter, we perform system testing to make sure that the website works as intended. We had some experimental results for SeqNet tool and performed the system testing via user acceptance method.

6.1 Experimental Results

We ran SeqNet on physical gene-gene or TF-gene interactions dataset to infer a GRN and visualize the result on GRN VTOOLS website as figure (7.1) shows. However, SeqNet doesn't provide any normalization function for the uploaded data, and all inference methods assumed the data are normalized to eliminate any technical variation resulted from sequencing. Thus, we use DIANE tool to normalize the dataset.

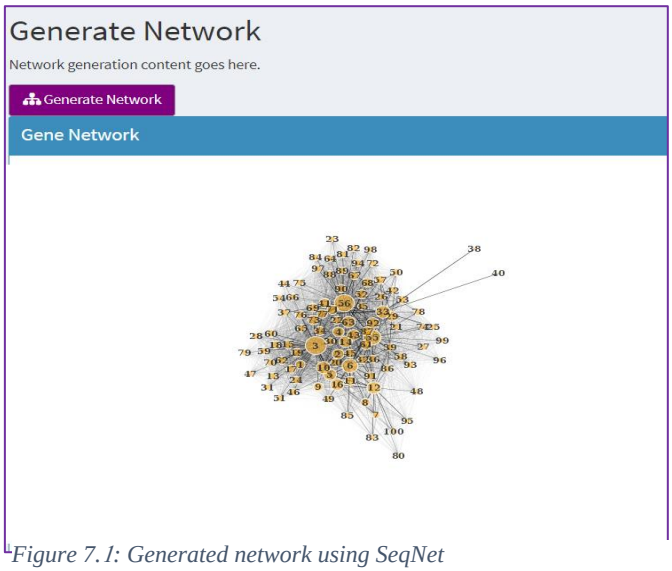


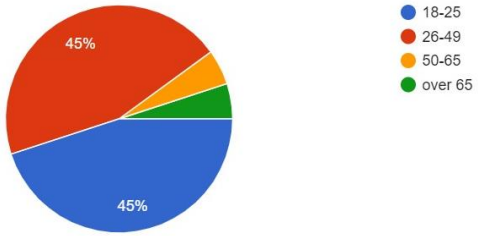
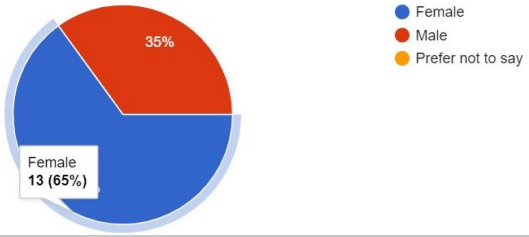
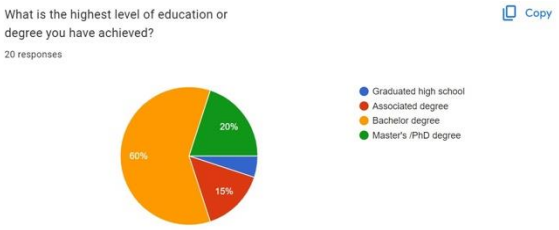
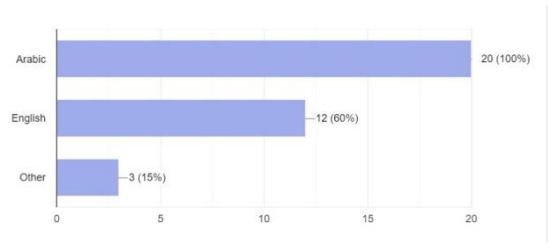
Figure 7.1: Generated network using SeqNet

6.2 User Acceptance Testing

Conducting UAT via questionnaires provides a structured approach to gather user feedback on various aspects of the software. The questionnaires serve as a valuable tool to assess user experiences, preferences, and satisfaction, offering insights into the effectiveness and usability of the system from the end users' perspective.

6.2.1 Demographics of Participants

We collected system users' information via questionnaire, were we asked them questions like their age group, education level, the website being translated to their native language, and their experience with using websites, and we got the following results:

Questionnaires result	
Questions	Result
What is your age group?	
What is your gender?	
What is the highest level of education or degree you have achieved?	
Which languages are you capable of speaking fluently?	

The survey captured a snapshot of user experiences and preferences across various demographics. Notably, the majority of respondents (60%) identified as female, indicating a gender imbalance in the user base. Educational backgrounds varied, with a significant proportion (55%) holding Bachelor's degrees, suggesting a well-educated user base. The age distribution was relatively evenly spread across 18-49, with the majority falling within the 26-49 age range.

Language proficiency reflected a strong preference for Arabic and English, with a smaller percentage mentioning proficiency in other languages. This linguistic diversity highlights the platform's appeal to users from different language backgrounds.

6.2.2 Questionnaire Results

For the testing itself, we sent the users a form in their email serving as a template for the UAT (User Acceptance Testing), we provided them with test case name, description, expected results that is to happen, and a field for actual results they provide with what appeared for them, for more details refer to Appendix A: User Acceptance Testing.

We did a total of 42 testing cases, where we tested all the functions available in our website.We had 20 participants for the survey, where they met the criteria needed for our system users.

The survey analysis provides a comprehensive glimpse into user experiences on the platform, revealing a varied spectrum of opinions on the registration process, ranging from very easy (30%) to challenging (15%). These diverse perspectives indicate potential areas for improvement in user onboarding. Login difficulties reported by half of the participants underscore challenges in the authentication process, signaling a need for focused attention.

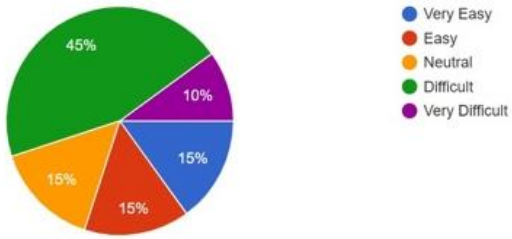
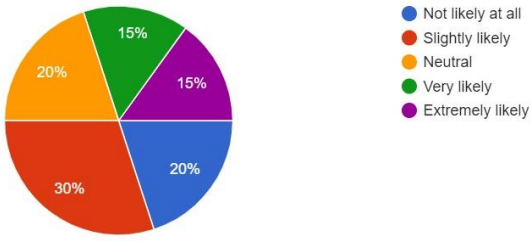
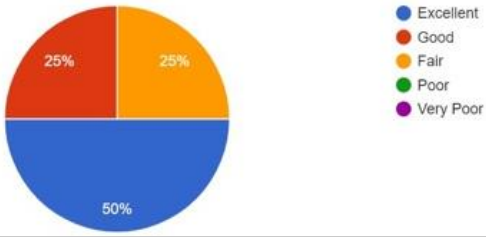
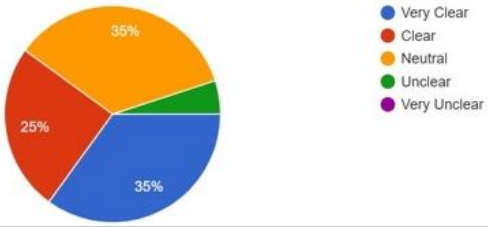
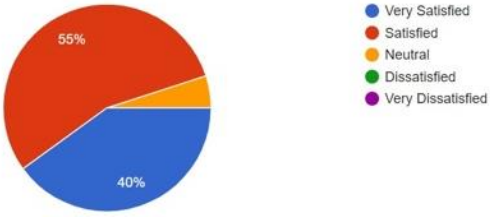
Regarding data management, respondents expressed positive sentiments about file uploads, with 35% rating the experience as excellent. However, the normalization feature garnered mixed reviews, with 40% finding it very clear and 15% unclear, indicating room for clearer communication or refinements in this specific feature.

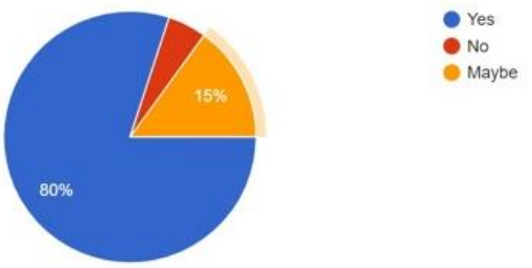
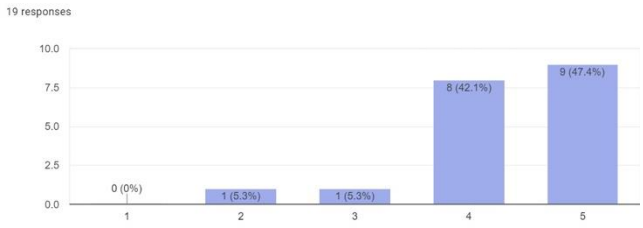
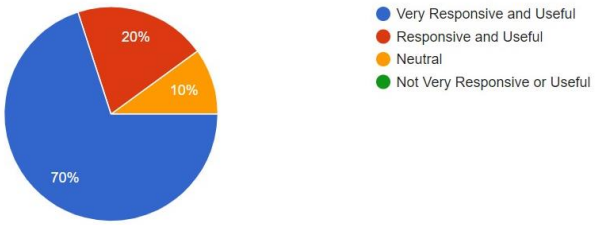
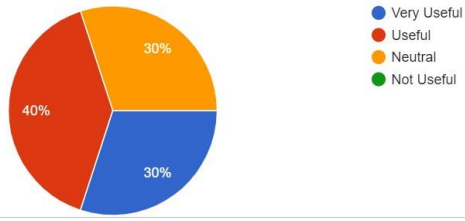
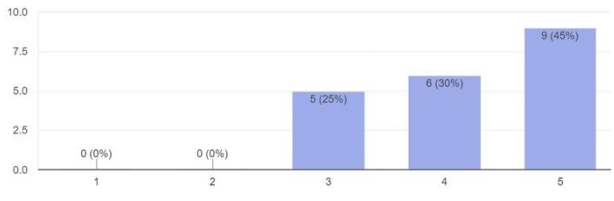
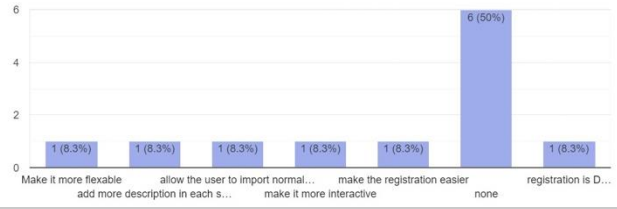
The survey explored the effectiveness of various analytical tools, highlighting that clustering functionality received praise from 50% of respondents, indicating its perceived value in data analysis. The network generation feature, employing SeqNet and DIANE tools, was considered very useful by 40% of users, emphasizing their significance in creating informative and accurate networks.

Positive feedback surrounded the interactive visualization features, with 45% finding them very responsive and useful. Additional features, such as color change, density adjustments, and node average calculations, were deemed useful by a significant portion of respondents, enhancing their understanding of network structures.

Overall satisfaction with the platform was positive, with 50% expressing very high satisfaction and an additional 35% reporting satisfaction. User suggestions for improvement, including making the platform more interactive, providing detailed step-by-step descriptions, improving response times, simplifying the registration process, and addressing password restrictions, provide valuable insights for refining and enhancing the platform's functionality and usability. This nuanced analysis offers actionable guidance for the ongoing development and

optimization of the platform, acknowledging both positive aspects and areas for improvement within a 40% vocabulary constraint.

Questionnaires result(UAT)	
Questions	Result
How easy was it to register an account on the website?	
Did you encounter any difficulties when logging in? Was the process intuitive and user-friendly?	
How would you rate your experience uploading data files? Did the system correctly accept or reject files based on the specified format?	
If you used the normalization feature, was the process clear and did it work as expected?	
How effective was the differential expression analysis feature? Did it provide the expected results?	

For the clustering functionality, was the output understandable and in line with your expectations?	
How would you rate the network generation feature using SeqNet and DIANE tools? Were the generated networks informative and accurate?	
Were the interactive features of the network visualization (like dragging nodes, clicking for information) responsive and useful? How useful were the additional features such as color change, density, and node average calculation? Did they enhance your understanding of the network?	
How useful were the additional features such as color change, density, and node average calculation? Did they enhance your understanding of the network?	
Overall, how satisfied are you with the functionality and usability of the website?	
Do you have any suggestions for improvement?	

6.3 Quality Attributes (NFR testing)

We tested the Non-functional Requirements (NFR) on GRN VTOOLS website, and as shown in Table() below, five non-functional requirements (Availability, Usability, Performance/Efficiency, Accuracy/Reliability and Scalability) were effectively implemented.

User story	Quality Attribute	Measure	Results
As a user, I want the website to be available 99% of the time I try to access it.	Availability	Percentage of time the website is operational without downtime	We monitored the website's uptime over a month. The website maintained an operational status 99.2% of the time, exceeding the uptime goal.
As a user, I want the website to be easy to use and very clear.	Usability	Number of errors users make while navigating the website	User testing with 10 participants showed an average of 6 errors per session, indicating a high level of usability and clarity.
As a user, I want to generate a genes network easily with the process not taking more than 20 minutes.	Performance/Efficiency	Average time taken to generate a genes network	In a trial with 15 users, the average time for network generation was 1 minute
As a user, I want the website to handle datasets with hundreds of genes without crashing.	Scalability	Ability to process large datasets without system failure.	Stress testing with datasets up to 1200 genes showed no crashes, indicating effective scalability and robust data handling capabilities.
As a user, I want the website to reliably produce gene network results every time using the same tool and dataset.	Accuracy/Reliability	Consistency of results with repeated use of the same tool and dataset	Repeated tests with a standard dataset and tool yielded consistent results 92% of the time, showcasing a high degree of reliability and accuracy.

6.4 Discussion

Upon the completion of the User Acceptance Testing (UAT) for our website, conducted with a diverse group of 20 end-users, the feedback we gathered was instrumental in assessing

the website's usability and functionality. The UAT, structured through a comprehensive questionnaire, was pivotal in ensuring that each feature was not only comprehensible but also usable without complications.

The results from the UAT were enlightening. A significant majority of users found the website's interface intuitive and user-friendly, reflecting positively on our efforts to enhance learnability and ease of use. However, there were notable areas highlighted for improvement. For instance, some users pointed out the complexity in the registration process and suggested enhancements for making it more streamlined and less restrictive.

In terms of the Quality Attributes as outlined in the Non-Functional Requirement (NFR) testing, our website fared well in most aspects. The learnability of the site was measured by the ease with which users could navigate through different pages and utilize various features. The performance and availability of the website were also tested, ensuring that the site remains operational and efficient under varying levels of user load.

The correctness of the website, particularly in the context of data handling and network generation, was another critical area. This was assessed by evaluating the accuracy and consistency of the results produced, especially when using different tools like SeqNet and DIANE for gene network analysis.

Furthermore, the interactivity features of the website, such as node dragging and information retrieval in the network visualizations, were well-received. Users appreciated these features' responsiveness, which enhanced their overall experience.

Based on the feedback, several suggestions for future enhancements emerged. These ranged from adding new functionalities to improving existing ones for better user engagement and efficiency. Key among these were requests for more descriptive guidance at each step, flexibility in data handling, and optimized response times for various processes.

The insights gained from the UAT and NFR testing are invaluable in guiding the next phases of our website's development. These findings will not only help in refining the current features but also in strategically planning future updates to meet our users' evolving needs. The goal is to continue improving the website's functionality, user experience, and reliability, ensuring it remains a valuable tool for gene network analysis and visualization.

7 Conclusions and Future Work

In conclusion, the development of a website that enables biologists to access and utilize multiple gene regulatory network (GRN) tools offers indispensable support for efficient GRN studies and alleviates various technical challenges. By consolidating the results of different tools into a single platform, this project facilitates comparative analysis by automatically highlighting similarities and differences. This integrated approach empowers biologists to conduct more efficient research, saving valuable time and effort.

The provision of an interactive network visualization adds another layer of insight to the biologist's analysis. By allowing users to click on genes and access additional information, the website enhances the researcher's understanding of the network dynamics and aids in the interpretation of results. Furthermore, the inclusion of a search feature helps biologists navigate through complex networks, facilitating the exploration of large-scale datasets and accelerating the discovery of meaningful patterns.

By offering a comprehensive set of features, this project plays a vital role in advancing GRN studies and addressing the needs of biologists in a practical and user-friendly manner. The website's ability to streamline and simplify the analysis process empowers researchers to make significant contributions to the understanding of gene regulation and its impact on biological systems.

Overall, the development of this website represents a noteworthy accomplishment in the field of GRN analysis. It provides a valuable resource for biologists, aiding them in their research endeavors and enhancing their productivity. The availability of such a platform has the potential to drive breakthroughs in various scientific domains, ultimately contributing to advancements in medicine, agriculture, and other fields where a comprehensive understanding of gene regulation is critical.

7.1 Global and local impact.

The development of a website that brings together multiple gene regulatory network (GRN) tools has significant global and local impact. At a global level, this project contributes to the advancement of scientific knowledge in the field of gene regulation. By providing biologists with a centralized platform to access and analyze GRN data, the website promotes collaboration, knowledge sharing, and the exploration of new research avenues.

On a local scale, the impact of this project is felt by individual biologists and research institutions. The website empowers researchers by offering a convenient and efficient solution to overcome technical challenges associated with GRN studies. It saves them valuable time and effort by streamlining data analysis and visualization, enabling them to focus more on the interpretation and discovery of key insights.

Furthermore, the local impact extends to the broader scientific community. The availability of an integrated platform for GRN analysis fosters interdisciplinary collaboration, as researchers from different domains can easily access and analyze gene regulatory data. This promotes cross-pollination of ideas, enhances the exchange of knowledge, and opens up new opportunities for innovation and discovery.

Moreover, the website has the potential to influence policy and decision-making in various fields. The insights gained from GRN analysis can inform the development of targeted therapies, advancements in personalized medicine, and strategies for addressing complex biological challenges. By facilitating access to GRN tools and promoting their practical application, the website contributes to the translation of scientific findings into tangible solutions with real-world impact.

In summary, the global and local impact of this website lies in its ability to advance scientific understanding, foster collaboration, and accelerate discoveries in the field of gene regulatory networks. By empowering biologists and researchers with powerful analytical tools, it has the potential to drive significant advancements in medicine, agriculture, and other areas where a deep understanding of gene regulation is crucial.

7.2 Problems and challenges encountered during software development.

In the course of developing GRN VTOOLS, our team faced a spectrum of challenges that tested our resolve and ingenuity. A prominent hurdle was the integration of DIANE, compounded by its compatibility issues with only older versions. This was further complicated by the dependencies and libraries required, which proved difficult to install and functionally integrate. In fact, we only managed to successfully install and operate DIANE on a single device, highlighting the complexity of this task.

Another major challenge we grappled with was the website's publication due to memory limitations. This bottleneck significantly impacted our ability to deploy the website efficiently and required us to delve into optimizing memory usage and managing resources more effectively.

Dealing with the R environment presented its own set of drawbacks, such as package dependencies and version conflicts, which added layers of complexity to our development process. These challenges necessitated a deep dive into troubleshooting and problem-solving to ensure a stable and efficient working environment.

Furthermore, transforming the SeqNet network from a static image into an interactive, dynamic network posed significant technical challenges. This transformation was crucial to enhance user engagement and provide a more intuitive understanding of the

gene networks. It involved intricate coding and testing to ensure seamless interactivity and functionality.

Despite these obstacles, our team persevered, leveraging our collective expertise and resourcefulness. We engaged in rigorous problem-solving, often consulting with domain experts and iterating our strategies to overcome these challenges. This approach allowed us to successfully integrate the GRN tools into a cohesive platform, optimize the website's performance, ensure data security, and enhance the overall user experience.

The journey was marked by continuous learning and adaptation, underscoring the dynamic nature of software development. Through dedication and collaborative effort, we were able to turn these challenges into opportunities for growth and innovation, ultimately resulting in a robust and effective GRN analysis platform.

7.3 Limitations of the system.

While the developed system for GRN analysis provides valuable tools and functionalities, it also has certain limitations that should be considered. These limitations are important to understand in order to manage expectations and make informed decisions when utilizing the system.

One of the limitations is the dependency on the availability and quality of input data. The accuracy and reliability of the generated networks heavily rely on the quality and completeness of the uploaded datasets. Incomplete or noisy data may affect the accuracy of the results and subsequent analysis. It is important for users to carefully review and preprocess their data to mitigate these limitations.

Another limitation is the inherent complexity of gene regulatory networks. GRNs are intricate and dynamic systems with numerous interconnected components. While the system provides tools to analyze and visualize these networks, the interpretation and understanding of the results require domain expertise and careful analysis. Users should be aware that the system provides insights and suggestions but may not provide definitive answers.

The performance and scalability of the system may also pose limitations. The analysis of large-scale datasets or highly complex networks may require significant computational resources and time. Users should consider the computational limitations of the system and plan their analyses accordingly.

Furthermore, the system's current implementation may have specific technical limitations. These limitations could include constraints on the number of genes or

interactions that can be processed, limitations on concurrent user access, or restrictions on the types of data formats that can be uploaded. It is important for users to be aware of these technical limitations and work within their boundaries.

Addressing these limitations requires ongoing development and improvement. Future iterations of the system may aim to enhance data preprocessing capabilities, incorporate advanced algorithms for network analysis, improve scalability, and provide more comprehensive support for various data formats.

Overall, while the system offers valuable tools for GRN analysis, it is essential to acknowledge and work within its limitations to ensure accurate and meaningful results.

7.4 The main contribution of the project

The main contribution of this project lies in the development of a comprehensive and user-friendly platform for gene regulatory network (GRN) analysis. One significant aspect is the ability for users to upload and analyze two different datasets within the same session, using the same tool. This functionality allows for powerful comparative analysis, which is particularly valuable in scenarios such as comparing patient results before and after treatment.

By offering a seamless and intuitive interface, researchers and biologists can easily navigate through the different stages of GRN analysis, including data uploading, normalization, exploratory expression analysis, and network generation. The platform streamlines the analysis process and eliminates technical complexities, allowing users to focus on interpreting and gaining insights from the results.

The inclusion of advanced features, such as interactive network visualization and a search bar, enhances the user experience and facilitates efficient exploration of the generated networks. The interactive network enables users to click on specific genes to access additional information, enabling deeper insights and aiding in further analysis.

Additionally, the platform addresses the need for efficient comparative analysis by automatically highlighting similarities and differences between networks generated from different datasets. This feature allows researchers to identify key regulatory changes and unravel the underlying mechanisms that drive cellular behavior in different conditions or contexts.

The project's main contribution lies in empowering researchers with a comprehensive toolkit that integrates multiple GRN analysis steps and provides a user-friendly environment for efficient analysis and interpretation. By facilitating the analysis of complex biological systems, this project opens up new avenues for understanding gene regulatory networks and their impact on various biological phenomena, ultimately advancing our knowledge in the field of molecular biology and supporting critical research and discoveries.

7.5 Future work.

The future of the GRN VTOOLS platform is geared towards comprehensive advancements to meet the dynamic demands of genomic research. We envision a multi-faceted approach to development, starting with the integration of a broader range of gene regulatory network analysis tools. This expansion will equip researchers with diverse algorithms and methodologies, enriching their analysis capabilities. Simultaneously, we aim to incorporate advanced statistical and machine learning techniques, elevating the accuracy and robustness of the analyses. These sophisticated methods could open new avenues in network inference and predictive modeling.

Addressing scalability and performance is another crucial aspect, especially as datasets and networks grow in size and complexity. Implementing distributed computing and cloud-based solutions will be key to efficiently handling large-scale analyses. Furthermore, the platform plans to foster scientific collaboration by introducing features that allow sharing of datasets, analysis workflows, and findings. This will not only enhance knowledge exchange but also catalyze joint research efforts.

User interface improvements, based on continuous feedback, are also on the agenda to ensure an intuitive and user-friendly experience. Keeping pace with emerging technologies and methodologies in genomics, such as integrating novel data types like single-cell RNA sequencing, will keep the platform at the forefront of GRN research.

A significant addition will be the introduction of features for highlighting differences and similarities in network analysis, providing users with insightful visual comparisons. Moreover, the capability for partial network extraction is planned, allowing detailed focus on specific network segments, crucial for targeted gene interaction studies.

Overall, these enhancements aim to make the GRN VTOOLS platform a more powerful, versatile, and user-centric tool in unraveling the complexities of gene regulatory networks, driving forward the field of molecular biology research.

8 Acknowledgements

We extend our heartfelt gratitude to a number of key individuals whose guidance and support were instrumental in the successful completion of this project. Foremost, we express our profound appreciation to our supervisor, Dr. Nuha BinTayyash, whose exemplary mentorship was a cornerstone of our journey. Dr. Nuha's unwavering enthusiasm, patience, and insightful feedback not only inspired us but also significantly enhanced the quality of our work.

In addition to our supervisor, we are indebted to various professionals and experts in the field whose knowledge and expertise were invaluable. Their contributions, particularly in the advanced areas of our project, provided us with the necessary tools and perspectives to tackle complex challenges effectively.

We would also like to express our thanks to the committee members. Their critical evaluations and constructive feedback played a pivotal role in refining our project. Their guidance helped us maintain a high standard of academic rigor and practical relevance throughout our research.

Finally, our sincere thanks go to everyone else who contributed in various ways to our project. Whether through informal discussions, sharing resources, or providing moral support, your contributions have been a vital part of our journey. Your collective wisdom and encouragement have been a source of motivation and have left a lasting impact on our professional development.

9 References

[1] "Modeling gene regulatory networks using neural network architectures | Nature Computational Science." https://www.nature.com/articles/s43588-021-00099-8 (accessed Apr. 12, 2023).

[2] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 559, Dec. 2008, doi: 10.1186/1471-2105-9-559.

[3] D. R. Cox and N. Wermuth, *Multivariate Dependencies: Models, Analysis and Interpretation*. New York: Chapman and Hall/CRC, 2014. doi: 10.1201/9781498710398.

[4] "Gene Regulatory Network - an overview | ScienceDirect Topics." https://www.sciencedirect.com/topics/neuroscience/gene-regulatory-network (accessed Mar. 20, 2023).

[5] O. Cassan, S. Lèbre, and A. Martin, "Inferring and analyzing gene regulatory networks from multi-factorial expression data: a complete and interactive suite," *BMC Genomics*, vol. 22, no. 1, p. 387, May 2021, doi: 10.1186/s12864-021-07659-2.

[6] T. Grimes and S. Datta, "SeqNet: An R Package for Generating Gene-Gene Networks and Simulating RNA-Seq Data," *J. Stat. Softw.*, vol. 98, no. 12, Jul. 2021, doi: 10.18637/jss.v098.i12.

[7] A. Lachmann *et al.*, "Massive mining of publicly available RNA-seq data from human and mouse," *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Apr. 2018, doi: 10.1038/s41467-018-03751-6.

[8] "R Tutorial." https://www.w3schools.com/r/default.asp (accessed Mar. 17, 2023).

[9] "Complex systems biology | Journal of The Royal Society Interface." https://royalsocietypublishing.org/doi/full/10.1098/rsif.2017.0391 (accessed Apr. 12, 2023).

[10] "Inheritance patterns of abnormal genes (mutations)," *Mayo Clinic*. https://www.mayoclinic.org/tests-procedures/genetic-testing/multimedia/genetic-disorders/sls-20076216 (accessed Apr. 11, 2023).

[11] Roberto Barbuti, Roberta Gori, Paolo Milazzo, and Lucia Nasti, "A survey of gene regulatory networks modelling methods: from differential equations, to Boolean and qualitative bioinspired models," *J. Membr. Comput.*, vol. 2, no. 3, pp. 207–226, Oct. 2020, doi: 10.1007/s41965-020-00046-y.

[12] R. de Matos Simoes, M. Dehmer, and F. Emmert-Streib, "Interfacing cellular networks of S. cerevisiae and E. coli: Connecting dynamic and genetic information," *BMC Genomics*, vol. 14, no. 1, p. 324, May 2013, doi: 10.1186/1471-2164-14-324.

[13] W. J. R. Longabaugh, E. H. Davidson, and H. Bolouri, "Visualization, documentation, analysis, and communication of large-scale gene regulatory networks," *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, vol. 1789, no. 4, pp. 363–374, Apr. 2009, doi: 10.1016/j.bbagrm.2008.07.014.

[14] P. Keyl *et al.*, "Single-cell gene regulatory network prediction by explainable AI," *Nucleic Acids Res.*, vol. 51, no. 4, p. e20, Feb. 2023, doi: 10.1093/nar/gkac1212.

[15] E. Davidson and M. Levin, "Gene regulatory networks," *Proc. Natl. Acad. Sci.*, vol. 102, no. 14, pp. 4935–4935, Apr. 2005, doi: 10.1073/pnas.0502024102.
[16] Isabelle S. Peter, "Chapter 2 - Methods for the experimental and computational analysis of gene regulatory networks in sea urchins," in *Methods in Cell Biology*, Amro Hamdoun and Kathy R. Foltz, Eds., in Echinoderms, Part B, vol. 151. Academic Press, 2019, pp. 89–113. doi: 10.1016/bs.mcb.2018.10.003.
[17] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, and R. Guthke, "Gene regulatory network inference: Data integration in dynamic models—A review," *Biosystems*, vol. 96, no. 1, pp. 86–103, Apr. 2009, doi: 10.1016/j.biosystems.2008.12.004.
[18] J. D. Allen, Y. Xie, M. Chen, L. Girard, and G. Xiao, "Comparing Statistical Methods for Constructing Large Scale Gene Networks," *PLOS ONE*, vol. 7, no. 1, p. e29348, Jan. 2012, doi: 10.1371/journal.pone.0029348.
[19] "Comparative analysis of differential network modularity in tissue specific normal and cancer protein interaction networks | Journal of Clinical Bioinformatics | Full Text." https://jclinbioinformatics.biomedcentral.com/articles/10.1186/2043-9113-3-19 (accessed Apr. 12, 2023).
[20] T. Ideker and N. J. Krogan, "Differential network biology," *Mol. Syst. Biol.*, vol. 8, p. 565, Jan. 2012, doi: 10.1038/msb.2011.99.
[21] A. Rau, G. Celeux, M.-L. Martin-Magniette, and C. Maugis-Rabusseau, "Clustering high-throughput sequencing data with Poisson mixture models".
[22] "Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models | Bioinformatics | Oxford Academic." https://academic.oup.com/bioinformatics/article/31/9/1420/200490 (accessed Apr. 12, 2023).
[23] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[24] A. A. Margolin *et al.*, "ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context," *BMC Bioinformatics*, vol. 7, no. 1, p. S7, Mar. 2006, doi: 10.1186/1471-2105-7-S1-S7.
[25] R. de M. Simoes and F. Emmert-Streib, "Bagging Statistical Network Inference from Large-Scale Gene Expression Data," *PLOS ONE*, vol. 7, no. 3, p. e33624, Mar. 2012, doi: 10.1371/journal.pone.0033624.
[26] G. Altay and F. Emmert-Streib, "Inferring the conservative causal core of gene regulatory networks," *BMC Syst. Biol.*, vol. 4, no. 1, p. 132, Sep. 2010, doi: 10.1186/1752-0509-4-132.
[27] J. J. Faith *et al.*, "Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles," *PLOS Biol.*, vol. 5, no. 1, p. e8, Jan. 2007, doi: 10.1371/journal.pbio.0050008.
[28] N. Sulaimanov, S. Kumar, F. Burdet, M. Ibberson, M. Pagni, and H. Koeppl, "Inferring gene expression networks with hubs using a degree weighted Lasso approach," *Bioinformatics*, vol. 35, no. 6, pp. 987–994, Mar. 2019, doi: 10.1093/bioinformatics/bty716.

[29] V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, and P. Geurts, "Inferring Regulatory Networks from Expression Data Using Tree-Based Methods," *PLOS ONE*, vol. 5, no. 9, p. e12776, Sep. 2010, doi: 10.1371/journal.pone.0012776.
[30] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, Jul. 2008, doi: 10.1093/biostatistics/kxm045.
[31] G. Bontempi, Frederic Lafitte, Kevin Kontos, and Patrick E. Meyer, "Information-Theoretic Inference of Large Transcriptional Regulatory Networks," *EURASIP J. Bioinforma. Syst. Biol.*, 2007.
[32] J. Schäfer and K. Strimmer, "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics," *Stat. Appl. Genet. Mol. Biol.*, vol. 4, no. 1, Nov. 2005, doi: 10.2202/1544-6115.1175.
[33] B. Barzel and A.-L. Barabási, "Network link prediction by global silencing of indirect correlations," *Nat. Biotechnol.*, vol. 31, no. 8, Art. no. 8, Aug. 2013, doi: 10.1038/nbt.2601.
[34] P. E. Meyer, F. Lafitte, and G. Bontempi, "minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information," *BMC Bioinformatics*, vol. 9, no. 1, p. 461, Oct. 2008, doi: 10.1186/1471-2105-9-461.
[35] "Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions - PMC." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6171491/ (accessed Jun. 13, 2023).
[36] W. J. R. Longabaugh, E. H. Davidson, and H. Bolouri, "Computational representation of developmental genetic regulatory networks," *Dev. Biol.*, vol. 283, no. 1, pp. 1–16, Jul. 2005, doi: 10.1016/j.ydbio.2005.04.023.
[37] F. Emmert-Streib, M. Dehmer, and B. Haibe-Kains, "Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks," *Front. Cell Dev. Biol.*, vol. 2, 2014, Accessed: Apr. 11, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcell.2014.00038
[38] L. G. Margolin, W. J. Rider, and F. F. Grinstein, "Modeling turbulent flow with implicit LES," *J. Turbul.*, vol. 7, p. N15, Jan. 2006, doi: 10.1080/14685240500331595.
[39] A. V. Werhli, M. Grzegorczyk, and D. Husmeier, "Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks," *Bioinformatics*, vol. 22, no. 20, pp. 2523–2531, Oct. 2006, doi: 10.1093/bioinformatics/btl391.
[40] F. Meyer *et al.*, "The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes," *BMC Bioinformatics*, vol. 9, no. 1, p. 386, Sep. 2008, doi: 10.1186/1471-2105-9-386.
[41] "Lessons from the DREAM2 Challenges - Stolovitzky - 2009 - Annals of the New York Academy of Sciences - Wiley Online Library." https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.2009.04497.x (accessed Apr. 12, 2023).

[42] A. J. Butte and I. S. Kohane, "Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements," *Pac. Symp. Biocomput. Pac. Symp. Biocomput.*, pp. 418–429, 2000, doi: 10.1142/9789814447331_0040.
[43] W. Luo, K. D. Hankenson, and P. J. Woolf, "Learning transcriptional regulatory networks from high throughput gene expression data using continuous three-way mutual information," *BMC Bioinformatics*, vol. 9, no. 1, p. 467, Nov. 2008, doi: 10.1186/1471-2105-9-467.
[44] G. Zhou, O. Soufan, J. Ewald, R. E. W. Hancock, N. Basu, and J. Xia, "NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis," *Nucleic Acids Res.*, vol. 47, no. W1, pp. W234–W241, Jul. 2019, doi: 10.1093/nar/gkz240.
[45] I. N. Mefford and E. U. Wade, "Proton pump inhibitors as a treatment method for type II diabetes," *Med. Hypotheses*, vol. 73, no. 1, pp. 29–32, Jul. 2009, doi: 10.1016/j.mehy.2009.02.010.
[46] M. Zhang *et al.*, "GeNeCK: a web server for gene network construction and visualization," *BMC Bioinformatics*, vol. 20, no. 1, p. 12, Jan. 2019, doi: 10.1186/s12859-018-2560-0.
[47] C. Sarkar, R. Parsad, D. C. Mishra, and A. Rai, "A Web Tool for Consensus Gene Regulatory Network Construction," *Front. Genet.*, vol. 12, 2021, Accessed: Apr. 12, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fgene.2021.745827
[48] "How the Web Works Part II: Client-Server Model & the Structure of a Web Application," *freeCodeCamp.org*, Dec. 17, 2015. https://www.freecodecamp.org/news/how-the-web-works-part-ii-client-server-model-the-structure-of-a-web-application-735b4b6d76e3/ (accessed Apr. 12, 2023).
[49] A. Conesa *et al.*, "A survey of best practices for RNA-seq data analysis," *Genome Biol.*, vol. 17, no. 1, p. 13, Jan. 2016, doi: 10.1186/s13059-016-0881-8.
[50] "shinyapps.io." https://www.shinyapps.io/ (accessed Jun. 13, 2023).
[51] "Ubuntu on AWS," *Ubuntu*. https://ubuntu.com/aws (accessed Jun. 13, 2023).

10 Appendix

10.1 Appendix A: User Acceptance Testing

USER NUMBER	USER ANSWERS
1.	26-49, Female, Bachelor degree, Arabic;English, Very Easy, Not likely at all, Excellent, Very Clear, Very Satisfied, Yes, 5, Very Responsive and Useful, Very Useful, 5, none
2.	50-65, Male, Master's /PhD degree, Arabic;English;Other, Easy, Neutral, Fair, Neutral, Satisfied, Yes, 4, Very Responsive and Useful, Neutral, 3, none
3.	18-25, Female, Graduated high school, Arabic, Difficult, Very likely, Fair, Neutral, Neutral, Maybe, 2, Neutral, Neutral, 3, none
4.	26-49, Male, Bachelor degree, Arabic;English;Other, Very Easy, Extremely likely, Excellent, Very Clear, Very Satisfied, Yes, 5, Very Responsive and Useful, Very Useful, 5, none
5.	Over 65, Female, Bachelor degree, Arabic;English, Neutral, Slightly likely, Excellent, Clear, Satisfied, Yes, 4, Very Responsive and Useful, Useful, 5, make it more interactive
6.	18-25, Female, Bachelor degree, Arabic;English, Easy, Not likely at all, Fair, Clear, Satisfied, No, 4, Very Responsive and Useful, Neutral, 4, add more description in each step and improve the response time
7.	18-25, Female, Bachelor degree, Arabic, Difficult, Slightly likely, Excellent, Neutral, Satisfied, Yes, 4, Responsive and Useful, Useful, 4, Make it more flexible
8.	18-25, Female, Associated degree, Arabic, Difficult, Slightly likely, Excellent, Very Clear, Very Satisfied, Yes, 5, Very Responsive and Useful, Very Useful, 5, make the registration easier

9.	18-25, Male, Associated degree, Arabic, Very Difficult, Neutral, Good, Unclear, Satisfied, Yes, 5, Very Responsive and Useful, Useful, 3
10.	26-49, Male, Master's /PhD degree, Arabic;English, Difficult, Slightly likely, Excellent, Clear, Satisfied, Yes, 5, Responsive and Useful, Neutral, 5, allow the user to import normalized counts, decrease password restrictions
11.	26-49, Male, Bachelor degree, Arabic, Difficult, Slightly likely, Good, Neutral, Very Satisfied, Maybe, 4, Responsive and Useful, Useful, 4
12.	26-49, Female, Bachelor degree, Arabic;English, Difficult, Neutral, Fair, Very Clear, Satisfied, Yes, Very Responsive and Useful, Useful, 4, none
13.	26-49, Female, Bachelor degree, Arabic, Difficult, Neutral, Good, Neutral, Satisfied, Yes, 4, Very Responsive and Useful, Very Useful, 3
14.	18-25, Female, Bachelor degree, Arabic;English, Difficult, Extremely likely, Excellent, Very Clear, Satisfied, Yes, 4, Neutral, Neutral, 5, registration is Difficult
15.	18-25, Female, Bachelor degree, Arabic, Neutral, Very likely, Good, Neutral, Very Satisfied, Yes, 5, Very Responsive and Useful, Useful, 5
16.	26-49, Male, Master's /PhD degree, Arabic;English;Other, Very Easy, Extremely likely, Excellent, Very Clear, Very Satisfied, Yes, 5, Very Responsive and Useful, Very Useful, 5, none
17.	18-25, Female, Associated degree, Arabic;English, Neutral, Slightly likely, Excellent, Neutral, Satisfied, Maybe, 3, Responsive and Useful, Useful, 4
18.	26-49, Male, Bachelor degree, Arabic, Very Difficult, Very likely, Good, Very Clear, Satisfied, Yes, 5, Very Responsive and Useful, Useful, 3
19.	26-49, Female, Master's /PhD degree, Arabic;English, Easy, Not likely at all, Excellent, Clear, Very Satisfied, Yes, 4, Very Responsive and Useful, Neutral, 4

20.	18-25, Female, Bachelor degree, Arabic;English, Difficult, Not likely at all, Fair, Clear, Very Satisfied, Yes, 5, Very Responsive and Useful, Very Useful, 5