**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Oh My Genes Documentation**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**By Machinda Zvazvinhu**

**(Dean and Golden)**

**March 27 2019**

**Table of Contents**

[**1.** **Introduction** 2](#_Toc4591265)

[**1.1.** **Purpose** 2](#_Toc4591266)

[**1.2.** **Overview** 2](#_Toc4591267)

[**1.3.** **User Characteristics** 2](#_Toc4591268)

[**1.4. Terminologies & Abbreviations** 2](#_Toc4591269)

[**2. Functional Requirements** 2](#_Toc4591270)

[**2.1. Input** 2](#_Toc4591271)

[**2.2. Output (Data Analyzing)** 2](#_Toc4591272)

[**2.2.1. Table Analyzing** 2](#_Toc4591273)

[**2.2.2. Plot Analyzing** 3](#_Toc4591274)

[**3. Non-functional Requirements** 4](#_Toc4591275)

[**3.1. Response Time** 4](#_Toc4591276)

[**3.2. Aesthetic Aspects** 4](#_Toc4591277)

[**3.3. Confidentiality Policy** 4](#_Toc4591278)

[**4. Constraints** 5](#_Toc4591279)

[**4.1. Browser Compatibility** 5](#_Toc4591280)

[**4.2. Space Complexity** 5](#_Toc4591281)

[**4.3. Budget** 5](#_Toc4591282)

[**4.4. System Downtime** 5](#_Toc4591283)

[**5. Change Cases** 6](#_Toc4591284)

[**6. Milestones** 7](#_Toc4591285)

[**7. References** 8](#_Toc4591286)

[**8. Group Information** 9](#_Toc4591287)

[**8.1. Group Name** 9](#_Toc4591288)

[**8.2. Group Members** 9](#_Toc4591289)

1. **Introduction**

This document is intended to specify a set of requirements for the project OMG, which means “Oh My Genes”, a web application for identifying genes’ expression differential.

* 1. **Purpose**

The main purpose is to help biologists analyze their experiment data by identifying differentially expressed genes from gene expression file containing two sample cells, automatically and quickly.

* 1. **Overview**

The web application is based on Flask framework for analyzing the difference between genes and it has a simple interface with a single button [Upload and Go] that accepts a gene expression in a text file with the correct format.

* 1. **User Characteristics**

The users are devoted biologists as well as researchers studying in the field of expression of genes and the website is maintained by technicians with basic knowledge on flask and python.

* 1. **Terminologies & Abbreviations**

Control sample - a cell sample prepared in its normal condition, without any defects.

Treatment sample - a cell sample treated by special chemicals, or in which some genes are altered.

Differentially expressed genes - the genes which have signiﬁcantly different expression levels between two samples.

Up-regulation - a gene is said to be up-regulated if it has higher expression in treatment than in control.

LogFC - log fold change of gene expression. log\_2 [T/C], where T is the gene expression level from a treatment sample, while C is the gene expression level from a control sample.

OMG – oh my genes, this is the name of the project.

**2. Functional Requirements**

**2.1. Input**

A valid submitted gene expression ﬁle has the following format:

• It is a TAB-delimited, plain text ﬁle with three columns;

• The ﬁle contains an optional head line, followed by each gene’s expression in a control sample (e.g., Control Sample) and in a treatment sample (e.g., Knock Out Sample).

The ﬁle pattern is as follows:

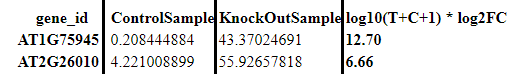
|  |  |  |
| --- | --- | --- |
| Gene \_id | Control Sample | Knock Out Sample |
| AT1G01010 | 1.198558083 | 2.036161827 |
| AT1G01020 | 13.75736234 | 13.370796 |
| AT1G01030 | 0.833779536 | 0.203616183 |
| AT1G01040 | 9.58846466 | 7.126566394 |
| AT1G01046 | 0 | 0 |
| AT1G01050 | 23.81482799 | 21.10821094 |
| AT1G01060 | 0.625334652 | 1.221697096 |
| AT1G01070 | 1.719670292 | 0.950208853 |
| AT1G01080 | 28.34850421 | 25.24840665 |
| AT1G01090 | 58.26034505 | 42.96301455 |
| AT1G01100 | 1066.508249 | 1308.030358 |
| AT1G01110 | 2.709783491 | 1.425313279 |

**2.2. Output (Data Analyzing)**

The web application displays a **table** and a **scatter plot** when given a gene expression file.

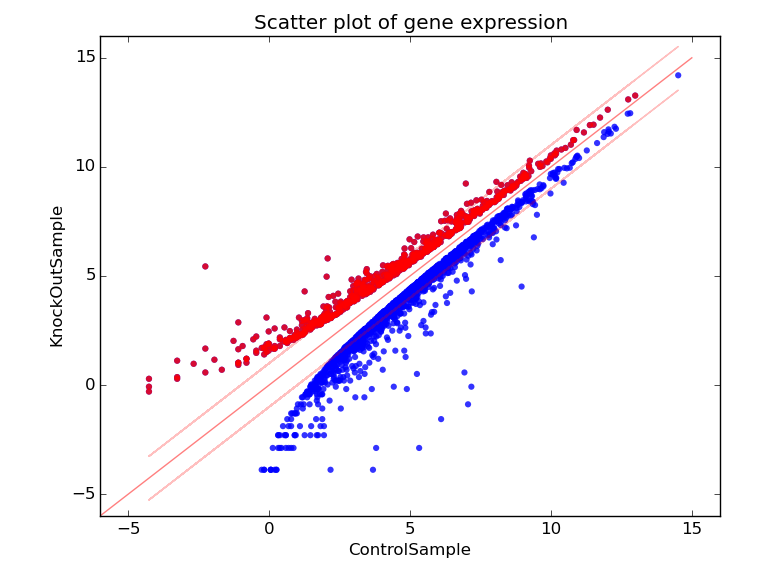
**2.2.1. Table Analyzing**

The table contains a list of differentially expressed genes in the following format.



**2.2.2. Plot Analyzing**

The scatter plot displays differentially expressed genes, with the X-axis and the Y-axis as the Control and the KnockOut respectively. The up-regulated genes are shown in red dots whereas down-regulated genes are shown in blue.



**3. Non-functional Requirements**

This section covers the non-functional requirements of the app for a preferable using experience.

**3.1. Response Time**

The web application is designed to serve biologists from the world over meaning the server will have to stand a heavy load of requests. This may weigh heavy on the system causing it to go down for some time or disable response to clients. Time complexity must be taken into consideration when the user has a huge amount of data to analyse. For good optimized experience, our algorithm should be sufﬁciently effective to deal with this situation. For most user cases we have decided the response time should be 5 seconds for a favourable experience in the worst-case scenario.

**3.2. Aesthetic Aspects**

The interface should be as simple as possible since the web application’s function is to solve biological analysis. The graph should reflect the distinction between the control and treatment samples on the graph.

**3.3. Confidentiality Policy**

As a public web application, the security of data must be a signiﬁcant part to be considered. For biologists, every experiment result is confidential due to its sensitivity, therefore when users are using the application, their data must be used and stored securely. In other words, the web application must maintain the experiment data’s reliability, integrity and secrecy and the experimental data will not be accessed or shared without the agreement of the owner.

There are functions we wish to improve or add in order to improve the general user all round experience. For starters the system can set the threshold log\_2FC automatically by itself. Also adding the capability to add multiple tables at once and then getting graph plotted for the different tables rather than having to enter one file at once. There is need also for a login and registration and administrative functions so that there is a hierarchy to show who accesses what and who can change what, this is so information integrity and security is maintained. Moreover, a database functionality would help so that our information is not scattered all over the web and previously accessed files can easily be retained without too much trouble. We would also like to be able to see the gene \_id when we move the cursor over a scatter point, with the ability to click and be able to view the gene information. The security issue in this case though takes precedence over everything else since this information might be too sensitive to be left unsecure.

**4. Constraints**

The application is designed to serve a large number of biologists all over the world, therefore, some constraints have to be taken into account.

**4.1. Browser Compatibility**

Since the application has a very coverage, a cross platform compatibility must be considered. Thus, the application should be accessible through Firefox, Chrome, and Safari.

**4.2. Space Complexity**

Space complexity should be considered in terms of the size of the entire web application as well as the space occupied when the application starts analyzing data. It must have a total size of at most 1GBT and when the application starts analyzing the data, memory occupancy must be limited to a certain level to make sure the system works smoothly.

**4.3. Budget**

The budget for the entire development process must be less than 10,000 USD.

**4.4. System Downtime**

The occurrence of system shutdowns should also be taken into consideration. The downtime must be less than less than 30 minutes per year, which maximizes the demand for the application due to its reliability.

**5. Change Cases**

1. In the near future, we might be able to support not only txt file, but Excel files as well

(TBA)

**6. Milestones**

1. Submit SRS for review by Mar 27th.

3. Get design done by May 16th.

4. Get coding done by June 6th.

**7. References**

*Omg project by 宋一豪 and 徐宇泽 :* [*https://oh-my-genes.readthedocs.io/en/latest/*](https://oh-my-genes.readthedocs.io/en/latest/)

**8. Group Information**

**8.1. Group Name**

Machinda Zvazvinhu

**8.2. Group Members**

1. Golden 201732120168: *goldenmc@zjnu.edu.cn*

2. Dean 201732120164: *deanmajaya@zjnu.edu.cn*