Table of Contents

[**1. Introduction** 2](#_Toc4575945)

[**1.1. Purpose** 2](#_Toc4575946)

[**1.2. Overview** 2](#_Toc4575947)

[**1.3. User Characteristics** 2](#_Toc4575948)

[**1.4. Terminologies & Abbreviations** 2](#_Toc4575949)

[**1.4.1. Terminologies** 3](#_Toc4575950)

[**1.4.2. Abbreviations** 3](#_Toc4575951)

[**2. Functional Requirements** 3](#_Toc4575952)

[**2.1. Input** 3](#_Toc4575953)

[**2.2. Output(Data Analyzing)** 5](#_Toc4575954)

[**2.2.1. Table Analyzing** 5](#_Toc4575955)

[**2.2.2. Plot Analyzing** 6](#_Toc4575956)

[**3. Non-functional Requirements** 7](#_Toc4575957)

[**3.1. Response Time** 7](#_Toc4575958)

[**3.2. Aesthetic Aspects** 7](#_Toc4575959)

[**3.3. Confidentiality Policy** 7](#_Toc4575960)

[**4. Constraints** 8](#_Toc4575961)

[**4.1. Browser Compatibility** 8](#_Toc4575962)

[**4.2. Space Complexity** 8](#_Toc4575963)

[**4.3. Budget** 8](#_Toc4575964)

[**4.4. System Downtime** 8](#_Toc4575965)

[**5. Change Cases** 8](#_Toc4575966)

[**6. Milestones** 9](#_Toc4575967)

**1. Introduction**

We have made this particular Software Requirements Specifications (SRS) to simulate a set of requirements for the project “Oh My Genes” (OMG), a web application for identifying genes’ expression differential.

**1.1. Purpose**

The main purpose of this document is to identify differentially expressed genes given a gene expression file containing two cell samples. By having OMG many biologists can analyze their data. Not only analyze their data but also, they can get an accurate scatter plot of their data.

**1.2. Overview**

The OMG has a very simple and friendly interface. When we open the OMG, we can see “Choose File” and “Upload & Go”. Scientists and biologists around the globe can choose the file they want to upload that contains gene expression levels from two samples, representing two experimental conditions. The OMG software will return a table of differentially expressed genes and a scatter plot of these genes whose X-axis is control and Y-axis is treatment. If an invalid gene expression is chosen, the software returns a page informing the user to provide another file.

**1.3. User Characteristics**

**Client User:** Scientists and Biologists who are working on or studying about genes expression.

**Site Maintainer:** Technicians who have the basic knowledge of flask and python to maintain a website.

**1.4. Terminologies & Abbreviations**

By having this section in our document (SRS), we shall explain the terminologies and abbreviations in a simple way for any user to use. It leads to a better understanding of the projects main purpose.

**1.4.1. Terminologies**

Control sample - A cell sample prepared in its normal condition.

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.

**1.4.2. Abbreviations**

OMG - Oh My Genes

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

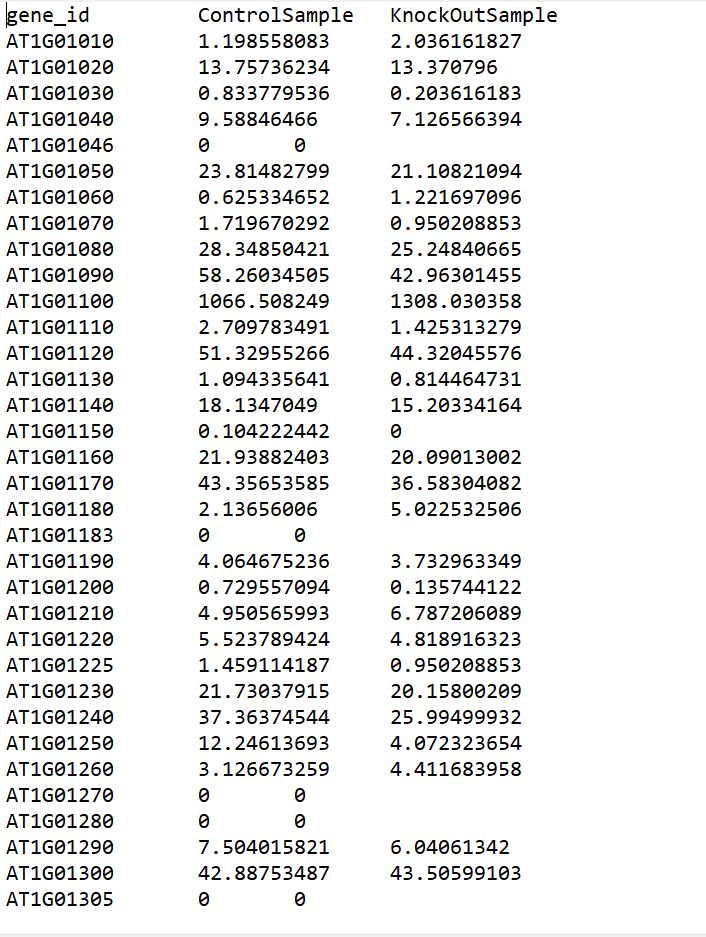
**2. Functional Requirements**

**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. Knockout Sample).



**2.2. Output (Data Analyzing)**

The software OMG adds a log fold change of gene expression on the first table we insert. The logFC will be displayed as the last column of the table. Not only logFC but we will also get a scatter plot when we upload gene expression file.

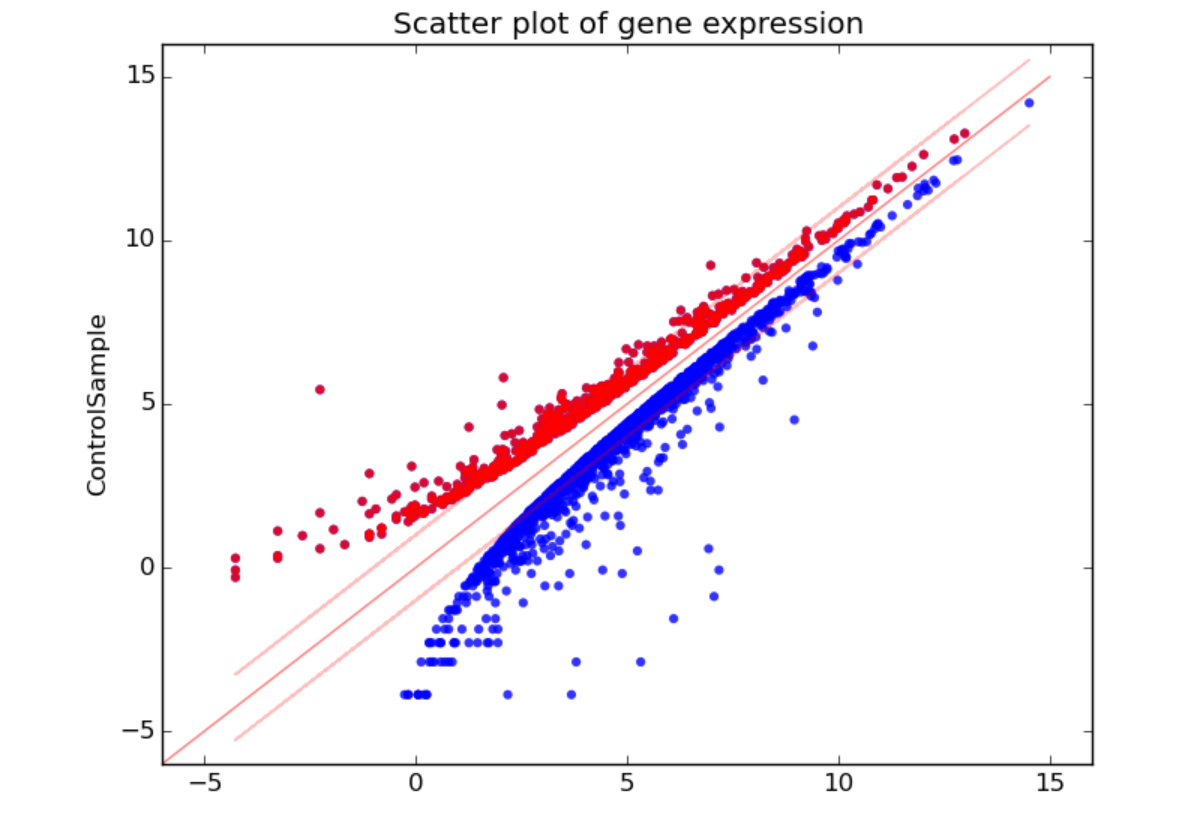
**2.2.1. Table Analyzing**

The table contains a list of differentially expressed genes with the following format:

| **gene\_id** | **control\_sample** | **knockout\_sample** | **log\_2[FC]** |
| --- | --- | --- | --- |
| AT1G01010 | 1.198558083 | 2.036161827 | 0.76 |
| AT1G01020 | 13.75736234 | 13.370796 | -0.04 |
| AT1G01030 | 0.833779536 | 0.203616183 | -2.03 |
| AT1G01040 | 9.58846466 | 7.126566394 | -0.43 |
| AT1G01046 | 0 | 0 |  |
| AT1G01050 | 23.81482799 | 21.10821094 | -0.17 |
| AT1G01060 | 0.625334652 | 1.221697096 | 0.97 |
| AT1G01070 | 1.719670292 | 0.950208853 | -0.86 |
| AT1G01080 | 28.34850421 | 25.24840665 | -0.17 |
| AT1G01090 | 58.26034505 | 42.96301455 | -0.44 |
| AT1G01100 | 1066.508249 | 1308.030358 | 0.29 |
| AT1G01110 | 2.709783491 | 1.425313279 | -0.93 |

**2.2.2. Plot Analyzing**

The scatter plot displays differentially expressed genes. The X-axis is Control, and Y-axis is KnockOut. Replace ‘Control’ and ‘KnockOut’ with appropriated column names if provided in the uploaded file. The up-regulated genes are shown in red dots, and down-regulated genes are shown in blue.



**3. Non-functional Requirements**

For this section, we are going to clarify some non-functional requirements for a better using experience.

**3.1. Response Time**

In technology, response time is the time a system or functional unit takes to react to a given input. This software is used by very many scientists and biologists all around the globe so there is a great chance of response time delay. The server maybe has to stand heavy load of requests thus may cause some problems. We should also consider the time we are going to need for large amount of data analyze. For the optimization aspects, our algorithm should be sufficiently effective to deal with this situation. I and my colleagues , have decided, the limit of response time, should **less than 5 seconds** for the most cases.

**3.2. Aesthetic Aspects**

As a web application to solve biological analyzing. The interface should be designed as simply as possible and make the scatter plot and table more distinct.

**3.3. Confidentiality Policy**

Since this web application is public, the security of data must be a significant part to be considered. Biologists have taken quiet a good time to get their experiment result. Therefore, when users are using the application, their data must be used and stored safely and properly. In other words, the web application must maintain the experiment data’s reliability, integration and secrecy and the experimental data will not be divulged without the agreement of the owner.

**4. Constraints**

For scientific and analyze uses, this application is orientated to various users(biologists) all over the world. Before developing the project, we have to take some main constraints into account so that we can come up any useful ideas to deal with them in the future.

**4.1. Browser Compatibility**

Our application is designed for a world wide application usage so it should have a cross platform compatibility to satisfy various users. Thus, the application should be **accessible through Firefox, Chrome, and Safari**.

**4.2. Space Complexity**

Not only the time complexity should be considered, but also the space complexity. There are two aspects for space complexity. One is the size of the whole web application. **The application should not larger than 1GB.** For another aspect, when the application starts data analyzing, memory occupancy must be limited under a certain level to make sure the system work in a proper way.

**4.3. Budget**

Budget less than 13,000 USD.

**4.4. System Downtime**

System **downtime less than 25 minutes per year** in order to satisfy users’ large demand of data analyzing.

**5. Change Cases**

1. We are working on how we can support Excel files.
2. Analyzed result downloading can be supported in the long run.
3. More functions will be provided from the application.

**6. Milestones**

1. Submit SRS for review by **March 27th.**
2. Get SRS approved by **April 15th**.
3. Get design done by **May 10th**.
4. Get coding done by **May 15th**.
5. Acceptance tests by **June 1st**.
6. Release by **June 6th**

**7. Appendices**

| **Date** | **Change Log** |
| --- | --- |
| March 27, 2019 | 1. Specified basically functions and interface for the application. |
| 1. Get non-functional requirements in detail. |
| 1. Clarified some constrains about the application. |
| 1. Imagined some change cases and tried to find ways to solve them. |
| 1. Decided the project developing milestones. |