

Oh my genes

- 1. Introduction
 - 1.1. Purpose
 - 1.2. Overview
 - 1.3. User characteristics
 - 1.4. Terminology&Glossary
- 2. Functional requirements
 - 2.1. Input
 - 2.2. Output
- 3. Non-functional requirements
 - 3.1. Response Time
 - 3.2. Aesthetic Aspects
 - 3.3. Confidentiality Policy
- 4. Constraints
 - 4.1. Compatibility
 - 4.2. Servet space
 - 4.3. Budget
 - 4.4. System downtime
- 5. Change cases
 - 5.1. Information technology aspects
 - 5.2. Genetics aspects
- 6. Milestones
- 7. Appendices
- 8. References

Group Member:

王雪洁(201732120106) 249795581@qq.com

• 1. Introduction

- This document describes the Little Hill Lab's initial requirements for an online application (Oh My Genes) which allows our scientists to upload gene expression files and quickly get differentially expressed genes.

- **1.1. Purpose**

- To identify differentially expressed genes given a gene expression file containing two cell samples.

- **1.2. Overview**

- The web application has a simple interface with a single button [Upload and GO]. Our scientists upload a plain text file containing gene expression levels from two samples, representing two experimental conditions. Accepting the file, the 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 given, the web application returns a page informing the user to provide the correct format.

1.3. User characteristics

- **clients:** Biologists who are using the web to get result of gene expression.
- **maintainers:** Biologists who have the knowledge of python and github.

1.4. Terminology&Glossary

1.4.1. Terminology

- *Control sample* - a cell sample prepared in its normal condition.
- *Knock out 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. Glossary

- *logFC* - log fold change of gene expression. $\log_2 [T/C]$, where T is the gene

expression level from a knock out 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 (see the attached file for a full example).

The file contains an optional head line, followed by each gene's expression in a control sample (e.g., ControlSample) and in a knock out sample (e.g., KnockOutSample).

See the txt file pattern as following:

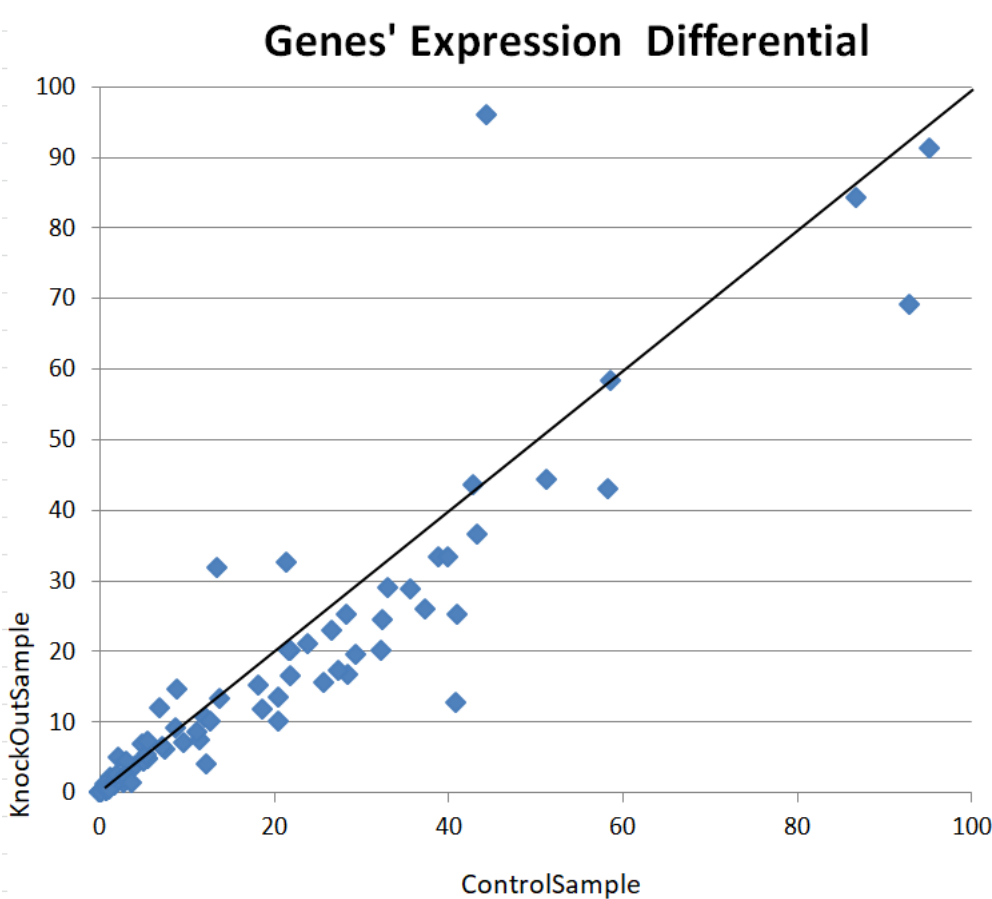
gene_id	ControlSample	KnockOutSample
AT1G01010	1.198558083	2.036161827
AT1G01020	13.75736234	13.370796
AT1G01030	0.833779536	0.203616183
AT1G01040	9.58846466	7.126566394
AT1G01046	0	0
AT1G01050	23.81482799	21.10821094
AT1G01060	0.625334652	1.221697096
AT1G01070	1.719670292	0.950208853
AT1G01080	28.34850421	25.24840665
AT1G01090	58.26034505	42.96301455
AT1G01100	1066.508249	1308.030358
AT1G01110	2.709783491	1.425313279

2.2. Output

The web application displays a table and a scatter plot given a gene expression file. The table contains a list of differentially expressed genes with the following format:

gene_id	control_sample	treat_sample	log_2[FC]
AT1G01020	13.75736234	13.370796	0.49

The scatter plot displays differentially expressed genes. The X-axis is Control, and Y-axis is Treatment. Replace 'Control' and 'KnockOut' with appropriated column names if provided in the uploaded file. The up-regulated genes are shown above the slash, and down-regulated genes are shown below the slash.



3. Non-functional Requirements

3.1. Response Time

Because this web application is being used by biologists around the world, servers can become overloaded. If the user has a large amount of data, time complexity must be considered. Overall, the response time should be less than five seconds.

3.2. Aesthetic Aspects

The page should be easy to understand.

the results should be clear.

3.3. Confidentiality Policy

For biologists, every result of data is precious, so it is crucial to protect the safety of genetic information. Our network program must ensure the correct reliability of data, security and confidentiality. The experimental data and results cannot be disclosed privately.

4. Constraints

This application is aimed at biologists around the world. Before development, we need to consider some constraints, such as how to solve the network congestion problem that many people use at the same time, etc. With the development of technology, these constraints will be solved, and then new constraints will be generated.

4.1. Compatibility

Because the application is used worldwide, it should be cross-platform compatible and accessible through most browsers, such as Firefox.

4.2. Servet space

Web space less than 1GB

4.3. Budget

Budget less than 10,000 USD

4.4. System downtime

System downtime less than 30 minutes per year

5. Change cases

With the progress of information technology and genetics, there are some potential problems we may need to consider, they are divided into two aspects for us to talk about in following:

5.1. Information technology aspects

Information technology is developing very fast .No matter how it changes, the basic framework is the same. As the gene pool data increases, we need to optimize the gene pool, shorten the response time of the browser.

5.2. Genetics aspects

The future of genetics is uncertain, but it is certain that the gene pool will become larger and larger, and we must learn to organize and plan the gene pool.

6. Milestones

1. Submit SRS for review by **March 27th**.

7. Appendices

Date	Change Log
March 10 th --March 27th	1.Understand the purpose and direction of the application.
	2.Get non-functional requirements in detail.
	3.Knew some constrains about the application.
	4.Imagined some change cases and tried to slove them.
	5.Decided the project developing milestones.

8. References

OMG's example of ReadTheDocs:

<https://oh-my-gene.readthedocs.io/en/latest/>