SE. ACHAAB GODWIN FERGUSON
201632120144
COLLEGE OF M.P.I-ZJNU

ONLINE APPLICATION (OH MY GENES) FOR OUR SCIENTISTS

CLIENTS:

Biologists in our lab (Potentially worldwide).

PRODUCT:

Gene Web Application.

PURPOSE:

Identifying differentially expressed genes given a gene expression fill containing two cell sample.

ABSTRACT:

Gene expression: The translation of information encoded in a gene into protein or RNA structures that are present and operating in the cell. Expressed genes include genes that are transcribed into messenger RNA (mRNA) and then translated into protein, as well as genes that are transcribed into RNA, such as transfer and ribosomal RNAs, but not translated into protein.

The process of gene expression is used by all known life—eukaryotes (including multicellular organisms), prokaryotes (bacteria and Achaea), and utilized by viruses—to generate the macromolecular machinery for life.

Several steps in the gene expression process may be modulated, including the transcription, RNA splicing, translation, and post-translational modification of a protein. Gene regulation gives the cell control over structure and function, and is the basis for cellular differentiation, morphogenesis and the versatility and adaptability of any organism. Gene regulation may also serve as a substrate for evolutionary change, since control of the timing, location, and amount of gene expression can have a profound effect on the functions (actions) of the gene in a cell or in a multicellular organism.

INTRODUCTION:

A typical differential expression analysis of RNA-Seq data consists of normalizing the raw counts and performing statistical tests to reject or accept the null hypothesis that two groups of samples show no significant difference in gene expression. This example shows how to inspect

the basic statistics of raw count data, how to determine size factors for count normalization and how to infer the most differentially expressed genes using a negative binomial model.

The dataset for this example comprises of RNA-Seq data obtained in the experiment described by Brooks et al. [1]. The authors investigated the effect of siRNA knock-down of pasilla, a gene known to play an important role in the regulation of splicing in Drosophila melanogaster.
The dataset consists of 2 biological replicates of the control (untreated) samples and 2 biological replicates of the knock-down (treated) samples

DOMAIN(S):

- Control Sample.
- Treatment Sample.
- Differentially Expressed Genes.
- Up-Regulation.
- LogFC.

Control Sample:
Genetic control may be on the transcriptional or translational level. Transcriptional control works by controlling the number of RNA transcripts of a region of DNA, indirectly controlling protein synthesis. Translational control of protein synthesis works by regulating the step of translating RNA into protein.

Treatment Sample:
A cell sample treated by special chemicals, or in which some genes are altered.

Differentially Expressed Genes:
The unused genes in differentiated cells are not destroyed or mutated, and they retain the potential for being expressed. Only a small percentage of the genome is expressed in each cell, and a portion of the RNA synthesized in the cell is specific for that cell type.

Up-Regulation:
A gene is said to be up-regulated if it has higher expression in treatment than in control.
Includes a wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products (protein or RNA), and is informally termed gene regulation. Sophisticated programs of gene expression are widely observed in biology, for example to trigger developmental pathways, respond to environmental stimuli, or adapt to new food sources. Virtually any step of gene expression can be modulated, from transcriptional initiation, to RNA processing, and to the post-translational modification of a protein. Often, one gene regulator controls another, and so on, in a gene regulatory network.

Gene regulation is essential for viruses, prokaryotes and eukaryotes as it increases the versatility and adaptability of an organism by allowing the cell to express protein when needed. Although

as early as 1951, Barbara McClintock showed interaction between two genetic loci, Activator (Ac) and Dissociator (Ds), in the color formation of maize seeds, the first discovery of a gene regulation system is widely considered to be the identification in 1961 of the lac operon, discovered by François Jacob and Jacques Monod, in which some enzymes involved in lactose metabolism are expressed by E. coli only in the presence of lactose and absence of glucose.

In multicellular organisms, gene regulation drives cellular differentiation and morphogenesis in the embryo, leading to the creation of different cell types that possess different gene expression profiles from the same genome sequence. This explains how evolution actually works at a molecular level, and is central to the science of evolutionary developmental biology ("evo-devo").

The initiating event leading to a change in gene expression includes activation or deactivation of receptor

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.

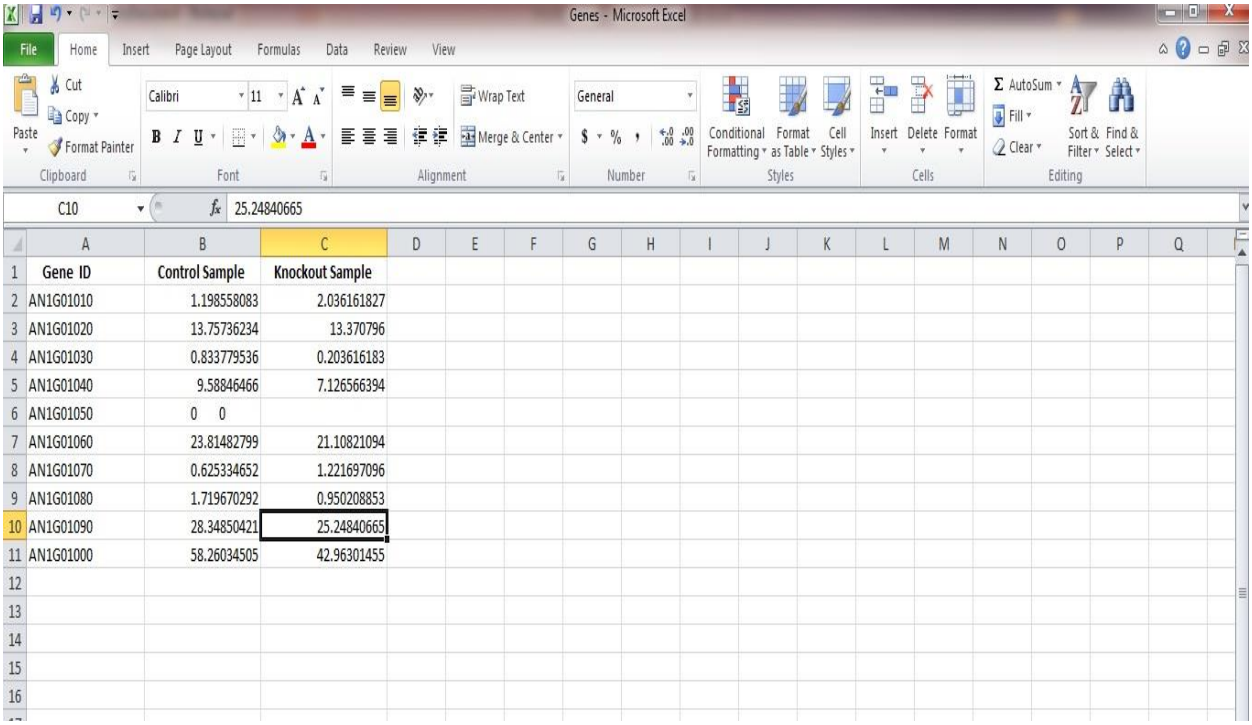
FUNCTIONALITY:

The web application is a simple genes Web App design for our scientists to upload a plain text file containing levels from two samples, representing two experimental conditions. The App has a provision for five samples of gene, which is the Control sample, Treatment sample, Differentially Expressed Genes, Up-Regulation and LogFC. And also a button (upload and go), which would be clicked anytime genes are being uploaded into the App to reach with each other.

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.

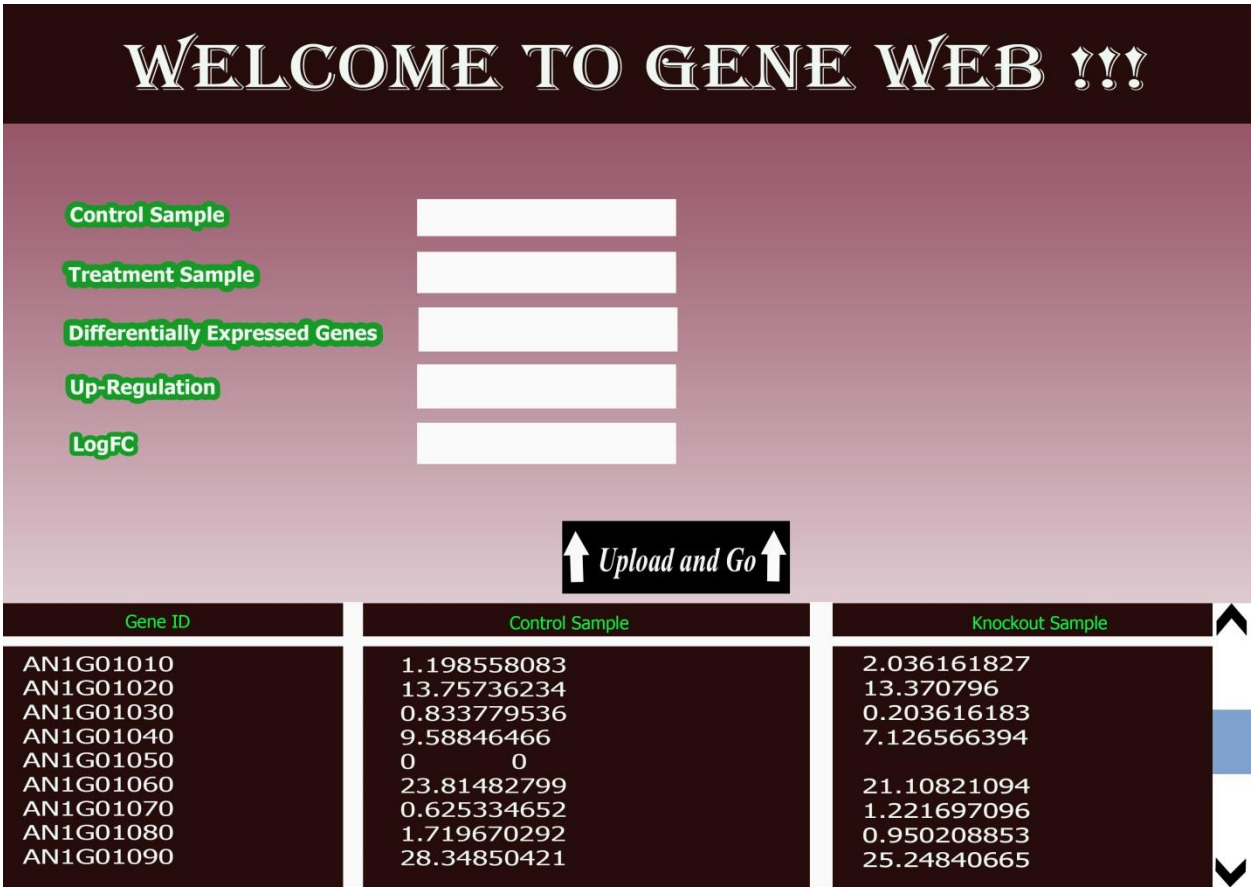
```
<!DOCTYPE html>
<html lang="en">

<head>
    <title>Python Flask Bucket List App</title>


    <link href="http://getbootstrap.com/dist/css/bootstrap.min.css" rel="stylesheet">

    <link href="http://getbootstrap.com/examples/jumbotron-narrow/jumbotron-narrow.css" rel="stylesheet"


</head>

<body>

    <div class="container">
        <div class="header">
            <nav>
                <ul class="nav nav-pills pull-right">
                    <li role="presentation" class="active"><a href="#">Home</a>
```

EXCEL FILE:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Gene ID	Control Sample	Knockout Sample														
2	AN1G01010	1.198558083	2.036161827														
3	AN1G01020	13.75736234	13.370796														
4	AN1G01030	0.833779536	0.203616183														
5	AN1G01040	9.58846466	7.126566394														
6	AN1G01050	0 0															
7	AN1G01060	23.81482799	21.10821094														
8	AN1G01070	0.625334652	1.221697096														
9	AN1G01080	1.719670292	0.950208853														
10	AN1G01090	28.34850421	25.24840665														
11	AN1G01000	58.26034505	42.96301455														
12																	
13																	
14																	
15																	
16																	

App Interface



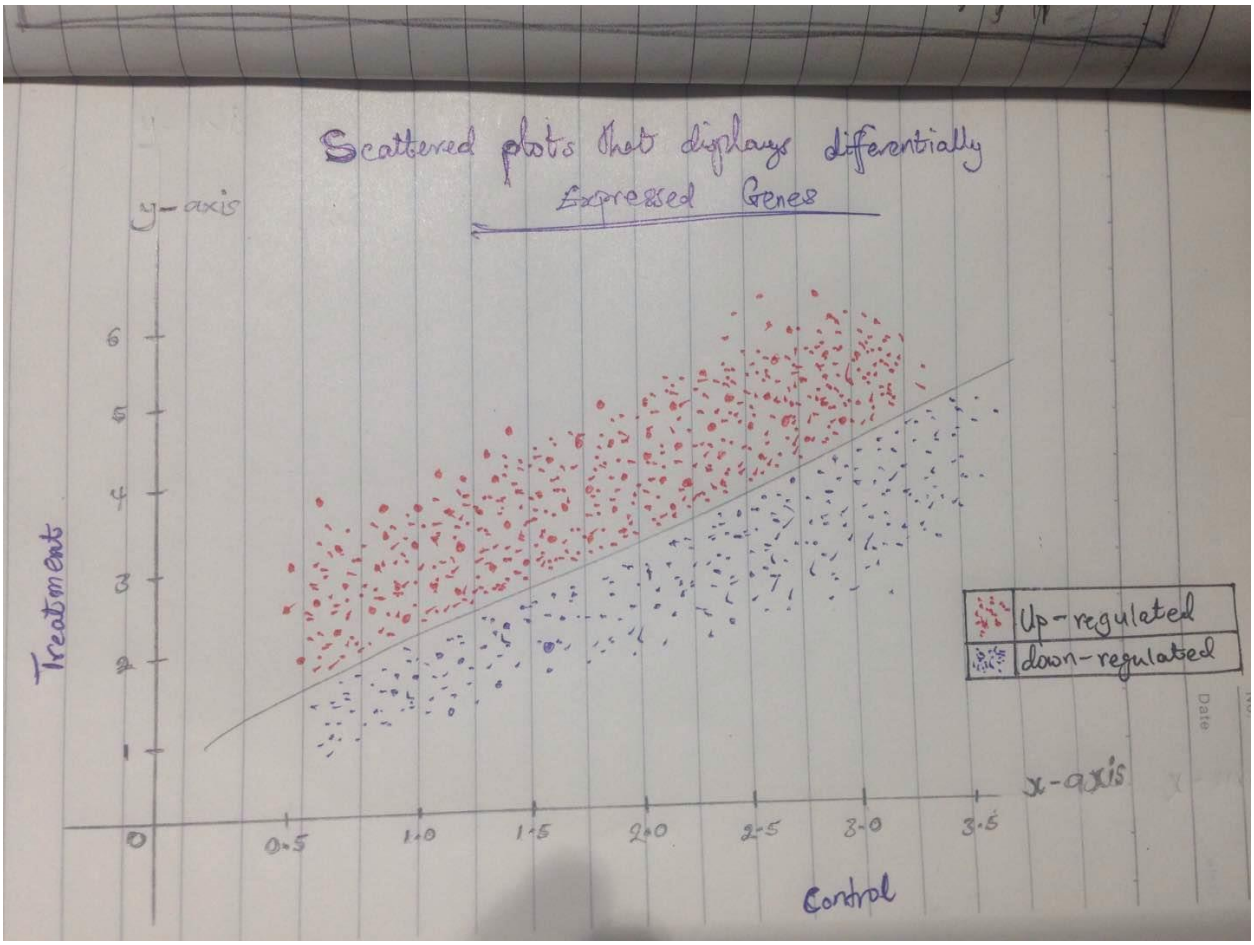
What the App is displaying is the input when a valid submitted gene expression file has the following format. It is a TAB-delimited, plain text file with three columns. The file contains an optional head line, followed by each gene's expression in a control sample (Control Sample) and in a treatment sample (Knockout Sample).

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	Treatment Sample	Log_2[FC]
AN1G01010	1.198558083	2.036161827	0.76

The scatter plot displays differentially expressed genes. The X-axis is Control, and Y-axis is Treatment. The up-regulated genes are shown in red dots, and down-regulated genes are shown in blue.



References:

- Marioni et al. RNA-Seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Research 2008. 18:1509-1517.
- https://www.ncbi.nlm.nih.gov
- Mortazavi et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 2008. 5:621-628.
- http://bioinfo.au.tsinghua.edu.cn
- http://bioinfo.au.tsinghua.edu.cn/software/degseq