**Mansoura University**



**Faculty of Computers and Information**

**Department of Computer Science**

**Project Proposal**

# Arabic Title

# 

### 

##### English Title

### 

### **Submitted by:**

#### 

|  |  |  |
| --- | --- | --- |
| Student Name | Student Email | Section |
| Noor Mohamed raafat | nourraafat51@gmail.com | Section 9 |
| Manar Ashraf Elemetwaly | Manarashraf692@gmail.com | Section 5 |
| Nada Salah Abd\_Elwahab | snada4846@gmail.com | Section5 |

Project Abstract:

The incidence of both type I and II diabetes continues to increase globally and the effectiveness of pharmaceutical treatments for the disease remains limited. A significant proportion of basic research into the disease has focused on understanding the transcriptional programs that regulate the development and function of the endocrine pancreas. Of the numerous research tools and approaches that have been utilized to address these fundamental questions of development, global gene expression profiling has become one of the most powerful techniques employed to date.

Project Objectives:

producing differential expressed gene in diabetes with infection and no infection

Who are the project **competitive**? and how will your project be **different**?

This tool takes as input a table of raw counts. The count table has to be associated with a phenodata file describing the experimental groups. These files are best created by the tool "Utilities / Define NGS experiment", which combines count files for different samples to one table, and creates a phenodata file for it.

DESeq2 performs an internal normalization where geometric mean is calculated for each gene across all samples. The counts for a gene in each sample is then divided by this mean. The median of these ratios in a sample is the size factor for that sample. This procedure corrects for library size and RNA composition bias, which can arise for example when only a small number of genes are very highly expressed in one experiment condition but not in the other.

As small numbers of replicates make it impossible to estimate within-group variance reliably, DESeq2 uses shrinkage estimation for dispersions and fold changes. A dispersion value is estimated for each gene through a model fit procedure. You need to have biological replicates of each experiment condition in order to estimate dispersion properly. If there are no replicates, DESeq will estimate dispersion using the samples from the different conditions as replicates.

DESeq2 fits negative binomial generalized linear models for each gene and uses the Wald test for significance testing. In addition to the group information, you can give an additional experimental factor like pairing to the analysis.

DESeq2 detects automatically count outliers using Cooks's distance and removes these genes from analysis. It also automatically removes genes whose mean of normalized counts is below a threshold determined by an optimization procedure. Removing these genes with low counts improves the detection power by making the multiple testing adjustment of the p-values less severe.

**different**

The ultimate aim of the gene expression profiling procedures described in this chapter is to generate a list of genes that demonstrate differential expression as a result of the biological model in question. Because microarrays measure the expression of thousands of genes simultaneously, identification of changes in gene expression between different biological states requires methods to determine the significance of these changes while accounting for the enormous number of genes tested. Numerous statistical analysis tools exist for this purpose,

Tools, Hardware and Software Resources:

**Tools :-R**

**Software:- R\_Studio**

**Hardware:- computer**

SCHEDULING PHASES:

|  |  |  |
| --- | --- | --- |
| **From** | **To** | **Activity** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

References:

[**https://doi.org/10.1073/pnas.091062498**](https://doi.org/10.1073/pnas.091062498)

[**https://link.springer.com/protocol/10.1007/978-1-59745-448-3\_16**](https://link.springer.com/protocol/10.1007/978-1-59745-448-3_16)

[**https://chipster.csc.fi/manual/deseq2.html**](https://chipster.csc.fi/manual/deseq2.html)

[**https://chipster.csc.fi/manual/deseq2.html**](https://chipster.csc.fi/manual/deseq2.html)

[**http://scholar.google.com/scholar\_lookup?title=Use%20of%20within-array%20replicate%20spots%20for%20assessing%20differential%20expression%20in%20microarray%20experiments&author=G%20K.%20Smyth&author=J.%20Michaud&author=H%20S.%20Scott&journal=Bioinformatics&volume=21%0A&pages=2067-2075&publication\_year=2005**](http://scholar.google.com/scholar_lookup?title=Use%20of%20within-array%20replicate%20spots%20for%20assessing%20differential%20expression%20in%20microarray%20experiments&author=G%20K.%20Smyth&author=J.%20Michaud&author=H%20S.%20Scott&journal=Bioinformatics&volume=21%0A&pages=2067-2075&publication_year=2005)

[**http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Abstract&list\_uids=12094230**](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Abstract&list_uids=12094230)

**. Edlund H. (2002) . Pancreatic organogenesis - developmental mechanisms and implications for therapy . Nat. Rev. Genet . 3 , 524 – 532**

**. Habener J. F. , Kemp D. M. , and Thomas M. K. (2005). Minireview: transcriptional regula-tion in pancreatic development . Endocrinology 146 , 1025 – 1034**

**. Lantz K. A. and Kaestner K. H. (2005) . Winged-helix transcription factors and pancre-atic development . Clin. Sci . 108 , 195 – 204**

**. Wilson M. E. , Scheel D. , and German M. S. (2003). Gene expression cascades in pancre-atic development . Mech. Dev . 120 , 65 – 80**