**NTCC REPORT**

**On**

**Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana***

Submitted to



Amity University Uttar Pradesh

In partial fulfilment of the requirements for the award of the degree of Master of Science

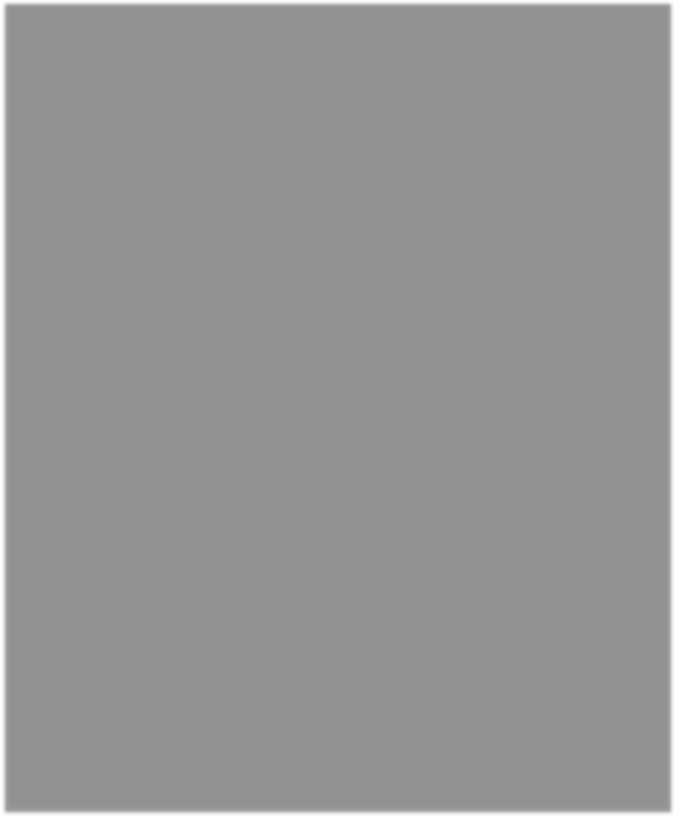
(Bioinformatics) By

# RUCHI BACHAL

**AMITY UNIVERSITY OF BIOTECHNOLOGY AMITY UNIVERSITY UTTAR PRADESH**

**NTCC Convener**

**COMPLETION CERTIFICATE**





**AMITY INSTITUTE OF BIOTECHNOLOGY DECLARATION**

I, **Ruchi Bachal** student of **MSc (Bioinformatics)** hereby declare that the project entitled “**Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana*”** which is submitted by me to **Amity Institute of Biotechnology**, Amity University, Uttar Pradesh, in partial fulfilment of requirement for the award of the degree in Masters of Science has not been previously formed the basis for the award of any degree, diploma, or other similar title or recognition.

Noida

Date: Name and Signature of Student



**AMITY INSTITUTE OF BIOTECHNOLOGY NTCC-2023**

**CERTIFICATE BY IFC**

On the basis of declaration submitted by **Ruchi Bachal**, student of **Master of Science**, I hereby certify that the project titled “**Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana***” which is submitted to **Amity Institute of Biotechnology**, Amity University, Uttar Pradesh, in partial fulfilment of **Master of Science (Bioinformatics)**, is an original contribution with existing knowledge and faithful record of work carried out by her under my guidance and supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Noida

Date: Mr Rajesh Kumar Jangade

Professor Amity Institute of Biotechnology



# AMITY INSTITUTE OF BIOTECHNOLOGY NTCC-2023

**PLAGIARISM CERTIFICATE**

This is to certify that the thesis entitled **“Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana*”** submitted by **Ruchi Bachal** for the partial fulfilment of the degree of Master of Technology in Biotechnology has been checked by **Turnitin** software for plagiarism. The thesis has **5% percent** plagiarism.

**Signature of IFC Signature of the NTCC Coordinator**



**AMITY INSTITUTE OF BIOTECHNOLOGY NTCC-2023**

**ACKNOWLEDGEMENT**

It is high privilege for me to express my deep sense of gratitude to those entire faculty members who helped me in the completion of the project entitled “**Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana***” under the supervision of my internal guide **Mr Rajesh Kumar Jangade** and external guide **Dr. Yashwant Patel**.

My special thanks to all other faculty members, batchmates & seniors of Amity Institute of Biotechnology, Amity University, Uttar Pradesh for helping me in the completion of project work and its report submission.

Ruchi Bachal A0500322012

# ABSTRACT

The Phenylpropanoid pathway, an important secondary metabolite pathway in plants produces a variety of secondary metabolites, like flavonoids, lignins. In terms of phenylpropanoid pathway, Arabidopsis thaliana is most extensively studied plant model which displays significant phenotypic variation. In this research, we performed a differential transcriptome analysis of Arabidopsis sample. In this study, we used RNA Seq to identify differentially expressed genes(DEGs) and different molecular, biological processes and pathways related with Arabidopsis. The study resulted in 50 upregulated genes and 71 downregulated genes by using the iDEP tool and enrichment and annotation was performed using Metascape. Our result indicates various other pathways other than Phenylpropanoid pathway. Our research provides an insightful information on the phenylpropanoid pathway and other pathways present in Arabidopsis. Additionally, our study demonstrates biological processes and molecular processes related with Arabidopsis.

Keywords: Phenylpropanoid pathway, Arabidopsis, RNA Seq, DEGs

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **S.NO** | **CONTENTS** | **PAGE NUMBER** |
| **1.** | COMPLETION CERTIFICATE | i |
| **2.** | DECLARATION | ii |
| **3.** | CERTIFICATE BY IFC | iii |
| **4.** | PLAGIARISM CERTIFICATE | iv |
| **5.** | ACKNOWLEDGEMENT | v |
| **6.** | ABSTRACT | vi |
| **7.** | LIST OF FIGURES | viii |
| **8.** | LIST OF TABLES | ix |
| **9.** | CHAPTER-1: INTRODUCTION | 1-2 |
| **10.** | CHAPTER-2: OBJECTIVE | 3 |
| **11.** | CHAPTER-3: LITERATURE REVIEW | 4-5 |
| **12.** | CHAPTER-4: METHODOLOGY | 6-10 |
| **13.** | CHAPTER-5: RESULTS | 11-28 |
| **14.** | CHAPTER-6: DISCUSSION | 29 |
| **15.** | CHAPTER-7: CONCLUSION | 30 |
| **16.** | REFERENCES | 31 |
| **17.** | WEEKLY PROGRESS REPORTS | 33 |
| **18.** | SYNOPSIS | 41 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE**  **NO** | **DESCRIPTION** | **PAGE**  **NO.** |
| **1.** | Workflow of RNA Seq | 9 |
| **2.** | A graph from the Output of FastQC report | 12 |
| **3.** | A graph from Output of Trimgalore report | 13 |
| **4.** | Bar plot of case and control samples | 14 |
| **5.** | Box plot of case and control samples | 14 |
| **6.** | Density plot case and control samples | 15 |
| **7.** | Heatmap of *Arabidopsis thaliana*'s top 100 genes. Genes that are upregulated are shown in Red, downregulated genes are shown in Green, and genes that are unchanged are shown in Black | 16 |
| **8(a).** | PCA Non-interactive plot case and control samples | 17 |
| **8(b).** | PCA 3D plot (left orientation) | 17 |
| **9.** | Output showing Differentially expressed genes case and control  samples | 18 |
| **10.** | Venn Diagram plot of DEGs case and control samples | 19 |
| **11.** | Upset plot of DEGs case and control samples | 19 |
| **12.** | Volcano plot of DEGs case and control samples | 20 |
| **13.** | MA plot of DEGs case and control samples | 20 |
| **14.** | Scatter plot of DEGs case and control samples | 21 |
| **15.** | Bar graph of enriched terms across input gene lists, colored by p-  values in Molecular process | 21 |
| **16.** | Bar graph of enriched terms across input gene lists, colored by p-  values in biological process | 22 |
| **17.** | Bar graph of enriched terms across input gene lists, colored by p-  values in KEGG pathway | 23 |

# LIST OF TABLES

|  |  |  |
| --- | --- | --- |
| **TABLE**  **NO.** | **DESCRIPTION** | **PAGE**  **NO.** |
| **1.** | Metadata consisting of sample description and replicate of the  case and control sample | 6 |
| **2.** | Top 7 clusters with their representative enriched terms in  molecular process | 22 |
| **3.** | Top 20 clusters with their representative enriched terms in  biological process | 23 |
| **4.** | Top 6 clusters with their representative enriched terms in KEGG  pathway | 23 |
| **5.** | Annotated genes generated by Metascape | 24 |

**CHAPTER-1 INTRODUCTION**

Arabidopsis is a plant species which belongs to mustard family Brassicaceae has become a useful model organism for scientific studies in biology of plant. Various characteristics features are present in Arabidopsis such as small size of the plant, short life cycle of about 6-8 weeks, genome is smaller as compared to genome of other species and also it was the first plant species whose genome was fully sequenced making it a useful model. There are about 20,000 genes of Arabidopsis and more than 30 megabases of annotated genomic sequences present in GenBank database[1]. Arabidopsis thaliana is a model plant species that has been extensively studied to cover the various molecular and biological mechanisms governing the phenylpropanoid pathway.

Phenylpropanoids are a group of secondary metabolites present in plant which are derived from phenylalanine. Phenylpropanoid pathway is a rich source of metabolite which serves as biosynthetic metabolism for lignin and also serves as a precursor to form various compounds which includes flavonoids, lignins and esters. Studies on many plant species and also on Arabidopsis have shown that phenylpropanoids and their derivatives are essential for survival of plants being classified as secondary metabolites. Using Arabidopsis as a model species, researchers have been able to understand how phenylpropanoid pathway works in plants, various enzymes and their function in plants and how phenylpropanoid pathway interacts and react with plant metabolism in different conditions[2].

In Arabidopsis, phenylpropanoids are produced from the amino acid phenylalanine and. Cinnamic acid is first product produced by deamination of phenylalanine. After formation of Cinnamic acid, cinnamic acid undergoes hydroxylation to from p-coumaric acid. From p-coumaric acid this pathway breaks and branches into various other branches which lead to formation of different classes of phenylpropanoid which includes flavonoid biosynthesis, lignin biosynthesis and hydroxycinnamic acid derivates[3].

Above reactions are regulated by different genes which are when upregulated and downregulated may lead to several changes or may alter any step in the pathway. These changes are diagnosed by differential gene expression when two contrasting phenotypes of Arabidopsis are compared.

Transcriptome refers to the complete set of all RNA molecules that are expressed in cell, tissue. It represents whole set of RNA that includes mRNA, non-coding RNA and other RNA molecules present. Differential transcriptome is basically an approach which is used to identify differentially expressed genes present in different conditions or groups and it shows significant differences or changes present in different conditions. To perform differential transcriptome many approaches are there such as Microarray techniques and RNA Sequencing. Nowadays RNA Sequencing is used for differential

transcriptome analysis as it provides in-depth and precise data regarding differentially expressed genes[4].

RNA Seq is a type of next-generation sequencing which is a high throughput technique used to see presence and quantity of RNA in a sample. It provides information not only about Transcriptome but also of gene expression, alternative splicing. Here in Arabidopsis for RNA Seq, Tuxedo protocol was performed to obtain differentially expressed genes to compare the phenotypes[4][5].

# CHAPTER-2 OBJECTIVE

The objective of studying elucidation of cell signalling pathways through RNA seq in contrasting phenotypes of Arabidopsis thaliana is to identify the genes that are differentially expressed in contrasting phenotypes. It is possible for researchers to assess the levels of gene expression in cells or tissues by using the RNA sequencing (RNA-seq) technology. Researchers can determine the genes that cause variations in phenotypes in Arabidopsis thaliana by comparing the levels of gene expression between various phenotypes. This knowledge may help in the understanding of the molecular processes that underlie phenotypic variations and set the way for the discovery of novel disease therapies. Understanding how phenylpropanoid pathway function in different phenotypes of Arabidopsis thaliana could be useful in our understanding of how this pathway function in other species of plants. Many genes that are important for phenylpropanoid pathway in Arabidopsis thaliana are likewise important for other secondary metabolite pathways. As a result, the knowledge we obtain from researching these pathways in Arabidopsis thaliana may be used to better understand diseases in other species.

# CHAPTER- 3 REVIEW OF LITERATURE

One of the study on model plant Arabidopsis discuss about Arabidopsis thaliana is a model organism that has been extensively utilized in plant biology research. It is a small, rapidly growing plant native to Europe and Asia[1][6]. Arabidopsis has a small genome, which makes genetic study possible. It is also a rather well-studied species, with a wealth of data on its genetics, biochemistry, and physiology. The report contends that studying model organisms from an ecological standpoint can enrich the research. This entails considering how model organisms interact with their surroundings and how these interactions affect their molecular activities. The paper's authors give various examples of how an ecological viewpoint might be used to acquire fresh insights into plant biology[6].

One of the study discuss about the the modification of the phenylpropanoid pathway in plants using metabolic engineering. A significant anabolic system in plants, the phenylpropanoid pathway has multiple branches that are involved in the production of a wide range of specialised metabolites, such as lignin, phenolic acids, curcuminoids, coumarins, stilbenes, and the huge group of flavonoids. These substances play a wide range of physiological roles, including controlling microbial symbiosis, stress tolerance, and growth and development. The potential of metabolic engineering to alter the phenylpropanoid pathway for plant defence is then covered by the authors. They point out that phenylalanine ammonia lyases (PALs), important enzymes in the route that produce a variety of antibacterial chemicals, are a good target for metabolic engineering[7].

Another study related to phenylpropanoid pathway in Arabidopsis thaliana shows the function and importance of pathway and also various derivates formed. The phenylpropanoid pathway is an important metabolic system in plants that produces a wide range of chemicals such as lignin, flavonoids, and coumarins. These chemicals provide a variety of roles, including structural support, pathogen and herbivore defense, and signalling in response to environmental stimuli. The pathway begins when phenylalanine is converted to cinnamic acid. Cinnamic acid is subsequently broken down into a number of other chemicals, including coumarins, flavonoids, and lignin. The study next goes into how the phenylpropanoid pathway is regulated. A multitude of factors, including environmental cues, developmental signals, and hormones, influence the pathway[2].

Another research on Arabidopsis to investigate the presence of DEGs in response to biotic and abiotic stresses. They performed meta-analysis of transcriptome data from 12 research that studied Arabidopsis thaliana responses to biotic and abiotic stressors. The study's purpose was to find common and distinct genes, pathways, and transcriptome mechanisms involved in plant responses to these stressors. The meta-analysis revealed that 933 genes were significantly differently expressed in

response to biotic and abiotic stressors. In reaction to biotic stresses, 436 genes were upregulated, while 476 were upregulated in response to abiotic stresses. The study's authors discovered many pathways that were frequently altered in response to biotic and abiotic stressors. The phenylpropanoid pathway, the jasmonic acid pathway, the ethylene pathway, and the abscisic acid pathway were among these[8].

A research on RNA Seq tells about the how pipeline works by using Tuxedo protocol. The study describes the usage of HISAT2 and Cufflinks for RNA-seq differential gene and transcript expression analysis. HISAT2 is a read alignment tool for RNA-seq readings that aligns them to a reference genome. Cufflinks is a transcript assembly and quantification tool that assembles transcripts and quantifies their expression using HISAT2 alignments. The study opens with an introduction to RNA- seq and the challenges of differential expression analysis. The authors next go over how HISAT2 and Cufflinks can be used for differential expression analysis. They emphasize the benefits of employing these technologies, such as their capacity to handle high-throughput data and measure transcript expression properly[4].

# CHAPTER- 4 METHODOLOGY

* 1. **RETRIEVAL OF DATA**

Initially, a detailed literature review was done to see about the different pathways involved in cell signalling pathways. The whole raw FastQC sequence was provided as the training datasets for training purpose for the two contrasting phenotypes consisting of 3 cases of case and control along with their replicates by the institution which had to be used for RNA seq purpose. Further the reference genome and the gff file were retrieved from NCBI database through Firefox browser. 2 Fastqz files were downloaded from NCBI.

Metadata

|  |  |
| --- | --- |
| **Sample description** | **Replicate** |
| Case | Case\_Rep1 |
| Case | Case\_Rep2 |
| Case | Case\_Rep3 |
| Control | Control\_Rep1 |
| Control | Control\_Rep2 |
| Control | Control\_Rep3 |

Table 1:Metadata consisting of sample description and replicate of the case and control sample.

# DATA PREPROCESSING

The preprocessing of the data was done through Linux.

# INSTALLATION OF DIFFERENT PACKAGES TO PERFORM RNA SEQ:

Different packages of Tuxedo suite for RNA Seq were installed on Linux system. The downloaded packages are- FASTQC, Trimgalore, HISAT2, SAMTOOLS, CUFFLINKS, CUFFMERGE and CUFFDIFF.

# RUNNING QUALITY CHECK:

A specialized tool called fastqc is used to conduct quality checks and produce fastqc reports that include various parameters. The tool's goal is to produce a quality control report that can identify issues or biases that come from either the sequencer or the initial library content. Fastqc was performed on 12 different datasets to check the quality whether it was good or not. If the files were bad they were

further trimmed using Trimgalore. As a result HTML file was generated in which different types of graphs were present.

**Command line:** /home/ruchi10/Downloads/Example\_file/FastQC/fastqc \*.fastq.gz --noextract -t 6 -o output/

# TRIMMING:

It was required to eliminate low-quality readings from the data in order to do additional research. For eradication of bad quality readings Trimgalore was utilised. TrimGalore is a wrapper script for RRBS sequence files that automates quality and adapter trimming as well as quality control, with the added feature of deleting biased methylation regions (for directional, non-directional (or paired end) sequencing). It employed Cutadapt for trimming purposes. One of the features is the ability to trim sequences if they become too short and fall below a specified threshold. Quality inspection was done on the resultant reduced files once more. Then the HTML files of the files were obtained so that it could be read.

**Command line:** /home/ruchi10/Downloads/TrimGalore-0.5.7/trim\_galore -q 25 -j 3 -- length 36 -- clip\_R1 10 --clip\_R2 10 --three\_prime\_clip\_R1 5 --three\_prime\_clip\_R2 5 --pairedCase1\_R1.fastq.gz Case-1\_R2.fastq.gz -o output

# READ ALIGNMENT:

For this step, a General Featured Format (GFF) reference genome file for Arabidopsis thaliana was downloaded from NCBI. Because it was in gz format, the file components were extracted. The GFF file was extracted as a fna file, which was converted to a fa file using the 'cp' command. HISAT2 was utilized to align readings to the reference genome. HISAT2 is a tool built specifically for alignment. First, an index file was created from the fa file, and then alignment was conducted. The generated alignment file, in the form of a Sequence Alignment Map (SAM), was then transformed to a Binary Alignment Map (BAM) using the samtools utility. The BAM files were then sorted and indexed.

# Command lines: -

**For building hisat2 index files -** path/hisat2-build ref\_fa ref\_index

**For alignment-** path/hisat2 -p 7 --sensitive -x ref\_index -1 Case-1\_R1 \_val\_1.fq.gz Case- 1\_R2\_val\_2.fq.gz -S Case1-RF.sam --rna-strandness RF --dta cufflinks -

**Sam to bam command line -** samtools sort Case1\_RF.sam -o sorted\_Case1-RF.bam

**Sorting bam files -** samtools sort sorted\_Case1-RF.bam -o sorted\_Case1.bam

**Indexing -** samtools index sorted\_Case1.bam

# ASSEMBLING OF TRANSCRIPTS:

Cufflinks were used to compile all transcripts from the six BAM files. Cufflinks is a transcript assembly tool with a variety of capabilities for processing RNA-Seq data. Using these programmes, transcripts are constructed from aligned RNA-Seq reads, and their abundances are estimated as well as examined for differential expression and regulation. Cufflinks contains a number of tools, some of which can be used separately, while others are intended to be used in tandem. Cufflinks also includes cuffmerge and cuffdiff for deeper RNA-Seq research. Each cufflink results folder had four files: gene.fpkm\_tracking, isoforms.fpkm\_tracking, skipped gtf, and transcripts.gtf. Transcripts.gtf was one of these files that was used for additional investigation. The resulting GTF file is an abbreviation for General Transfer Format.

**Command line:** /home/ruchi10/Downloads/ Cufflinks/cufflinks-2.2.1.Linux\_x86\_64/cufflinks -p3 -o cufflink-results/ -g /path/reference genome -b /path/bam file

# MERGING OF TRANSCRIPT ASSEMBLIES:

The transcript assemblies generated with cufflinks were then merged to make a single merged file in gtf format. A path2gtf text file was created with the path to all transcripts gtf files found in each cufflink folder.

**Command line -** python2.7/path/cuffmerge -o merged transcripts/ -g /reference genome file

-p 2 -s path/gtf text file

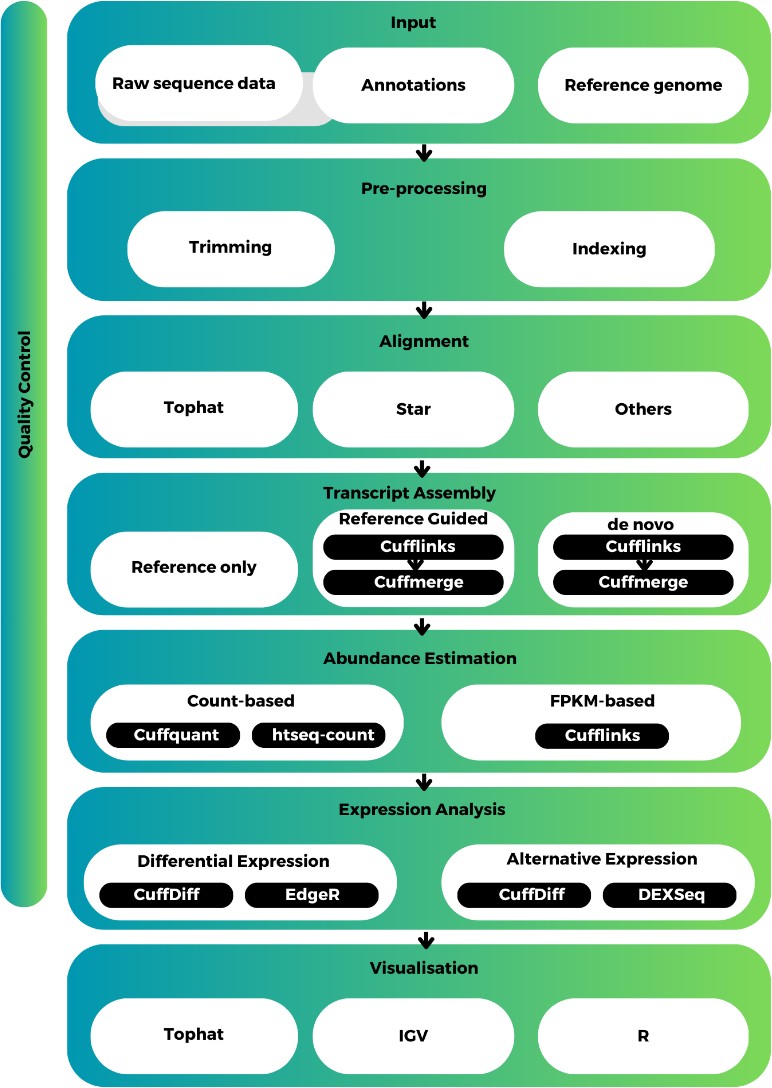


Figure 1: Workflow of RNA Seq

# EXTRACTION OF UNIQUE GENE LIST:

The unique gene list was taken from a merged gtf file created by combining transcript assemblies. To navigate to the merged folder directory, use the 'cd' command. The contents of the merged gtf file were then displayed using the'more' command. The column holding gene names was printed from the contents, and a gene list text file was formed. The created gene list was sorted, and a unique gene list was derived from it.

**Command line –**awk ‘{print $16}’ merged.gtf > genelist.txtmore genelist.txt

sort genelist.txt | uniq > uniq\_genelist.txtmore uniq\_genelist.txt

# MANUAL CURATION OF RESULTING DATA:

Each case and control's unique gene list and gene.fpkm\_tracking file were copied and pasted into an excel sheet. A function in Excel called 'VlOOKUP' was used for manual curation of the data. It is an abbreviation for Vertical Lookup, which is used to find a given value in a specific column. Using this programme, all fpkm values for a certain gene were collected in a single row of various cases and controls.

**Formula**: =VLOOKUP(A2,'Case 1'!E2:J38475,6,FALSE)

# DIFFERENTIALLY EXPRESSED GENES (DEGS):

**iDEP** ([**http://bioinformatics.sdstate.edu/idep11/**](http://bioinformatics.sdstate.edu/idep11/)**):** It stands for Integrated Differential Expression and Pathway Analysis[9]. It is, as the name implies, primarily developed for identifying DEGs along with plots. iDEP analyses differential gene expression in every pair of sample groups determined by parsing sample names[10]. The prepared excel file was initially uploaded, and the necessary parameters were set accordingly. Essentially, there are two major parameters that must be set. The first is the False Discovery Rate (FDR), while the second is the minimum cutoff. The rate at which key features become truly null is referred to as the FDR. According to the data, it was set at 0.7 and the minimum cutoff was set at 4.

# 4.2.9 ENRICHMENT AND ANNOTATION ANALYSIS:

**Metascape (**[**https://metascape.org/gp/index.html#/main/step1**](https://metascape.org/gp/index.html%23/main/step1)**):** is a gene annotation and enrichment analysis tool for evaluating pathways and processes. Enrichment refers to the genes that are either upregulated or downregulated in the pathways shown. To find prospective therapeutic targets, the bioinformatics program Metascape combines information from several sources, such as gene expression, protein-protein interactions, and genetic variation. It performs enrichment and annotation analysis[11].

In metascape, the curated gene list was copied and pasted. The list was subjected to a custom analysis. The parameters were then set. The Min Overlap was set to 2, the P value was set to 0.05, and the Min Enrichment was set to the default value of 1.5. The Analysis Report page was downloaded, and enriched pathways or processes were displayed on that page. All of the outcomes were documented and evaluated. After enrichment, Annotation was performed where Gene description, Symbol, ID and Kegg pathway were obtained and then the Analysis Report page was downloaded where excel sheet was obtained in which the phenylpropanoid pathway was searched.

# CHAPTER-5 RESULTS

* 1. **POST QUALITY CHECK**

As the raw fastq sequence reads were obtained and then FASTQC was applied on them, different graphs were generated which gave a summary of the quality of the reads.

The graphs generated were:

* + - **Basic Statistics -** Basic Statistics will give the summary of the contents in the file like Filename, File type, Encoding, number of total sequences, no of sequences that are flagged as poor quality, minimum and maximum length of the sequences and percentage of GC content in the sequence provided.
    - **Per base sequence quality -** This plot depicts the distribution of base quality scores across the entire read sequence per base sequence. A high average base quality score indicates a high quality read.
    - **Per tile sequence quality -** The "Per Tile Sequence Quality" graph in the FastQC report shows the quality scores of sequencing reads across different tiles of the flow cell. The X-axis represents the tile number, which corresponds to the physical position of the read on the flow cell. The Y-axis represents the average quality score per tile. A good per tile sequence quality graph will show a uniform distribution of colors, with no tiles showing consistently poor quality.
    - **Per sequence quality scores -** This plot depicts the distribution of average base quality ratings for each read. A high average base quality score indicates a high quality read.
    - **Per base sequence content -** This graph depicts the distribution of all four bases in a sequence.
    - **Per sequence GC content -** Per Base GC Content plots out the GC content of each base position in a file.
    - **Per base N content -** This graph plots out the percentage of base calls at each position for which an N was called.
    - **Sequence length distribution –** This plot depicts the distribution of read lengths in the dataset. A high-quality dataset will have a consistent distribution of read lengths.
    - **Sequence duplication levels -** This module counts the degree of duplication for every sequence in the set and creates a plot showing the relative number of sequences with different degrees of duplication.
    - **Overrepresented sequences -** This particular graph lists all of the sequence which make up more than 0.1% of the total.
    - **Adapater content -** This plot depicts the distribution of adapter sequences in the dataset. Short sequences that are introduced to the ends of reads during library preparation are known as adapter sequences.

Trimgalore is used to eliminate the low-quality readings and generate good quality reads.

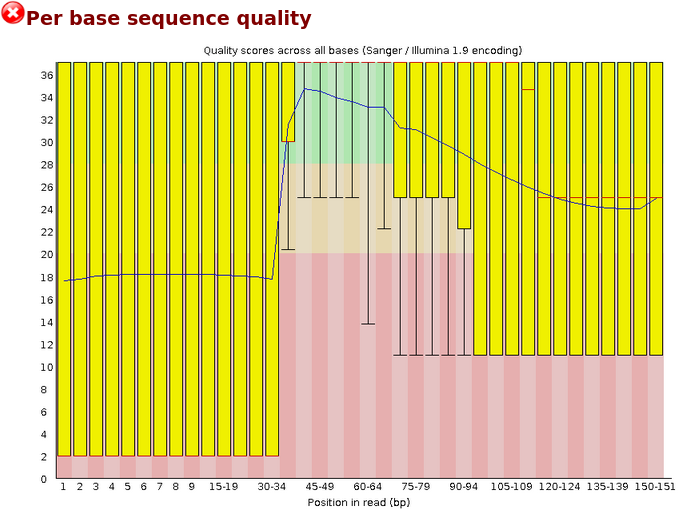


Figure 2: A graph from the Output of FastQC report

Here we can see the blue line representing mean quality is going through the red region, then it falls into the green region depicting good quality region but as it was not completely in the green region therefore bases that are lying in green area are of good quality and ones that are lying below this region needs to be discarded.

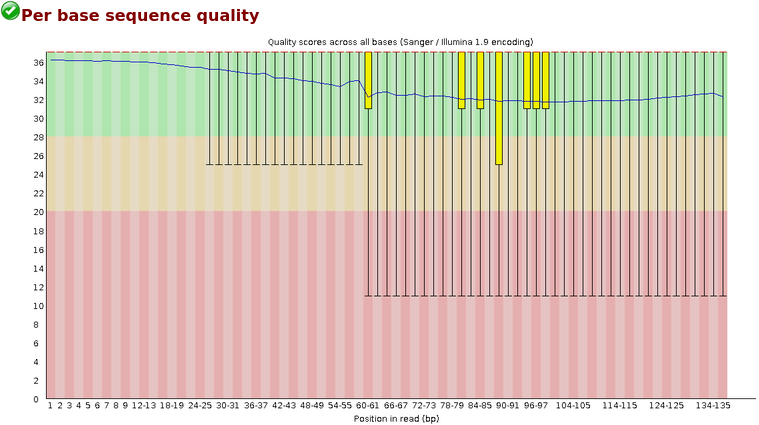


Figure 3: A graph from Output of Trimgalore report

The mean quality improved much after trimming. The blue line is now falling in green region which means that the quality of the reads has been good now. Sequence quality per base has become good.

# DIFFERENTIAL EXPRESSION ANALYSIS:

* + - After sorting the unique genes and fpkm values of case and control conditions on excel sheet, the file was saved in .csv format.
    - The .csv file was uploaded on iDEP with data type as the normalized expression data and differential gene expression analysis was performed and species was also selected.
    - Firstly, pre-processing was done where the minimum level was kept as 1000 and number of libraries as 1 by default.
    - In our data, 26175 genes are present in 6 samples, 823 genes were filtered out, 530 genes were converted to Ensembl gene IDs in our database. The remaining 293 genes were kept in the data using original IDs.

After pre-processing, we get following graphs:

**Bar Plot:** It shows the total read counts of the data. A bar plot is commonly used in iDEP to illustrate gene expression levels or differential expression patterns across circumstances or groups. It displays the expression levels or fold changes for a set of genes visually.

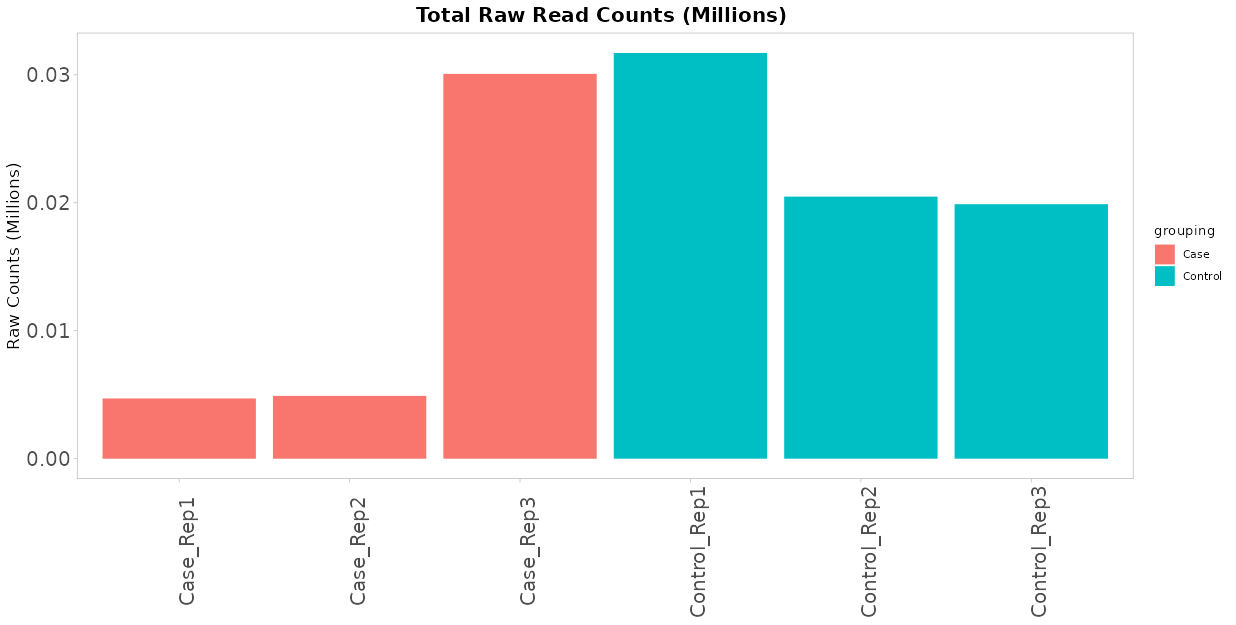
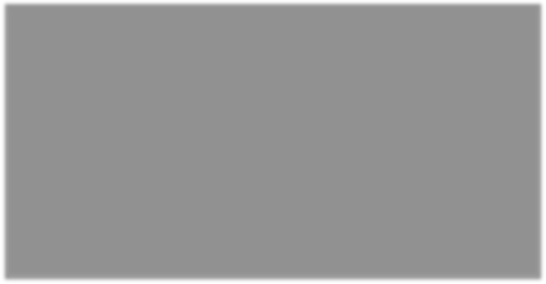


Figure 4: Bar plot of case and control samples

Here we can we 6 samples, 3 of case and 3 of control. A large number of variations are seen between case and control conditions. Looking at the graph, it is interpreted that control conditions have higher raw counts.

**Box Plot:** A box plot is a visual representation that shows the distribution and change of gene expression values or differential expression patterns in RNA-seq data across different circumstances or groups. It shows data’s interquartile range in box, median which is the line inside the box and the outliers that points outside the whiskers.

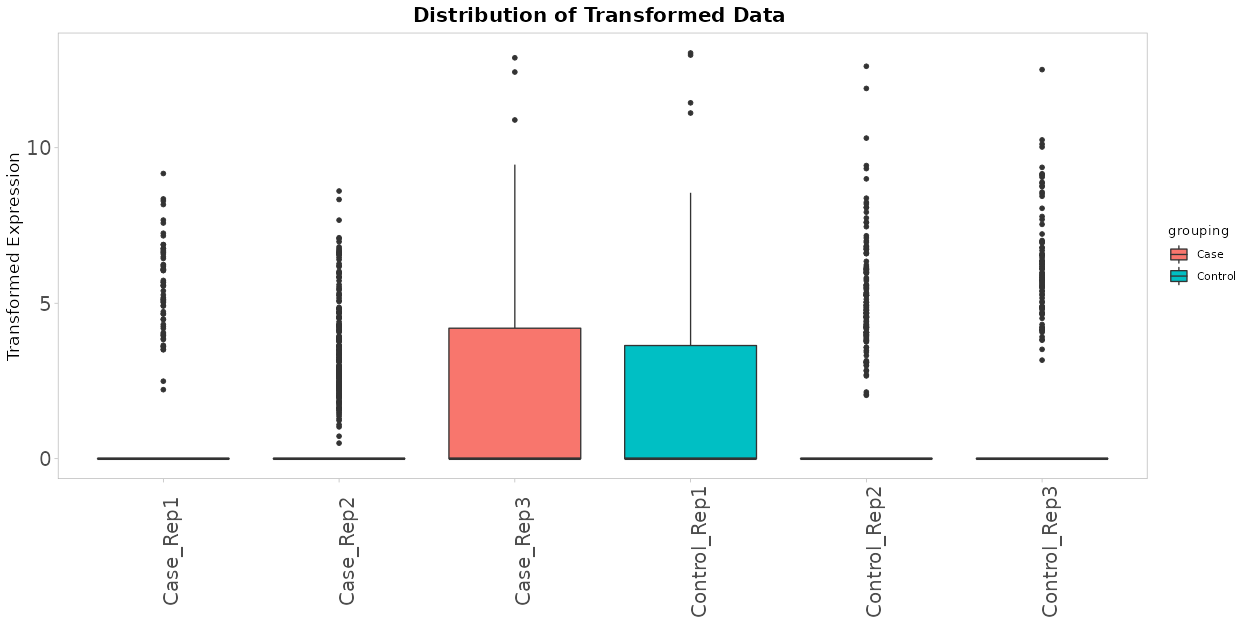
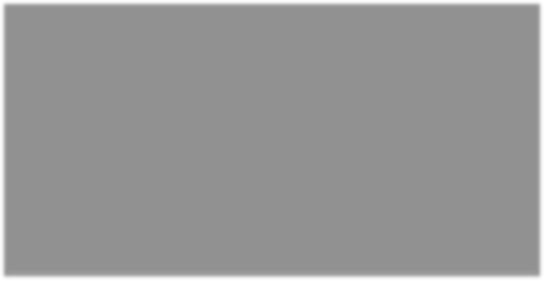


Figure 5: Box plot of case and control samples

Here we can see that case3 have highest box which indicates higher spread of data. Here except for case3 and control1 there is no such median value or interquartile value but they have a lot of outliers.

**Density Plot:** A density plot is a visualisation technique that is used to illustrate the density or distribution of gene expression values or differential expression patterns in RNA-seq data across various conditions or groups. It gives information on the probability density of expression values or differential expression scores.

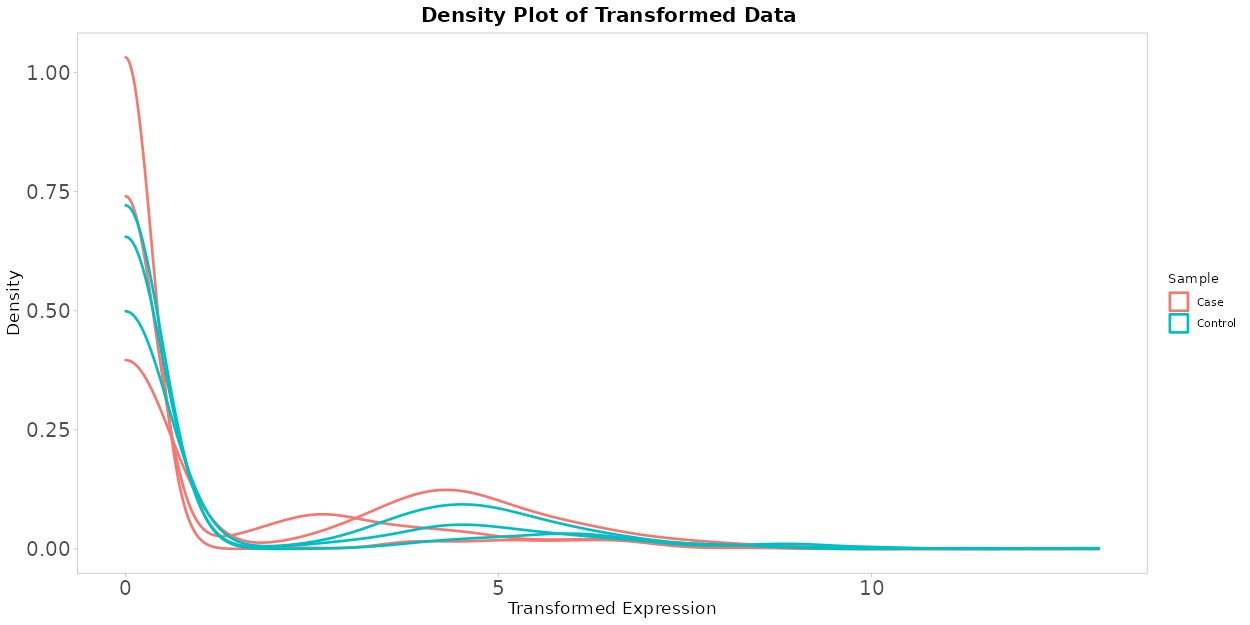
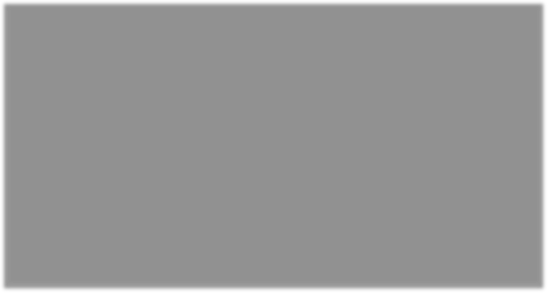


Figure 6: Density plot of case and control samples

* After pre-processing, we went to Clustering where Heatmap is generated

**Clustering:** In the clustering step, various clustering algorithms are applied to the pre-processed gene expression data to group genes or samples based on their similarity in expression patterns. The clustering procedure helps identify gene clusters or groups that have similar expression patterns across samples.

After Clustering Heatmap is generated.

**Heatmap:** The heatmap shows graphically how gene expression levels vary between samples. Since it uses colour coding to indicate the expression values, you may observe patterns of gene expression and compare the relative expression levels between samples. Heatmaps can be used to identify gene clusters with comparable co-expression patterns or expression profiles.

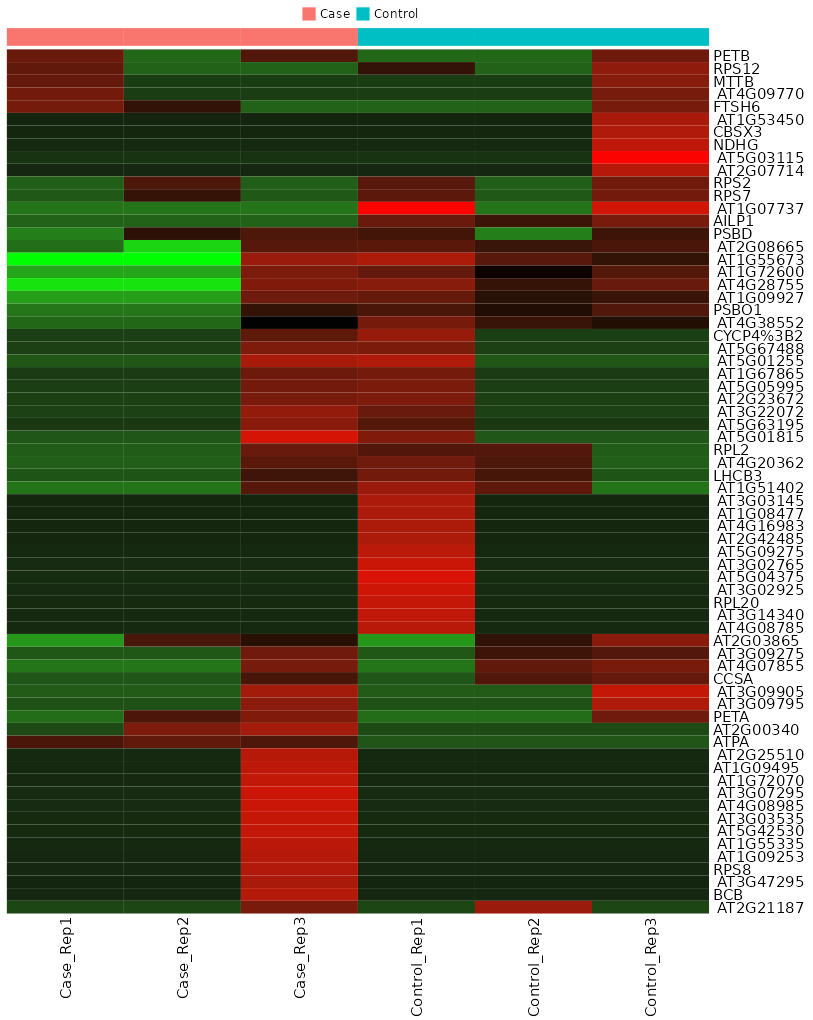
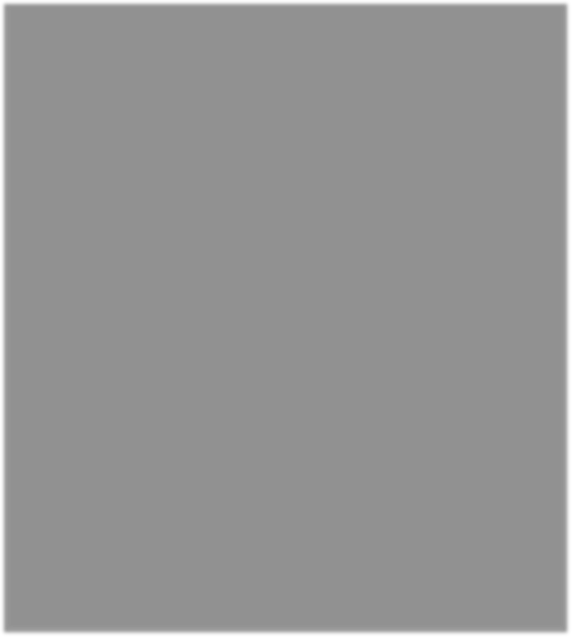


Figure 7: Heatmap of Arabidopsis thaliana's top 100 genes. Genes that are upregulated are shown in Red, downregulated genes are shown in Green, and genes that are unchanged are shown in Black.

**PCA:** Principal component analysis (PCA) is a popular technique for studying large datasets, enhancing data interpretation while maintaining the most information, and enabling the presentation of multidimensional data. PCA has a high number of dimensions or features per observation.

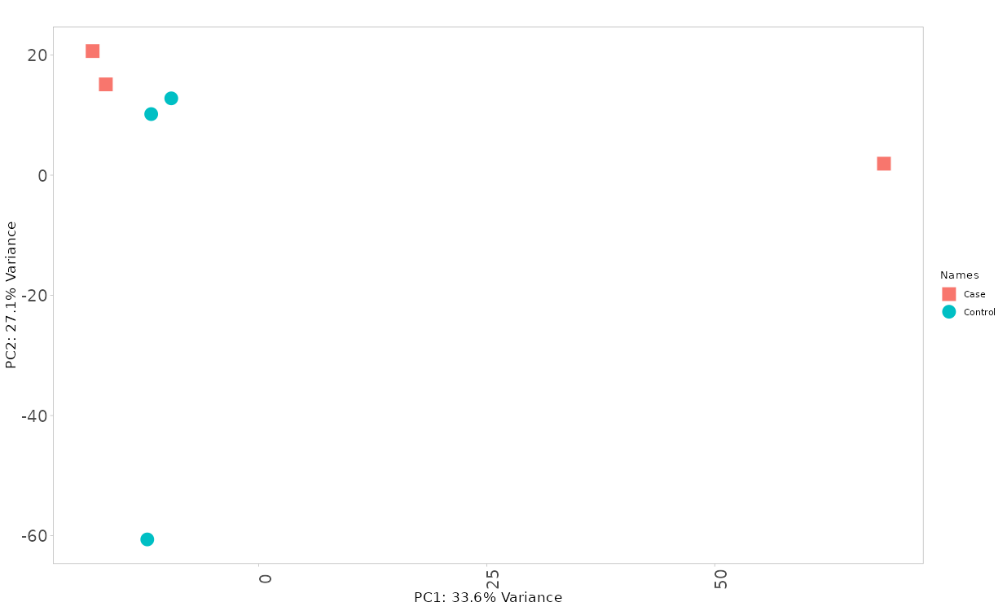
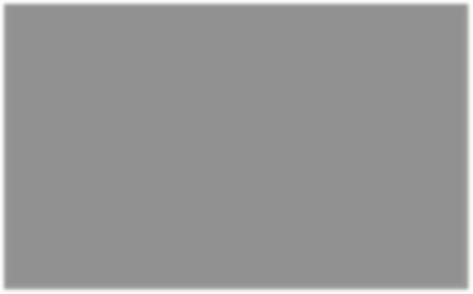


Figure 8(a): PCA Non-interactive plot of case and control samples

Here we can see that case and control condition are close to each other that represent that they have similar gene expression level. One red and blue dot are far that represents unique gene expression.

# PCA 3D Graph:

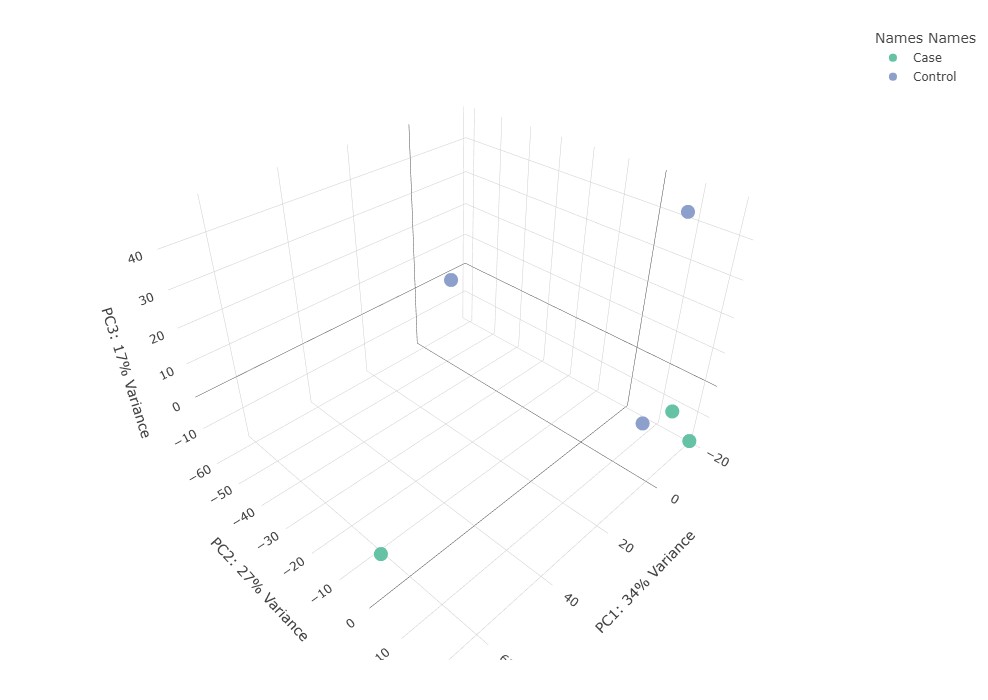
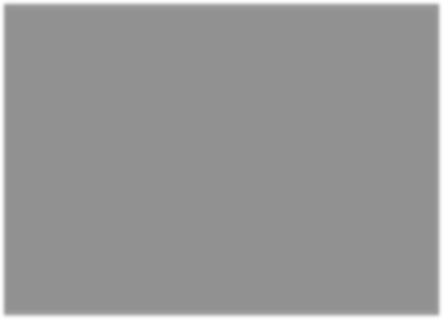


Figure 7(b): PCA 3D plot (left orientation) of case and control samples

# Differentially Expressed Genes:

This plot below shows the number of upregulated and downregulated genes which are basically the differentially expressed genes.

Here the FDR cutoff value was kept as 0.7 and min fold change value as 4.

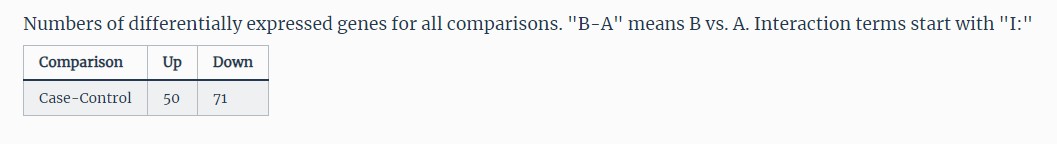
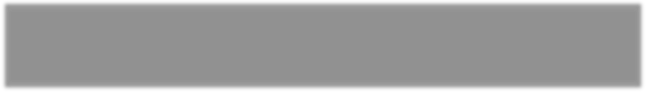
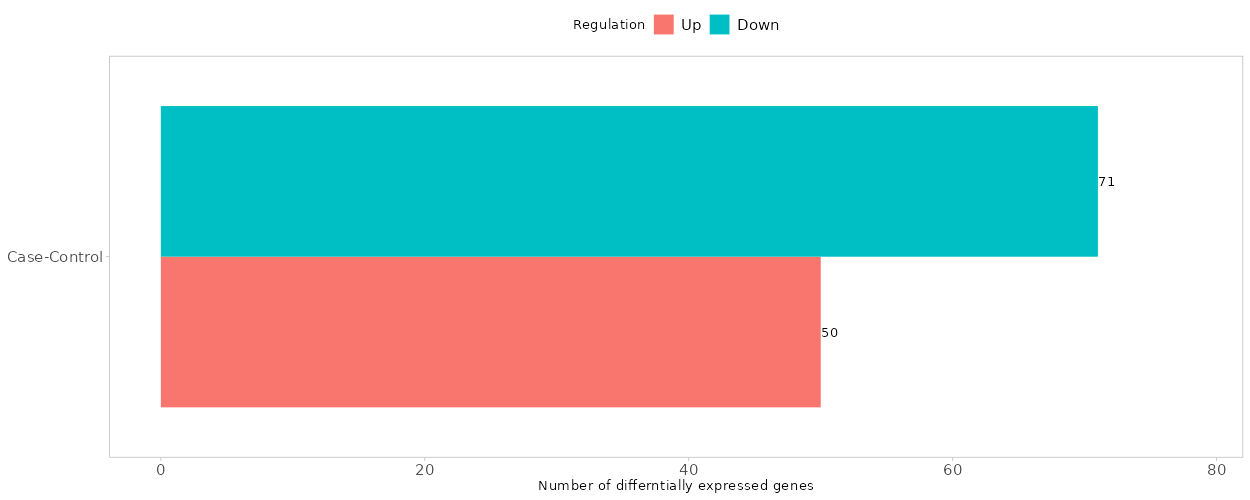
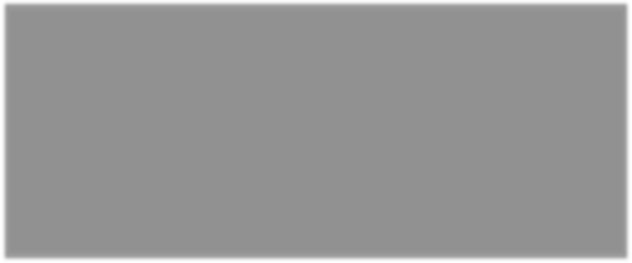


Figure 8: Output showing Differentially expressed genes considering FDR cutoff 0.7 and minimum fold change 4 of case and control samples

Here we can we see that 71 genes are downregulated and 50 genes are upregulated. It indicates that more genes are getting downregulated that shows they are differential expression of the genes.

# Venn Diagram:

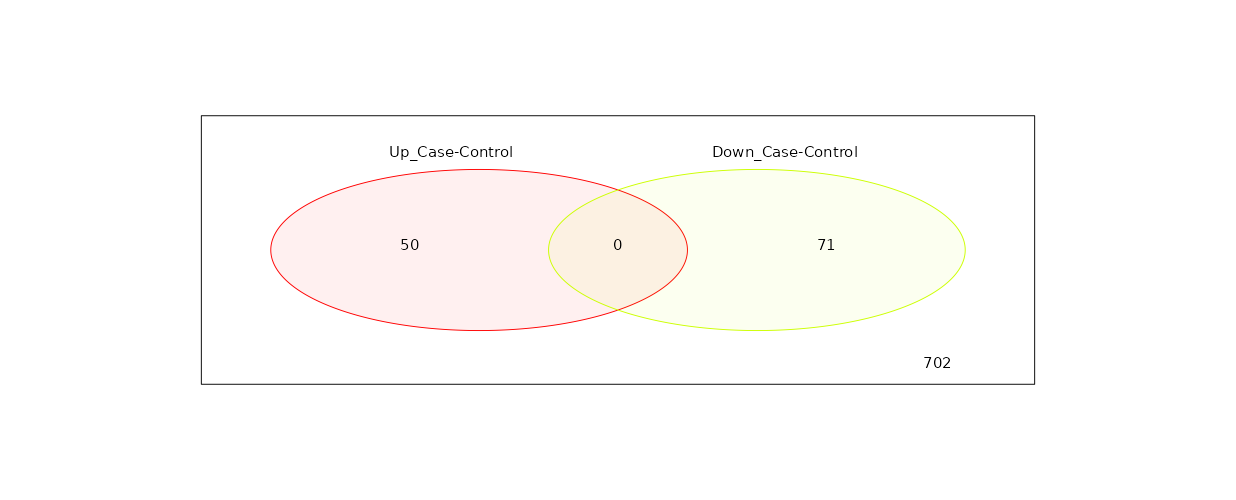
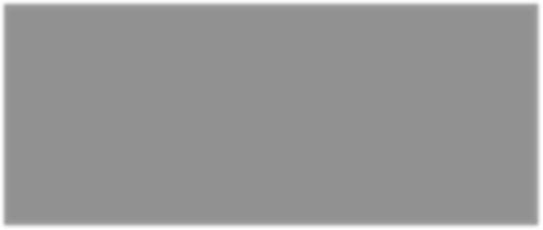


Figure 9: Venn Diagram plot of DEGs of case and control samples

# Upset Plot:

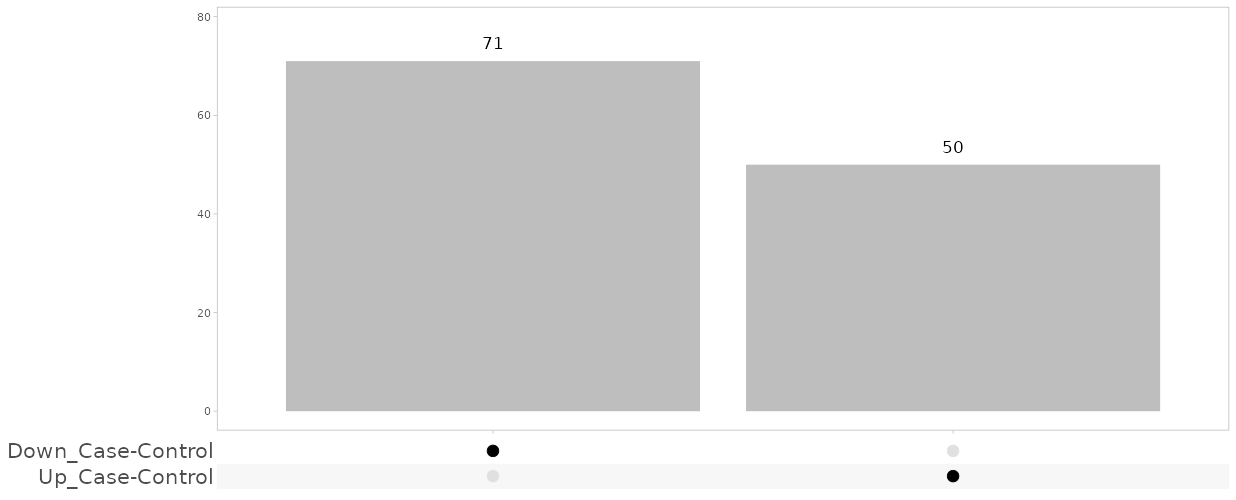
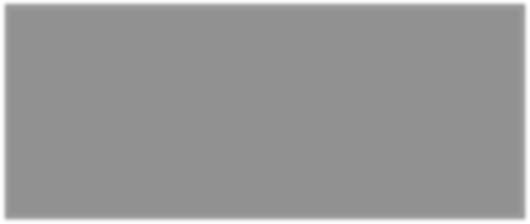


Figure 10: Upset plot of DEGs of case and control samples

**Volcano Plot:** In genomics and transcriptomics, a volcano plot is a popular visualisation tool for displaying the results of differential expression research. It combines statistical significance (often expressed as the negative logarithm of the p-value) and the fold change in gene expression between two conditions or groups.

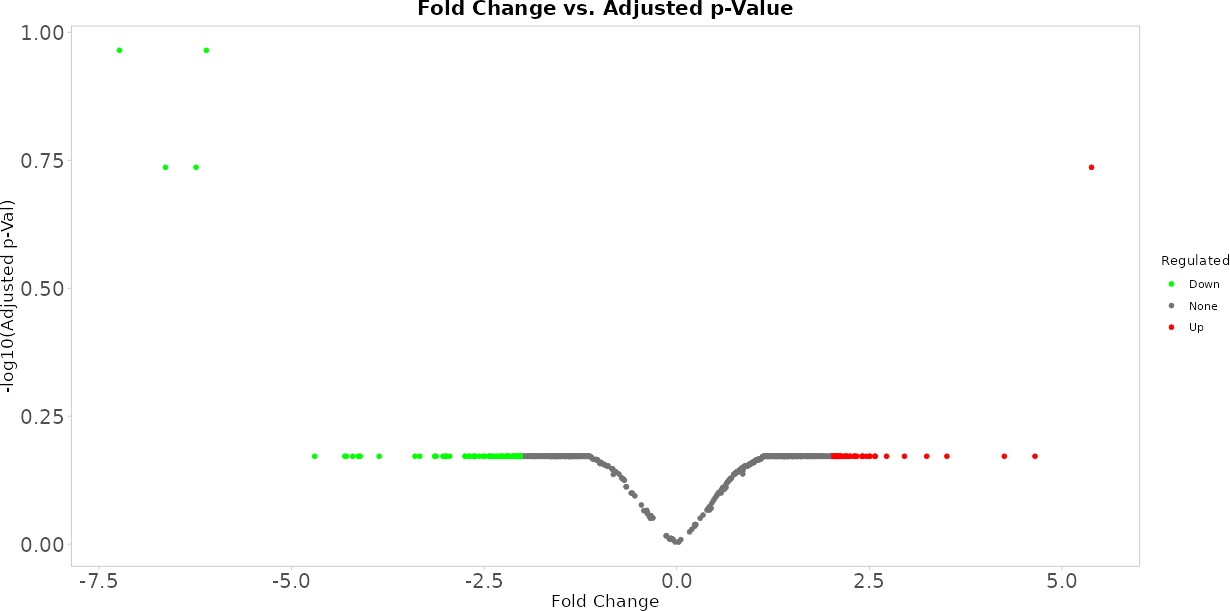


Figure 11: Volcano plot of DEGs of case and control samples

It represents three colours for dataset. Green colour indicates group of downregulated genes. Red colour indicates upregulated genes. Grey colour indicates genes that are not regulated.

**MA plot:** A MA plot is a visual depiction of genetic data using a Bland-Altman plot. By converting the data to M (log ratio) and A (mean average) scales and then showing these values, the figure illustrates the differences between measurements made in two samples.

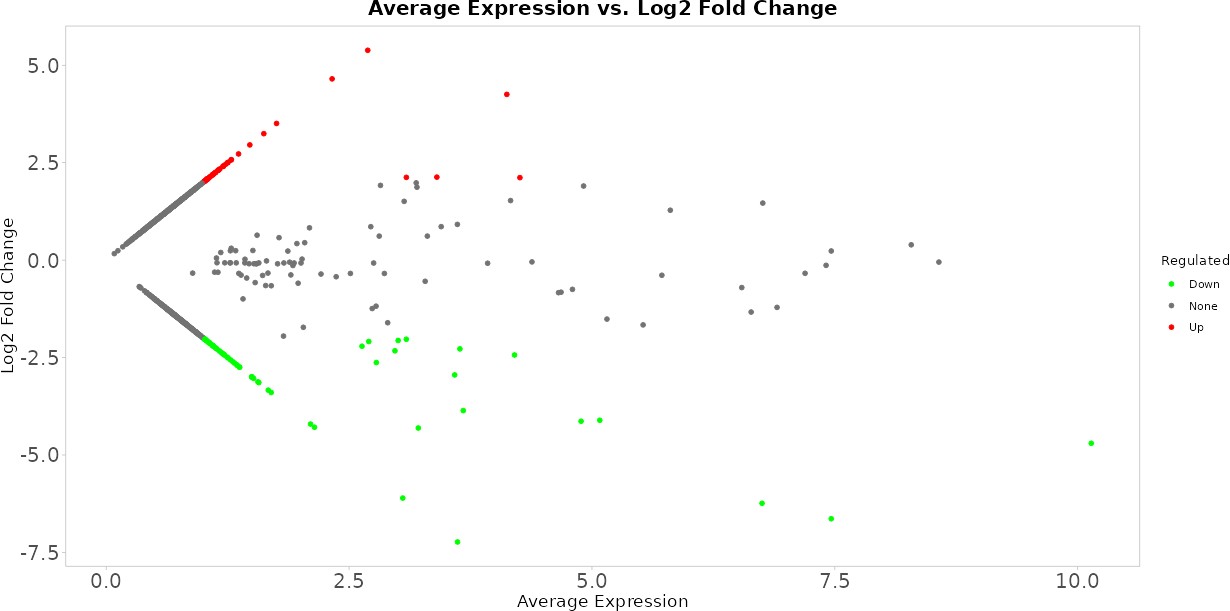


Figure 12: MA plot of DEGs of case and control samples

The graph shows the difference between the measurements of the upregulated and downregulated genes.

**Scatter plot:** Dots are used in a scatter plot to show the values of two separate numerical variables. Each dot's location on the horizontal and vertical axes represents a data point's values. To view relationships between variables, utilize scatter plots.

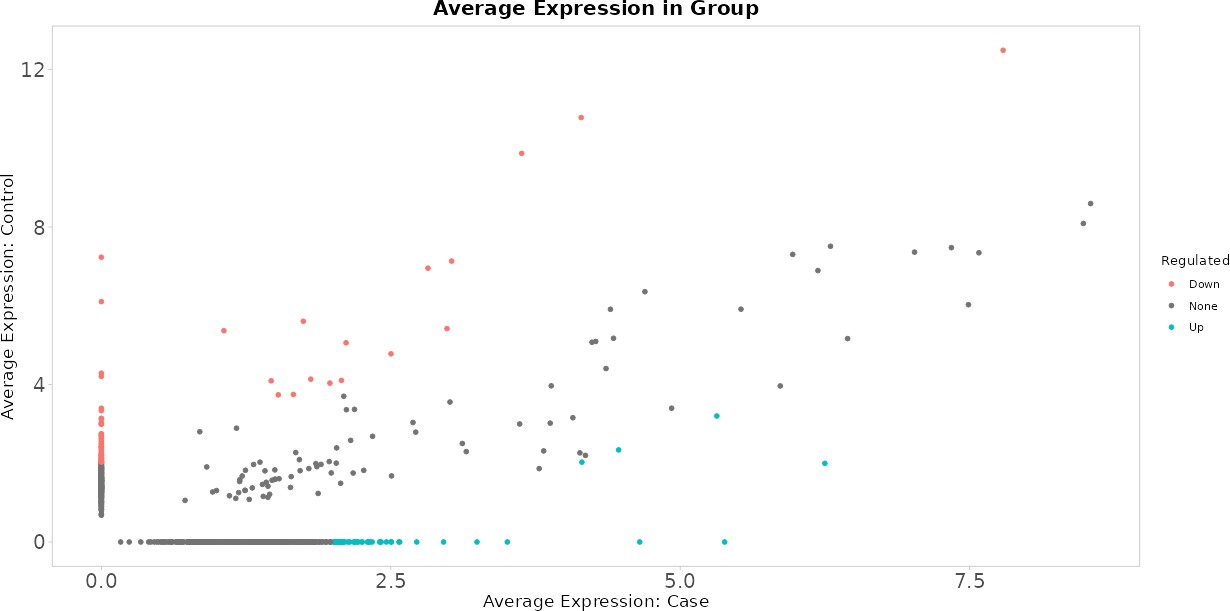


Figure 13: Scatter plot of DEGs of case and control samples

The blue and red dots indicate the expression values of the upregulated and downregulated genes respectively.

# ENRICHMENT ANALYSIS

1. **GO Molecular process:**

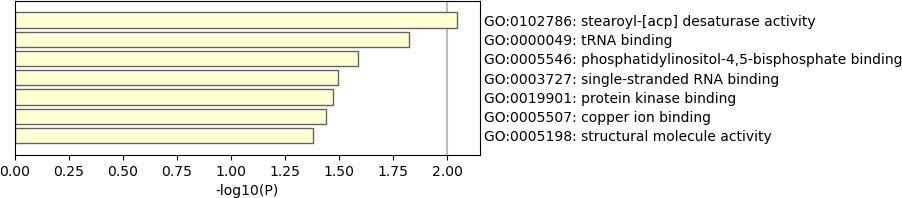


Figure 14: Bar graph of enriched terms across input gene lists, colored by p-values

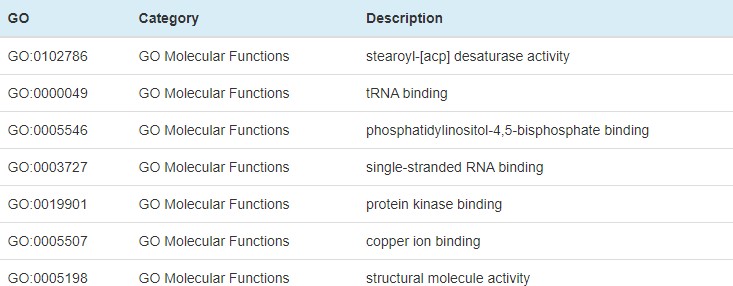


Table 2: Top 7 clusters with their representative enriched terms

# GO Biological process:

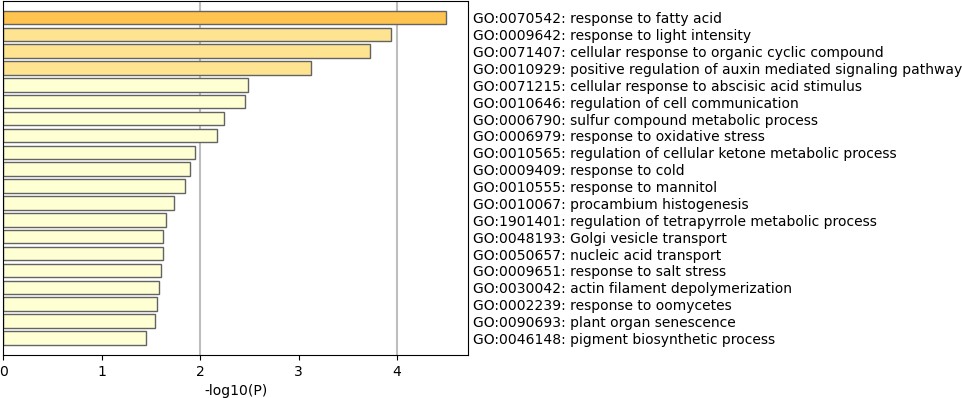


Figure 15: Bar graph of enriched terms across input gene lists, colored by p-values

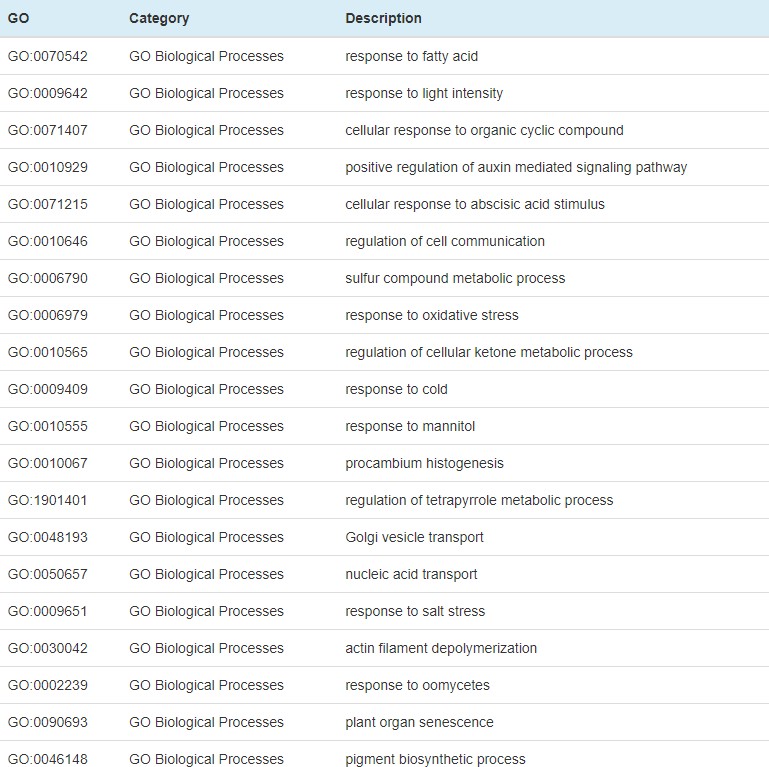


Table 3: Top 20 clusters with their representative enriched terms

# KEGG Pathway:

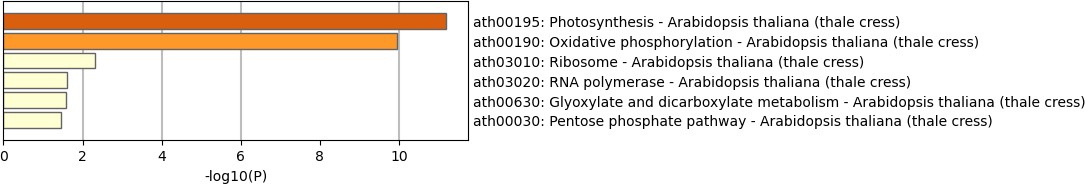


Figure 16: Bar graph of enriched terms across input gene lists, colored by p-values

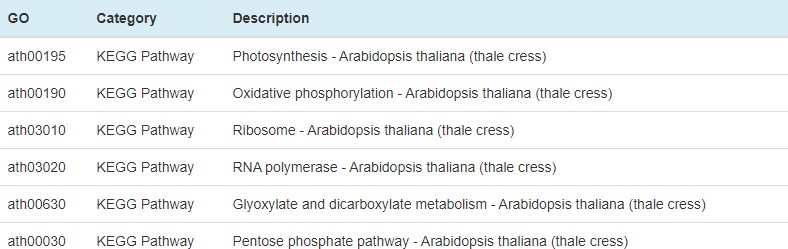
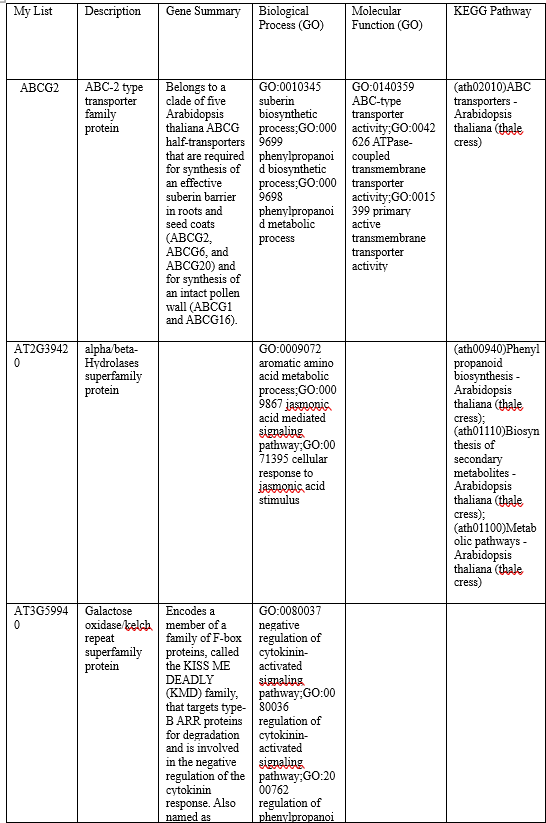
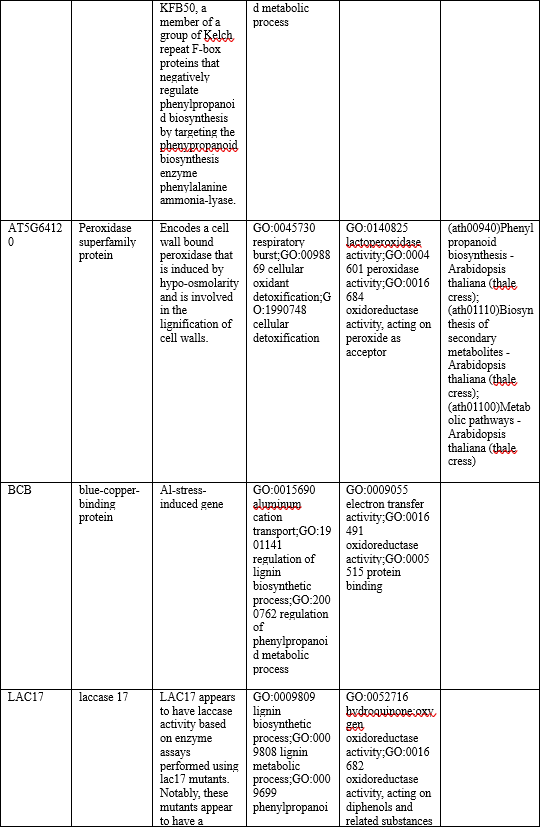


Table 4: Top 6 clusters with their representative enriched terms

# ANNOTATION

The unique genes were put through an annotation process and an excel sheet was generated which gave certain information about the unique genes. Those were like Gene ID, Tax ID, Gene Symbol, Description, Gene location and many more. From the table below we can see that there are unique genes which are differentially expressed in phenylpropanoid pathway.





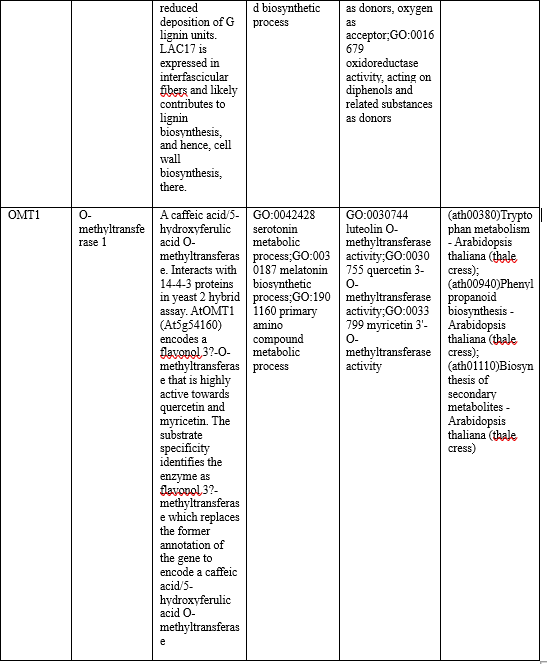


Table 5: List of annotated DEGs in phenylpropanoid pathway

The comparison is made on the basis of enrichment result and the genes present in phenylpropanoid pathway obtained after annotation.

The ABCG2 gene is present in Arabidopsis thaliana photosynthesis. It belongs to the ATP- binding cassette (ABC) transporter family. It was found in 2014 research that ABCG2 which plays a role in phenylpropanoid pathway is required for photosynthesis. ABCG2 encodes a transporter which is involved in synthesis of suberin barrier in roots and seed coat. As a result, ABCG2 may influence photosynthetic efficiency through affecting the plant's water balance and stress tolerance.

According to a 2012 research, AT2G39420 which plays a role in phenylpropanoid pathway is required for oxidative phosphorylation as encodes a monoacylglycerol lipase (MAGL8) involved in galactolipid breakdown in the chloroplast membrane. As a result, AT2G39420 may influence oxidative phosphorylation via modifying galactolipid metabolism in the chloroplast and influencing the plant's photosynthetic efficiency.

Role of ABCG2, BCB, LAC17, OMT1 genes role are not clear in oxidative phosphorylation.

A research 2018 discuss that AT3G59940 which plays a role in phenylpropanoid pathway also encodes a copper transporter (COPT6) involved in high-affinity copper ion absorption at the plasma membrane. Copper ions are crucial micronutrients for plants because they participate in a variety of enzymatic activities, including electron transport, free radical scavenging, hormone perception, and signalling. As a result, AT3G59940 may influence copper ion binding via affecting copper homeostasis and distribution in plants.

A research 2012 discuss that AT5G64120 which plays a role in phenylpropanoid pathway also encodes a copper chaperone (CCH) involved in the transport of copper ions to specific target proteins such as SOD and ETR1. These proteins require copper ions as cofactors because they participate in oxidative stress defence and ethylene signalling, respectively. As a result, AT5G64120 may influence copper ion binding in plants via altering copper utilisation and activation of cuproproteins.

ABCG2 encodes a half-transporter involved in the formation of a functional suberin barrier in roots and seed coverings. Cold stress induces the expression of ABCG2, implying that suberin biosynthesis is part of the plant's adaptation to low temperatures.

# CHAPTER-6 DISCUSSION

After carrying out thorough research on the phenylpropanoid pathway of two contrasting phenotypes of Arabidopsis thaliana through RNA Seq, we found a number of differentially expressed genes of which 50 genes were upregulated and 71 were downregulated. After annotation, 7 genes were found which are related to phenylpropanoid pathway.

After conducting our own research, we discovered that studies have been done on the relationship between these genes and various processes. Genes OMT1, AT2G39420, AT5G64120 are present in phenylpropanoid pathway. For example, it was found in a research that ABCG2 gene is present in Arabidopsis thaliana photosynthesis which may influence the photosynthetic efficiency of the plant[12]. According to a 2012 research, AT2G39420 which plays a role in phenylpropanoid pathway is required for oxidative phosphorylation as encodes a monoacylglycerol lipase (MAGL8) which may affect oxidative phosphorylation[13]. Genes AT3G59940 and AT5G64120 codes for COPT6 and CCH respectively which are involved in high-affinity copper ion absorption at the plasma membrane and transport of copper ions to specific target proteins may influence copper ion binding in plants[14][15]. A research also links that may be Cold stress induces the expression of ABCG2, implying that suberin biosynthesis is part of the plant's adaptation to low temperatures[12].

Studying the genetic basis of phenylpropanoid pathway will allow us to determine specific genes that may be improved to increase agricultural attributes like yield, drought resilience, and disease resistance. As a result, new crop types that are more productive and resistant to environmental pressures may be created.

# CHAPTER-7 CONCLUSION

In conclusion, the discovery of phenylpropanoid pathway by RNA Sequencing in two distinct phenotypes of Arabidopsis thaliana is a significant area of study with broad implications for genetics, evolutionary biology and medicine. We can create novel therapies for diseases and learn more about various species evolve and adapt to their environments by comprehending how phenylpropanoid pathway impact gene expression.

Research on Arabidopsis thaliana can help us understand the genetic basis of phenylpropanoid pathway in other plants, like agricultural plants. RNA sequencing (RNA seq) is a powerful technique for identifying the genes involved in various signalling pathways. By contrasting the levels of gene expression in diverse tissues or environmental contexts, RNA seq can determine whether genes are activated or repressed in response to a certain stimulus. Understanding the operation and regulation of phenylpropanoid pathway may be done using this knowledge. As the technology develops, we might expect seeing a lot more RNA seq applications in this area.

# REFERENCES

1. David W. Meinke, et al., Arabidopsis thaliana: A Model Plant for Genome Analysis.*Science***282**,662-682(1998).DOI:[10.1126/science.282.5389.662](https://doi.org/10.1126/science.282.5389.662)
2. Christopher M. Fraser and Clint Chapple "The Phenylpropanoid Pathway in Arabidopsis," The Arabidopsis Book 2011(9), (1 December 2011). <https://doi.org/10.1199/tab.0152>
3. <https://www.kegg.jp/pathway/ath00940>
4. Chung, M., Bruno, V.M., Rasko, D.A. *et al.* Best practices on the differential expression analysis of multi-species RNA-seq. *Genome Biol* **22**, 121 (2021). <https://doi.org/10.1186/s13059-021-02337-8>
5. Trapnell, Cole; Roberts, Adam; Goff, Loyal; Pertea, Geo; Kim, Daehwan; Kelley, David R; Pimentel, Harold; Salzberg, Steven L; Rinn, John L; Pachter, Lior (2012). *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. , 7(3), 562–578.* doi:[10.1038/nprot.2012.016](file://localhost/C:/Users/hp/Downloads/10.1038/nprot.2012.016)
6. Ute Krämer(2015) The Natural History of Model Organisms: Planting molecular functions in an ecological context with *Arabidopsis thaliana eLife* 4:e06100. <https://doi.org/10.7554/eLife.06100>
7. Yadav, V.; Wang, Z.; Wei, C.; Amo, A.; Ahmed, B.; Yang, X.; Zhang, X. Phenylpropanoid Pathway Engineering: An Emerging Approach towards Plant Defense. *Pathogens* **2020**, *9*, 312. <https://doi.org/10.3390/pathogens9040312>
8. Biniaz, Y.; Tahmasebi, A.; Afsharifar, A.; Tahmasebi, A.; Poczai, P. Meta-Analysis of Common and Differential Transcriptomic Responses to Biotic and Abiotic Stresses in *Arabidopsis thaliana*. *Plants* **2022**, *11*, 502.

<https://doi.org/10.3390/plants11040502>

1. <http://bioinformatics.sdstate.edu/idep11/>
2. Ge, S.X., Son, E.W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, 534 (2018). <https://doi.org/10.1186/s12859-018-2486-6>
3. [https://metascape.org/gp/index.html#/main/step1](https://metascape.org/gp/index.html%23/main/step1)
4. Vandana Yadav and others, ABCG Transporters Are Required for Suberin and Pollen Wall Extracellular Barriers in *Arabidopsis* , *The Plant Cell*, Volume 26, Issue 9, September 2014, Pages 3569–3588, <https://doi.org/10.1105/tpc.114.129049>
5. Kim, R.J., Kim, H.J., Shim, D. and Suh, M.C. (2016), Molecular and biochemical characterizations of the monoacylglycerol lipase gene family of *Arabidopsis thaliana*. Plant J, 85: 758-771. <https://doi.org/10.1111/tpj.13146>
6. Sanz, A., Pike, S., Khan, M.A. *et al.* Copper uptake mechanism of *Arabidopsis thaliana* high-affinity COPT transporters. *Protoplasma* **256**, 161–170 (2019). <https://doi.org/10.1007/s00709-018-1286-1>
7. Lung-Jiun Shin and others, Copper Chaperone Antioxidant Protein1 Is Essential for Copper Homeostasis , *Plant Physiology*, Volume 159, Issue 3, July 2012, Pages 1099– 1110, <https://doi.org/10.1104/pp.112.195974>

**WEEKLY PROGRESS REPORT**

# AMITY INSTITUTE OF BIOTECHNOLOGY WEEKLY PROGRESS REPORT-1

**Enrolment No.:** A0500322012 **Student Name:** Ruchi Bachal **Program:** MSc Bioinformatics

**Project Title:** Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana.*

# Target set for the week:

To do a thorough reading of previously available literature to get a better understanding of the project and learn about the methodologies that might be used in this investigation.

# Progress/Achievements of the week:

* Read research papers and review articles regarding NGS, RNA-seq and phenylpropanoid pathway.
* Learned RNA-seq technique and prepared a work-flow chart.
* Studied transcriptome analysis and the tools required to perform said analysis and the relevance of this analysis.
* Learned about various computational tools that might help in our investigation like DESeq2, FastQC, edgeR.
* Got to know about Tuxedo Protocol used in RNA-seq and tools like TopHat, Cufflinks, Cuffdiff, CummeRband etc.
* Got familiarized with online web-based-applications like iDEP that will help to analyze the data generated by RNA-seq and Metascape that is a gene annotation application.
* Used provided test sample data to visualize the data generated by RNA-seq analysis along with various plots and graphs like volcano plot, scatter plot, Heat maps in iDEP and Functional enrichment analysis in Metascape. It helped me to understand the good and bad data generated what is to be expected when we carry out the transcriptome analysis.
* Learned about Differential Gene Expression and upregulation and downregulation and their impact on the phenotype.
* Installed Oracle VM virtual box that would help in getting acquainted with Linux.
* Installed Rstudio and additional tools based on R like DESeq2, limma, edgeR and ggplot.

# Future Tasks:

To learn more about R and how to use R Studio. Get familiarized with DESeq2, edgeR etc. Will use sample data to try out the tools. Work further on literature review regarding the project.

# AMITY INSTITUTE OF BIOTECHNOLOGY WEEKLY PROGRESS REPORT-2

**Enrolment No.:** A0500322012 **Student Name:** Ruchi Bachal **Program:** MSc Bioinformatics

**Project Title:** Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana.*

# Target set for the week:

To get familiar with RStudio and Linux system.

# Progress/Achievements of the week:

* Revisited MS Excel and used different formulae that could make our work easy.
* Learned about various computational tools that might help in our investigation like DESeq2, limma, FastQC, ggplot2, edgeR. Installed these packages in RStudio.
* Installed various packages in Linux in the previously installed Oracle VM. The packages: - FastQC, Trimgalore, Hisat2, Samtools, Cufflinks, Cuffdiff and Cuffmerge.
* Used provided test sample data downloaded from EAM (European Nucleotide Archive) to get familiar with the above-mentioned packages.

# Future Tasks:

Step wise analysis of the pipeline on the dataset using Linux.

# AMITY INSTITUTE OF BIOTECHNOLOGY WEEKLY PROGRESS REPORT-3

**Enrolment No.:** A0500322012 **Student Name:** Ruchi Bachal **Program:** MSc Bioinformatics

**Project Title:** Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana.*

# Target set for the week:

To perform stepwise analysis of the pipeline on the provided dataset using Linux.

# Progress/Achievements of the week:

* + Obtained the datasets of *Arabidopsis thaliana* on which the investigation must be done.
  + The data was analysed using FastQC and analysis reports were obtained.
  + The data was trimmed as per the requirement using Trimgalore in Linux.
  + Allignment was done using HISAT2.
  + SAM files were generated for further investigation.

# Future Tasks:

To find the Differentially Expressed Genes.

# AMITY INSTITUTE OF BIOTECHNOLOGY WEEKLY PROGRESS REPORT-4

**Enrolment No.:** A0500322012 **Student Name:** Ruchi Bachal **Program:** MSc Bioinformatics

**Project Title:** Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana.*

# Target set for the week:

To perform the cufflinks package on provided dataset using Linux.

# Progress/Achievements of the week:

* + BAM files were generated from the SAM file.
  + Cufflinks was performed on the BAM file and transcripts file in GTF format was obtained.
  + Cuffmerge was performed on the transcripts file and merged transcripts file was obtained.
  + Cuffdiff was performed on the merged file and different files related to differential expression were generated.

# Future Tasks:

To analyze the differentially expressed genes using iDep.

# AMITY INSTITUTE OF BIOTECHNOLOGY WEEKLY PROGRESS REPORT- 5

**Enrolment No.:** A0500322012 **Student Name:** Ruchi Bachal **Program:** MSc Bioinformatics

**Project Title:** Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana.*

# Target set for the week:

To analyse the unique genes through iDep.

# Progress/Achievements of the week:

* + An error was found in the gff file, therefore performed all the steps again from FastQC till Cuffmerge.
  + The gene list obtained was manually curated.
  + The gene list was uploaded on iDep software and different graphs and results for the same were obtained.

# Future Tasks:

To find the differentially expressed genes in the phenylpropanoid pathway through Metascape software.

# AMITY INSTITUTE OF BIOTECHNOLOGY WEEKLY PROGRESS REPORT-6

**Enrolment No.:** A0500322012 **Student Name:** Ruchi Bachal **Program:** MSc Bioinformatics

**Project Title:** Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana.*

# Target set for the week:

To enrich the differentially expressed genes and annotate them using Metascape.

# Progress/Achievements of the week:

* + The unique gene list was uploaded on Metasacpe and Enrichment analysis was done.
  + The metascape gave us the GO molecular, biological process and KEGG pathways.
  + Then the unique gene list was annotated and information about the genes was obtained.

# Future Tasks:

* + Draft report will be prepared.

|  |
| --- |
| **AMITY UNIVERSITY**  -----UTTAR PRADESH----- |
| **Amity Institute of Biotechnology** |
| Summer Internship |

Student Name – Ruchi Bachal Enrolment Number – A0500322012 Programme – MSc (Bioinformatics)

Name of the organisation and address - Amity Institute of Biotechnology, J-3 Block, Amity University, Sector 125, Noida. 201303

# Guide:

Name – Mr. Rajesh Kumar Jangade Designation – Professor

Contact No. – 9312579601

# PROJECT INFORMATION:

1. **DURATION: (40 Days)**
   1. Date of Summer Internship commencement (05/06/2023)
   2. Date of Summer Internship commencement (14/07/2023)

# TITLE:

Differential transcriptome analysis of phenylpropanoid pathway in two contrasting phenotypes of *Arabidopsis thaliana.*

# OBJECTIVE:

The basic purpose of this study is to understand the application of next generation sequencing techniques in analyzing and studying the transcriptome of the chosen organism *Arabidopsis thaliana*. With the help of RNA seq Analysis, a detailed qualitative and quantitative study of the transcriptome will be performed to determine the enrichment analysis, differential expression of genes and various regulatory mechanisms involved in differential transcriptome of *Arabidopsis thaliana*.

# METHODOLOGY

The dataset will be obtained using next generation sequencing techniques such as illumina to obtain reads which are fragments of the RNA molecules. The reads are then preprocessed to check for the quality of the reads using tools like FASTQC which involves trimming of the sequence, removal of adaptor contaminated reads, etc.

The Tuxedo protocol will be used to analyze the RNA seq data with the help of various tools. The RNA seq reads will be mapped to a reference genome by identifying the unaligned segments using tools like TOPHAT. BOWTIE will be used as an aligner in the Tuxedo protocol which is used for aligning the reads that do not contain the splice junctions. The aligned reads are then taken as an input by CUFFLINKS for transcript assembly and expression levels of the RNA transcript so obtained. Followed by this, CUFFDIFF will be used for differential gene and transcript analysis in different phenotypes of *Arabidopsis thaliana*. The RNA seq data obtained will now be visualized through R Language to generate heat maps, density plots ,PCA, Bar graphs etc. and enrichment analysis is done to get an insight about the Transcriptomic data analyzed with the help of the tuxedo protocol.

# SUMMARY

Differential transcriptome analysis of phenylpropanoid pathway in Arabidopsis thaliana can be understood by the study of changes in transcriptome level specifically related to phenylpropanoid pathway. Phenylpropanoids are derived from phenylalanine which are a group of plant secondary metabolites which provides a variety of functions to plants such as lignin synthesis, structural modifications and signaling molecules. Phenylalanine undergoes deamination to produce cinnamic acid which is then converted to coumaric acid to produce aldehydes and alcohols. These alcohols are known as monolignols, which starts lignin biosynthesis.

RNA seq analysis helps us to identify genes that are upregulated or downregulated in different conditions of *Arabidopsis thaliana.* For example, Phenylalanine ammonia-lyase which catalyses the first step of pathway by converting phenylalanine to cinnamic acid has 4 isoforms PAL1, PAL2, PAL3, PAL4. Studies of the PAL genes have shown that PAL3 is expressed at basal levels in stems., PAL1, PAL2 and PAL4 are expressed at relatively high levels in stems during the latter stages of development with PAL1 expression which is localized to vascular tissue and PAL2 and PAL4 both expressed in seeds. An upregulation of PAL1 in PAL2 mutant suggests that PAL1 is overexpressing in absence or mutation of PAL2 and vice-versa. PAL4 gets upregulated in absence of PAL1 and PAL2.

The hypothesis that PAL1 and PAL2 serves as primary isoforms of PAL, to support this hypothesis an organism phenotype having double mutant of PAL1 and PAL2 is compared with phenotype of wild-type or single mutant of PAL. The analysis revealed significant changes in phenotype of double mutant in comparison of wild type and single mutants.