6-1-2017

Transcriptional response to stressors in arabidopsis thaliana

Aya Abd Elsalam Galal

School of Sciences and Engineering

Transcriptional Response to Stressors in *Arabidopsis thaliana*

A Thesis Submitted to
the Biotechnology Master's Program
in partial fulfillment of the requirements for
the degree in Master of Science

By: Aya Abd Elsalam Mahmoud Galal

Advisor: Dr. Ahmed Moustafa, Associate Professor, Department of Biology, American University in Cairo (AUC)
Advisor: Dr. Magdy Mahfouz, Assistant Professor, Biological and Environmental Sciences & Engineering Division, King Abdullah University of Science and Technology (KAUST)

THESIS DEFENSE

Student Full Name: Aya Abd Elsalam Mahmoud Galal

Student ID: 900082512

Thesis Title: Transcriptional Response to Stressors in *Arabidopsis thaliana*

This thesis has been successfully defended.

Thesis Supervisor:
Name: _____
Title: _____
Department: _____
Institution: _____
Signature: _____Date: _____

First reader:
Name: _____
Title: _____
Department: _____
Institution: _____
Signature: _____Date: _____

Second reader:
Name: _____
Title: _____
Department: _____
Institution: _____
Signature: _____Date: _____

Department Chair:
Name: _____
Title: _____
Signature: _____Date: _____

The American University in Cairo

Transcriptional Response to Stressors in *Arabidopsis thaliana*

A Thesis Submitted by
Aya Abd Elsalam Mahmoud Galal

To the Biotechnology Masters' Program

In partial fulfillment of the requirements for
The degree of Master of Science

Has been approved by

Thesis Committee Supervisor/Chair _____

Affiliation _____

Thesis Committee Reader/Examiner _____

Affiliation _____

Thesis Committee Reader/Examiner _____

Affiliation _____

Thesis Committee Reader/External Examiner _____
(If required by dept.)

Affiliation _____

_____ _____ _____ _____

Dept. Chair/Director Date Dean Date

Acknowledgements

The journey towards this point in time has been filled with tears and laughter, for that I am grateful for the company of my friends, family and colleagues during this journey. Without your continuous support encouragement and help this thesis could never have happened.

First and foremost, I would like to express my sincerest gratitude to my advisor Dr. Ahmed Moustafa for his continuous support during my studies. Your guidance, inspiration, patience, encouragement, and great knowledge helped me throughout the entire process. Through this process, I learned to think more critically, became more detail oriented and learned that no question is a stupid question despite how simple and small it may appear to be. Thank you so much for accepting me as your student and for providing a great environment for me to grow as a young bioinformatics analyst and be on my way to becoming a bioinformatian.

I would like to thank all my current and former colleagues and bioinformatics team mates for creating a great environment to bounce ideas off of each other and for being willing to provide ideas, tips and at times lengthy discussions. To my friends who have made the past few years full of memories and never ending support. Thank you to Dina Atef, Heba Zaza, Sarah Halawa, Dina Hesham, Moustafa Abou Howya and Sara Angela for your continuous cheers, encouragements and motivation. The journey would have been much harder without your support.

Finally I offer my special thanks to my family. My parents for their continuous support and encouragement while pursing my degree and my brothers for always being there when I needed.

ABSTRACT

Stress is defined as any external force that can trigger a defensive response from an organism. In plants, stress is something that has been shown to affect plant reproduction and productivity by activating a defensive response. It can be caused by various things including but not limited to biotic or abiotic conditions such as temperature, drought or salt stress. Exposure to stress leads to the production of various transcriptomes that are governed by signals released as a result of the exposed stress. *Arabidopsis thaliana* is characterized by its inability to tolerate any form of extreme stress and given its status as a model organism it is an ideal candidate to investigate the various effects of stress on plants. By studying the transcriptomes produced by *Arabidopsis thaliana* under different stress conditions, a more well-rounded profile of how plant systems react to different stress conditions is produced.

Experiments were carried out in KAUST by exposing the stress intolerant plant to Pladienolide B; a drug that is known to affect the slicing mechanism, RNA sequencing was used in order to produce the transcriptome profile of the plant in response to the stress over a series of time points. The classic tuxedo protocol for RNA sequencing analysis was used to assemble the transcripts and following differential gene expression analysis by CuffDiff, the R package CummeRbund was used to visualize the results. Functional analysis of the significant differentially expressed genes was carried out using PANTHER.

PANTHER was able to classify 12,646 genes; expressed at after exposure to the treatment for 6 hours, and 10,649 genes; expressed after exposure to the treatment for 24 hours, into functional classes. With around 50% of the differentially expressed gene having catalytic activity and around 25% having binding activity. Further investigation revealed that the alternatively spliced differentially expressed genes were heavily involved in various development and regulatory process that are essential for plant maturation. While a few functionally uncharacterized genes were expressed, some of which may hold valuable information in understanding plant stress response.

This research offers a deeper understanding of how plants are effected by stress through the characterization of the differentially expressed genes. Future investigation of the uncharacterized genes expressed is needed as it may provide deeper insights to the plant stress response.

Table of Contents

LIST OF TABLES AND FIGURES

Chapter 1: Introduction:

1.1 *Arabidopsis thaliana*

Arabidopsis thaliana a small flowering plant commonly known as the thale crest plant belonging to the Brassicaceae family is characterized by a small genome size of 125 mega base pairs and a short life cycle of 6 weeks (*Genome Assembly*, n.d.). It is an annual plant native to northwestern Africa, Asia and Europe; with its general appearance of leaves purple in color, maximum height of around 25 cm and a natural habitat of mostly rocky soils or roadsides, it is widely considered a weed (*Arabidopsis thaliana %93 Overview*, n.d., *Arabidopsis thaliana (thale cress)*, n.d.). Scientists have been researching *A. thaliana* since the 1900s and it has commonly been used as a model system for the study of plant science with an emphasis on development and behavior, genetics and evolution; owing to its relatively small genomic size, its short life cycle and the extensive research and investigation that has been performed on it in recent decades (Arabidopsis Genome Initiative, 2000; Coelho et al., 2007; Garber et al., 2011; *Genome Assembly*, n.d., "TAIR - About Arabidopsis," n.d.; López-Bucio et al., 2007; Meinke et al., 1998; Rensink & Buell, 2004). Over the course of the past decade research has shown that *A. thaliana* is stress intolerant, with its inability to tolerate any form of extreme stress. Stress can be due to environmental factors, biotic, or abiotic stress such as temperature, drought or even salt stress; in plants the main cause of damage is oxidative stress when a combination of environmental factors lead to the creation of reactive oxygen species (ROS) (Sunkar, Ã, & Kirch, 2003). With each stress condition a different transcriptome is produced; this transcriptome is governed by signals that are sent out as a result of the exposed stress factors. By studying the transcriptome produced under different stress conditions we are able to better characterize and create a more well-rounded profile of how plant systems react under different conditions. Previous research has been able to show that different levels of tolerance are observed when exposed to different levels of stress; this lends itself to supporting the cause of a case specific transcriptome being produced in response to various stressors (Sunkar et al., 2003).

1.2 Effect of stress on plants

Stress can be defined as any external force that can trigger a defensive response from an organism. In plants, stress is something that can affect plant reproduction and productivity by activating the plants defensive response; in the form of post-translational modification that eventually leads to changes in the transcriptome and triggering repair and defensive mechanisms that have been adapted over an extended period of time (Kranner et al., 2016). Stress can be abiotic otherwise referred to as environmental; including water avilability, temperature and salt stress. Biotic stress factors include pathogens or wild life that will ultimately lead to mechanical damage to the plant (Kranner et al., 2010).

Environmental stress can cause great damage to plants and can lead to the disruption of many physiological functions, leading to the increase in the production of reactive oxygen species (Kranner et al., 2010). ROS are key components of signaling networks where they regulate various developmental processes and are key in controlling processes such as abiotic stress response and systemic signaling (Mittler, 2002). The effects of many different types of stress on plants have been studied, some more extensively than others owing to the variations in the overall climate and its effect on crop plants. A study conducted by *Beck et al. 2004* on Scots pine (*Pinus sylvestris*) done to investigate the environmental signals triggered as a response to cold stress, shows that low temperatures causes dehydration of the cells and tissues when cellular water freezes leading to loss of function of the bio-membranes. This can be detrimental to many plants that are not able to assimilate to the cold temperatures (Beck et al., 2004).

Abiotic stress is the primary cause of crop loss worldwide and can lead to a series of morphological, physiological, biochemical and molecular changes that negatively affect plant growth and productivity (Wang et al. , 2003). The diverse environmental stresses lead to the activation of cell signaling pathways and the production of stress proteins and the upregulation of anti-oxidants; the complex nature of plant response to abiotic stress is more accurately explained in (Wang et al., 2003). The final result of the plants response to stress is the development of stress tolerance or stress resistance (Boyko & Kovalchuk, 2008). Stress tolerance refers to the plants inherent ability to temporarily withstand stress,

possibly through the modification in gene function; making gene transformation to improve stress tolerance and increase crop potential an area for investigation. Stress resistance refers to the plants ability to completely counteract against a particular stress (Osakabe et al., 2014; Pugnaire & Luque, 2001; Wang et al., 2003). Biotic and abiotic stress factors pose a major threat to the production of agriculture worldwide, through the impairment of physiological functions of the plant. As such extensive research has been conducted to better understand the complex nature of the stress response system and potentially combat the effects of stress (Durian et al., 2016; HanumanthaRao et al., 2016).

Previous research has focused on the response of plants to single stress treatment under controlled conditions, this does not accurately reflect the natural conditions that plants are exposed to. In the field, plants are exposed to a combination of stressors as such those discussed in (*Rizhsky et al., 2004)* where a study was conducted on *Arabidopsis* plants with the purpose of investigating the changes expressed following exposure to a combination of heat and drought stress, transcriptome analysis of the *Arabidopsis* plant showed that there are 454 transcripts specifically expressed as well as a combination of two multi-gene defense pathways contributing enhanced respiration, suppressed photosynthesis and a complex expression pattern of defense and metabolic transcripts leading to the damaging effect on plant growth and productivity (Rizhsky et al., 2004).

Chen et al., 2013 investigated the transcriptome changes in response to salt, osmotic and cold stress on *Arabidopsis*, where plants were subjected to stress treatments of 4^{o}C, 100 mM NaCl, or 200 mM mannitol respectively. RNA samples from the leaves and roots were collected following specific time points, results were able to show that around 30% of the transcriptome is sensitive to regulation by stress conditions and majority of the changes are specific to each stimulus; indicating that each stress condition leads to a specific transcriptomic change and that there is a potential overlap between different conditions, potentially identifying shared stress responses (Chen et al., 2013; Kreps et al., 2002).

1.3 Alternative splicing

Alternative splicing is a gene regulatory process that results in a single gene having the capability to code for multiple proteins. Levels of alternative splicing can be affected by stress; stress can be in the form of biotic stress or even abiotic stress. Standard constitutive splicing is a process by which introns are removed from a protein and only exons remain in order to produce a mature mRNA sequence; alternative splicing is a variation on constitutive splicing where introns can be incorporated into the mRNA producing an mRNA sequence that codes for a potentially different gene with potentially a function entirely different than the original gene. Alternative splicing has widely been identified as an important posttranscriptional regulatory mechanism that can increase the proteome diversity and enhance transcriptome plasticity (Filichkin et al., 2010).

Several forms of alternative splicing have been identified including (Black, 2003; Matlin et al., 2005; Pan et al., 2008; Sammeth et al. , 2008) 1) alternative donor site; where an alternative 5' splice junction is used leading to a change in the 3' upstream boundary, 2) alternative acceptor site; similar to alternative donor site it is where an alternative 3'splice junction is used leading to a modification in the 5' downstream boundary, 3) intron retention; where an intronic sequence may be retained and depending on the nature of the retained sequence this can lead to the incorporation of a stop codon or a shift in the reading frame (Sammeth et al., 2008), 4) mutually exclusive exons; where one of two possible exons is retained in the mature mRNA but not both, 5) exon skipping; where an exon is spliced out of the primary transcript (Sammeth et al., 2008) (B. Wang & Brendel, 2006). A recent revelation has contributed a new form of alternative splicing termed Exitron; where introns possess features from both protein coding introns and protein coding exons (Marquez et al., 2015; Staiger & Simpson, 2015).

Alternative splicing is prevalent in both humans and plants with a difference in the most common form of alternative splicing expressed. Based on cDNA analysis of mammalian systems almost 60% of the human genes were suggested to be alternatively spliced with 58% of alternative splicing events belonging to the exon skipping category; intron retention is the least prevalent form of alternative splicing in humans with only 5%

(Wang & Brendel, 2006). In plants the opposite holds true with intron retention being the pre-dominate form; perhaps due to the naturally shorter intron length (Wang & Brendel, 2006). Variation in the splicing pattern contributes to major consequences for the mRNA isoforms; where the encoded proteins consist of unique domains that have different functions (Staiger, 2015). Variations in the sequence of an alternatively spliced isoform can lead to change in the regulation by microRNAs, stability and can be directed into RNA decay pathways (Staiger, 2015). More than 30% of alternative splicing events result in a premature termination codon which are targets for nonsense-mediated mRNA decay (Lewis et al., 2002). Through the use of nonsense-mediated decay alternative splicing leads to an overall quantitative change in the transcript levels.

1.4 Regulation of Alternative splicing

In *A. thaliana* alternative splicing is controlled by a spliceosome, a complex mechanism found primarily in the nucleus of eukaryotic cells that is composed of five small nuclear RNAs and a variety of protein factors. The spliceosome is responsible for removing introns from a transcribed pre-mRNA and allowing for the production of mature mRNA (Will & Luhrmann, 2011). Riboswitches; small metabolites that are responsible for controlling alternative splicing via mRNA secondary structure, have been shown to have an impact on gene expression following binding of the metabolite. miRNA has also been suggested to have a role in alternative splicing regulation, where the splicing of pre-mRNA affects the miRNA target site within and can lead to the generation of miRNA resistant or susceptible transcripts (Staiger, 2015). Another link between miRNA and the regulation of alternative splicing is that primary transcripts of miRNA undergo alternative splicing themselves, thus affecting mature miRNA (Staiger, 2015).

Owing to the difference in prevalent form of alternative splicing between plants and humans, it can be concluded that each organism recognizes exons and introns in different ways and evidence suggests that each may regulate alternative splicing in a way that is unique to the organism.

1.5 Alternative splicing in *Arabidopsis thaliana*

Splicing is a more conserved mechanism in plants than it is in humans and despite alternative splicing being widely studied in mammals, studies examining splicing in plants have been limited but there has been an increase in the number of investigations in the field. As a process alternative splicing can occur at different developmental stages and can be linked to a certain tissue type or environmental condition such as temperature, salt stress or abiotic stress (Filichkin et al., 2010).

Alternative splicing creates genetic variation and is able to restore coding potential to previously hidden genes (Gan et al., 2011). Alternative splicing events are highly common in *A. thaliana* with a rather high incidence of around 16% in all intron retaining genes (Wang & Brendel, 2006); the change in environmental condition is the leading cause of alternative splicing and is assisted by various splicing factors. Recent data based on whole genome transcriptome sequencing has shown that 61% of all *Arabidopsis* genes to be alternatively spliced (Filichkin et al., 2010; Marquez et al. , 2012; Reddy, 2007; Rühl et al., 2012; Staiger, 2015).

In *A. thaliana* it has been shown that serine-arginine rich proteins; a family of pre-mRNA splicing factors that show a high level of conservation, are largely responsible for regulating alternative splicing (Richardson et al., 2011).

1.6 Computational transcriptomics

Transcriptomics is the study of the complete set of transcripts present in a cell with respect to a specific developmental stage or under the influence of a certain physiological condition. By understanding the transcriptome we are able to interpret various functional elements of the genome, its molecular constituents and understand the development of diseases. Transcriptomics aims to determine the transcriptional structure of genes, their splicing patterns and quantifying the changing expression levels of each transcript during different developmental stages and conditions (Wang et al., 2009). Computational transcriptomics aims to achieve the goals of transcriptomics through computational means;

by using various programs that can be used to simulate an experiment and produce results that can provide insights without the expense of an experimental procedure. Computational transcriptomics can also be used to analyze results obtained from an experiment that was performed in the lab, allowing for more in-depth analysis and investigations.

1.7 RNA-sequencing

RNA-sequencing (RNA-seq) also referred to as whole transcriptome shotgun sequencing, is a fairly new technology that allows the capturing of both the quantity and quality of RNA in a genome at any specific point in time using high throughput next generation sequencing technology. The commercially available RNA-seq platforms are Illumina, Roche 454, Helico BioSciences and Life Technologies (Ozsolak & Milos, 2011; Z. Wang et al., 2009). Owing to the dynamic nature of the transcriptome meaning that it is constantly changing, being able to look at the bases present at any time allows for the possible identification of alternative gene spliced transcripts, post-transcriptional modifications, gene fusion and changes in the gene expression. Some of the advantages of RNA-seq is its ability to provide insight on not only mRNA transcripts but also visualize different populations of RNA including total RNA and small RNA (miRNA & tRNA) with higher accuracy and an increased coverage of the transcriptome population, its ability to determine exon/intron boundaries by verifying previously annotated 5' or 3' gene boundaries or by amending them (Ingolia et al., 2012; Morin et al., 2008).

RNA-seq is an improvement on the previous method used to study transcriptome and gene expression, which was microarray. Microarray relies on the prior knowledge of the organism's genome while RNA-seq does not. This allows RNA-seq to identify novel exons. RNA-seq is a form of quantitative transcriptomics analysis that is performed using several tools each with a specific function; it has become the standard in studying gene and transcription expression. From an experimental point of view the basic protocol for RNA sequencing involves several steps starting with obtaining RNA from the organism under investigation after which a double stranded cDNA library is generated from the mRNA using oligo primers. The generated cDNA is then fragmented by DNase I, adapters can then be ligated onto one or both ends of the fragment. Each fragment can then be amplified

and placed in a high throughput sequencer in order to obtain short sequence reads from either one end; termed single-end sequencing, or both ends; termed pair-end sequencing (Nagalakshmi et al., 2010). These reads which range from 30-400 bp; depending on the sequencing platform that was used, go through a genome assemble step where the reads are analyzed using one of two methods, if the organism under investigation has a known reference genome or transcript; the reads are mapped against the aforementioned reference genome. If a reference genome is not available then the sequenced reads can be assembled de novo (Alamancos et al., 2012; Katz et al., 2010; Lu et al., 2013; O'Neil & Emrich, 2013; Z. Wang et al., 2009). [Figure 1]

This step produces a transcription map consisting of both the transcriptional structure and/or the level of expression of each gene (Z. Wang et al., 2009). Following genome assembly the produced transcription map can be analyzed for differential gene expression through various bioinformatics tools. Therefore RNA-seq is used to better analyze the different transcriptomes produced under various stress conditions. By utilizing its ability to capture both the quantity and the quality of RNA in a produced transcriptome within a specific timeframe and the ability to capture several different types of RNA all at once, we are able to investigate different aspects pertaining to the dynamic nature of the transcriptome and potentially garner new information on how the plant responds to different conditions. Studies have shown that RNA-seq has been able to resolve the start and ends of known genes and transcripts while providing a better understanding of the splicing isoforms of known genes and discovering novel transcribed regions (Haas & Zody, 2010; Z. Wang et al., 2009).

Studies have shown that the accuracy levels of RNA-seq rivals that of other established methods such as microarrays and quantitative polymerase chain reaction (Griffith et al., 2010; Griffith et al., 2015; S. Li et al., 2014). It has been reported that 85% of novel splicing events and 88% of the differentially expressed exons that are predicted by RNA-seq are validated by approaches such as reverse transcription polymerase chain reaction and quantitative polymerase chain reaction (Griffith et al., 2010, 2015). RNA-seq has contributed to discoveries in many fields ranging from fusion discoveries in cancer (Griffith et al., 2015; Honeyman et al., 2014; Maher et al., 2009; Singh et al., 2012), regulation of alternative splicing (de Klerk et al., 2015; Griffith et al., 2015; Sultan et al.,

2008; Wilhelm et al., 2008) and it is currently being transitioned to clinical applications in many human diseases (Griffith et al., 2015; Kalari et al., 2014; Van Keuren-Jensen et al., 2014).

1.7.1 RNA-seq data Analysis

In order to understand the raw data obtained from RNA-seq, the data must undergo several steps before a clear picture appears. A description of the standard pipeline is as follows, the raw sequenced reads obtained from the sequencing machine undergo a pre-processing step entailing a quality control (QC) step. This QC step involves trimming of the reads, two common trimming strategies include "quality trimming"; which involves removing the ends of the reads where the base quality scores have dropped below a certain level. The second trimming strategy is " adapter trimming" where the adapter sequence is removed by masking specific sequences during library construction (Griffith et al., 2015). Read trimming is followed by an indexing step and then a read alignment or assembly step, this step involves merging reads into larger contiguous sequences (Contigs) based on the sequence similarity; the most commonly used tools are TopHat and STAR (Dobin et al., 2013; Engström et al., 2013; Griffith et al., 2015; Martin & Wang, 2011; Trapnell et al., 2009b).

Alignment of the RNA-seq data involves a comparison between each read to a previously assembled reference genome in order to produce transcripts, the choice of tool in this step is dependent on whether a reference genome is available. If a reference genome is available then Cufflinks and it's 3 subprograms cuffmerge, cuffquant and cuffdiff can be used (Griffith et al., 2015; Trapnell et al., 2010). If *De novo* assembly is being performed; meaning a reference genome is not available, then Trinity can be used (Grabherr et al., 2011; Griffith et al., 2015; Haas et al., 2013; Lu et al., 2013; O'Neil & Emrich, 2013).

Following transcript assembly gene expression analysis can be performed this can be done via a variety of tools the most common of which are CuffDiff (a subprogram of Cufflinks) and EdgeR (Griffith et al., 2015; Robinson et al., 2010; Trapnell et al., 2012, 2013). The final step in the analysis pipeline is visualization of the results, this can be performed by a variety of tools ranging from an intuitive interface such as a genome

browser in the form of integrated genome viewer (IGV) (Thorvaldsdóttir et al., 2013) or integrated genome browser (IGB) (Helt et al., 2009). Another visualization platform is through the use of CummeRbund (*CummeRbund - An R package for persistent storage, analysis, and visualization of RNA-Seq from cufflinks output*, n.d.) An R package developed for the visualization and analysis of RNA-seq data. For a complete view of the RNA-seq analysis pipeline see Figure 2.

Studies aimed at analyzing RNA-seq data have identified a standard pipeline that can be used for analysis, this pipeline consists of various tools each with their own function with the final aim of result visualization. This classic pipeline has been termed "Tuxedo protocol" it consists of a sequence aligner TopHat (Trapnell et al., 2009b) which uses Bowtie (Langmead et al., 2009) as a sub program for alignment, Cufflinks (Trapnell et al., 2012) for transcript assembly and expression analysis and finally CummeRbund (*CummeRbund - An R package for persistent storage, analysis, and visualization of RNA-Seq from cufflinks output*, n.d.) for visualization. [Figure 3]

1.8 Challenges with RNA-seq

There are a few challenging steps that one must keep in mind when considering RNA-seq the first of which is during library construction; owing to the highly manipulative steps involved during the extraction of the RNA involved in the production of cDNA libraries which can complicate its use in profiling all types of transcript (Wang et al., 2009); the issue of sample purity, quality and quantity are a major concern given that RNA is unstable and prone to degradation, thus requiring very specialized handling and techniques (Griffith et al., 2015). Another challenge faced is the ability to discriminate between whether a specific RNA species is abundant or is simply PCR artifacts; this can be overcome by determining whether the sequence is observed in different replicates.

From a bioinformatics angle there is a challenge in developing methods to store, retrieve and process the large amount of data produced from the experiments; the development of more efficient tools can decrease errors in analysis and remove low-quality reads (Wang et al., 2009). For large transcriptomes a major challenge is the alignment of a certain portion of reads to multiple locations in the genome (Mortazavi et al., 2008);

making the read alignment stage the most challenging (Trapnell et al., 2009a). Large genomes have more complex transcriptomes and require more sequencing depth which is expensive. Generally the more complex the genome the more sequencing depth is required in order to obtain enough coverage and produce significant results.

1.9 Microarray

Microarray is an array that is able to assay large amounts of biological material using highly miniaturized detection methods, creating the lab-on-a-chip technique. It is a collection of microscopic DNA sequences that represent all the genes in an organism arranged in a grid like fashion, probes are designed to bind to specific sequences. This probe is fluorescently labelled in order to simplify detection. Following the hybridization step the chip is placed through a laser which can analyze the different level of gene expression based on the intensity of color produced (Schena et al., 1995). Standard microarray depends on hybridization of known regions and this does not allow for the discovery of spliced transcript variants or any novel variants.

1. 10 Comparison between transcriptome analysis methods.

The developed technologies currently used to quantify the transcriptome can be categorized into hybridization based techniques or a sequence based approach. Hybridization based approaches typically involve incubating fluorescently labelled cDNA alongside gene specific microarrays; this method is relatively inexpensive and can be designed to span specific portions of the genome in order to be able to detect and quantify spliced isoforms. Hybridization based methods are limited greatly by their reliance on existing genome knowledge, a limitation in the range of detection due to potential cross-hybridization and finally comparing expression levels across different experiments is difficult and requires complicated normalization methods (Wang et al., 2009). In comparison sequence based approaches are able to directly determine the cDNA sequence, examples of this method include Sanger sequencing, Tag-based methods and RNA-seq. Unlike hybridization based methods RNA-seq is not limited to previous genome

knowledge, has very low background signal, is able to study complex transcriptomes owing to its ability to reveal the precise location of transcription boundaries and it does not have an upper limit for quantification making it much more sensitive than microarray (Wang et al., 2009) [See Table 1]. However microarray remains a "go-to" technique for researchers owing to its proven consistency and accuracy in measuring gene expression this is shown by the numerous studies conducted using the microarray technique such as the genome wide survey of cold stress regulated alternative splicing in *Arabidopsis thaliana* (Barah et al., 2013) and microarray based analysis of stress regulated microRNAs in *A.thaliana* (Liu et al., 2008).

1.11 Macrolides

Macrolides are a class of natural products that contain a large macrocyclic lactone ring, commonly found in *Streptomyces.* Macrolides have been shown to have antibiotic or antifungal activity. Macrolides exhibit their function by binding to the 50S subunit of the bacterial ribosome and preventing ribosomal translocation, were protein biosynthesis is subsequently inhibited (Tenson et al., 2003). Some macrolides have been shown to have tissue penetrative ability against Gram-positive bacteria as such they have been used in eukaryotes to modulate inflammation and immunity (Tenson et al., 2003).

Studies investigating the effects of macrolides on humans have shown that macrolides are able to modulate the level of mitogen-activated protein kinase (MAPK) pathway eventually effecting mucin gene expression (Kanoh & Rubin, 2010), they also have an anti-inflammatory effect by decreasing the production of proinflammatory cytokines (Čulić et al., 2001; López-Boado & Rubin, 2008; Shinkai et al., 2008; Zalewska-Kaszubska & Górska, 2001) for example. Studies investigating the action of macrolides on plants are very few, one study investigated the effect of antibiotic macrolides in wastewater treatment plants (Cs et al., 2003), another investigate the effect of six antibiotics on plant growth (Liu et al., 2009).

1.12 Pladienolide B

A study conducted in 2004 by Saki et al. reported the identification of seven 12-membered macrolides from *Streptomyces platensis* Mer-11107; these were termed pladienolides A-G (Kanada et al., 2007; Substances & Antitumor, 2004). Pladienolides were discovered as a result of a hypoxia induced gene expression, the most potent of which are pladienolide B (Antibiotics, 2004; Kanada et al., 2007) and D (Arai et al., 2014; Kanada et al., 2007; Substances & Antitumor, 2004). These two compounds have been able to inhibit the growth of a variety of cancer cell lines *in vitro* and show tumor regression activity *in vitro* (Kanada et al., 2007) making it able to affect cell proliferation and splicing. The structure of pladienolide B has been confirmed by Kanada et al. (Arai et al., 2014; Okuda, & Kawamura, 2004) where a study was conducted to synthesize compounds pladienolide B & D; with the aim of preparation of novel synthetic analogues of the afore mentioned compounds. See Figure 4, for the chemical structure of pladienolide B.

The prominent anti-tumor activities of pladienolide B & D in both *in vitro* and *in vivo* systems, is a result of their binding to splicing factor SF3b (Kotake et al., 2007). Kotake et al., have shown that when pladienolide B binds to splicing factor SF3b leading to the inhibition of the spliceosome. This may impair the cellular mechanism of gene expression through any one of the several steps involved in the process including transcription, pre-mRNA processing, mRNA surveillance and mRNA export as well as any of their sub mechanisms (Kotake et al., 2007; Maniatis & Reed, 2002; Proudfoot et al., 2002). This effect lends evidence to the potential use of splicing machinery as an antitumor drug target (Kotake et al., 2007).

Another study has been able to validate the biological target of pladienolide B to be SF3b through the use of pladienolide-resistant clones from two colorectal cancer cell lines. Through the use of differential gene analysis the cell lines have been shown to possess a mutation in a gene coding for SF3b1; a subunit of SF3b, rendering both cell lines pladienolide resistant and as such not affected by the anti-proliferative effect of

pladienolide or its inhibitory effect on the spliceosome (Yokoi et al., 2011). This lends further proof to the use of pladienolide B as an anti-tumor drug in mammalian cells.

Pladienolide B was chosen as the drug for this study owing to its previously discussed proven effect on the splicing machinery in mammalian cells. In an investigatory fashion it was a point of interest to see what effect the compound may have on plants, if any.

1.13 Study objective & design.

As previously discussed, stress plays a major role in contributing to alternative splicing in plants. Understanding the intricate mechanisms of alternative splicing in *A. thaliana* and analyzing the subsequent transcriptomic changes will contribute greatly to the overall understanding of how plants react to stress.

The aim of this project is to assess the transcriptomic changes occurring in *A. thaliana* in response to pladienolide B, a molecule that has been shown to block the spliceosome and mimic the effects of stress. Allowing for the identification of the underpinnings between alternative splicing regulations in plants in response to stress cues. This will be done through the following steps:

1) Reporting the statistically and biologically differentially expressed genes.
2) Reporting the functional classification of the differentially expressed genes.
3) Inferring the differentially expresses alternatively spliced genes due to the treatment
4) Inferring the most enriched functional categories to which the differentially expressed alternatively spliced genes belong to.

Chapter 2: Methods

2.1 Experimental Setup

The experiments were carried out at King Abdulla University of Science and Technology (KAUST) in Dr. Magdy Mahfouz's lab. Seeds of wild-type *Arabidopsis thaliana* Col-0 were surface sterilized with 10% bleach for 10 minutes and used directly for seed germination assays or stored at 4ºC for 2 days. The seeds were plated on Murashige and Skoog (MS) medium agar plates supplemented with 1% sucrose and pladienolide B, the plates were then placed in a growth chamber under white light for 16 hours (~75 μmol m^{-2} s^{-1}) and 8 hour dark conditions at 22ºC for germination and seed growth.

One week old *Arabidopsis* seedlings were treated with 5.0 μM Pladienolide B for 6 and 24 hours, where total RNA was extracted using TRIzol Reagent, for a total of 2 conditions were used; a control condition (with 3 replicates for each of the 6 and 24 hour time point; denoted as C06 & C24) and a treatment condition (with 3 replicates for each of the 6 and 24 hour time points; denoted as P06 & P24). RNA-seq libraries were constructed using an Illumina Whole Transcriptome Analysis Kit following the standard protocol (Illumina HiSeq system) and sequenced on the HiSeq platform to generate high-quality pair-end reads.

2.2 Computational Analysis

All the computational analyses requiring a server were performed on the AUC server using the default parameter of each program. The raw RNA-reads for the 2 conditions (Control vs. Treatment) at 2 different time points (6 hr. & 24 hr.) each with 3 replicates; for a total of 6 files per condition and an overall of 12 files, were aligned using TopHat (v2.0.13) which used bowtie (v2.2.3) as an assembler (Trapnell et al., 2012). The

reads were aligned against the TAIR10 version of the *A. thaliana* genome. Transcript assembly was performed using Cufflinks (v2.2.1), this was conducted on each of the replicates for each of the time points. Following the assembly of transcripts for all conditions and replicates, differential gene expression was analyzed using CuffDiff (a subprogram of Cufflinks).

Differential expression analysis was performed using the R/Bioconductor package CummeRbund (v3.3); used for the visualization and exploration of cufflinks high-throughput sequencing data (Trapnell et al., 2012). The R version used is the Bug in your hair (3.3.1). The annotated IDs for the differentially expressed genes for each condition were extracted and uploaded on the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7). Functional analysis was implemented using the Gene Ontology Panther classification system (v11.1) (Mi et al., 2013; Mi et al., 2016; Mi & Thomas, 2009; Tang & Thomas, 2016) , the default parameters were used. For an overview of the workflow used see figure 5.

Chapter 3: Results and Discussion

Genome guided assembly via Cufflinks was used to assemble the transcripts from each of the 12 files (3 replicates per time point per condition). The number of transcripts assembled from the control condition at the 6 hour time point varied from 34092876 transcripts for the C06_0 replicate, 27470481 transcripts for CO6_1 replicate and 28909860 transcripts for CO6_2 replicate. Meanwhile the number of transcripts produced from the treatment condition at the 6 hour time point were as follows. 24410550 transcripts from the P06_0 replicate, 30688444 for the P06_1 replicate and 29625260 from the P06_2 replicate.

The transcripts produced from the 24 hour time point for the control condition ranged from 27474891 at the C24_0 replicate, 30154454 at the C24_1 replicate to 22128277 at the C24_2 replicate. While at the 24 treatment condition 28370379 transcripts were produced at the P24_0 replicate, 29331131 at the P24_1 replicate and 33418085 at the P24_2 replicate [Table 5]. A heat map was generated in order to visualize the correlation of gene expression between samples and replicates [Figure 21]. In it, it becomes clear that all control sample are grouped together and all treatment samples are grouped together. Thus verifying the relationship between the samples and each other.

Differential gene expression analysis of both conditions using CummeRbund revealed the differential expression of 32996 genes, 63972 isoforms and 237492 splicing events. Upon further investigation of the differentially expressed genes for each time point, the differentially expressed genes were identified and the annotation for the aforementioned genes were compiled and placed through an online tool for functional analysis. This allowed for the identification of the functional categories that were most affected by the plants exposure to the drug. Using different functions of CummeRbund a range of plots can be generated in order to visually analyze the differentially expressed genes from cuffdiff.

In the following section the different generated plots will be presented and analyzed.

3.1 The 100 most differentially expressed genes

Differential expression analysis identified a total of 12646 significant differentially expressed genes for the 6 hour time point; comparing the 6 hour control replicates with the 6 hour treatment replicates. For the 24 hour time point a total of 10649 genes were significantly differentially expressed; when comparing the control and the treatment conditions. All the significant differentially expressed genes were utilized to generate the plots seen below [Figure 6- Figure 14]. The functional analysis for the 100 most differentially expressed genes will be discussed in details in the following section.

3.2 Density Plot

Density plots are generated to assess the distributions of FPKM scores across the various samples; the frequency of each FPKM score in the sample is plotted in a density plot (Fukunaga, 1990; Racine et al., 2004). Figure 6 shows the distributions of the FPKM scores across the various samples in both conditions and figure 7 shows the FPKM distributions across all replicates of both conditions. From these figures we can deduce that genes are split into two main peaks one at a log10 (FPKM) of -2 and another at log10 (FPKM) of 2. At log10 (FPKM) 2, genes from the C06 sample show the highest distribution while at -2; genes from the P24 sample show the highest distribution scores.

3.3 Scatter Plot

Scatter plots are produced in order to perform a pairwise comparison between conditions where the data is displayed as a collection of points representing the normalized values (log10FPKM) between two conditions (Friendly & Denis, 2005; Fukunaga, 1990;

Jarrell, 1994). A scatter plot for both time points across both conditions is generated [Figure 8], each scatter plot is a comparison between two conditions.
The comparison between the different time points of the control condition show a similar expression trend, while the comparison between control and treatment conditions show a wide range of gene expression.

3.4 Box Plot

Boxplots are generated to study the distributional characteristics of a group of values and in turn show an overall response pattern of that group (Cleveland, 1993; Fukunaga, 1990; Jarrell, 1994). A total of two boxplots were generated in the analysis, [figure 9] (a comparison between the two conditions across the two time points); where the slightly larger size of the control shows a wide range of differential expression and all samples show a similar pattern of expression as per the positioning of the median line and the middle quartile (Frigge et al., 1989; McGill et al., 1978; Tukey, 1977).

Figure 10 (a comparison of all the replicates for both conditions across the two time points), shows that the widest range of expression values belongs to the zero replicate in both the 6 hour and 24 hour time point of the treatment condition. This points to the zero replicate having the widest range of differential expression throughout all the samples.

3.5 Volcano Plot

Volcano plots are types of scatter plots used to identify changes in large datasets that are composed of replicates. It is generated by plotting the negative log of the p-value representing the significance on the y-axis against the log of the fold change on the x-axis (Cui & Churchill, 2003; W. Li, 2012; Tukey, 1977). Each point is the fold-change versus the significance for each gene between the conditions specified in the analysis. The results of a volcano plot are data points of high statistical significance (low p-value) and high biological significance (high fold-change) appearing towards the top of the plot.

Figure 11 shows the volcano plot of –log10 (p-value) vs. log fold- change, showing both the statistical and biologically significant genes. The comparison between the control and treatment conditions show a wider range of differentially expressed genes.

3.6 Principal Component Analysis (PCA)

Principal component analysis is a statistical procedure that converts a set of potentially correlated variables into a smaller group of linearly uncorrelated variables called principle components (Jolliffe, 2002). This allows for the identification of the most important variables that contribute to the variability in the data through dimension reduction by correlating between the different conditions (Fukunaga, 1990; *Network component analysis: Reconstruction of regulatory signals in biological systems*, n.d.). For analysis purposes several PCA plots were generated with various combinations in an attempt to identify the variations between the datasets.

The PCA plots were able to group the samples belonging to the control condition together and the samples belonging to the treatment condition together. Verifying the variability in the data can be attributed to the introduction of a treatment condition and that all results are largely due to the aforementioned administrated treatment.

3.7 Functional Analysis using Gene Ontology PANTHER Classification system

In order to analyze the functions of the significant differentially expressed genes the generated annotations were run through PANTHER (Protein Analysis Through Evolutionary Relationships) (Mi et al., 2016) an online functional classification tool.

3.7.1 Functional analysis of significant differentially expressed genes

The significant differentially expressed genes were placed in PANTHER in order to deduce the function of each gene. A total of 12646 genes were placed in PANTHER for the 6 hour time point comparison, generating figure 15, where the genes were grouped into functional classes. The 6 hour time point (C06 vs P06) comparison showed that 48.9% of

the genes have known catalytic activity and 26.4% have known binding activity. Figure 16 shows the functional classes of the 10649 significant differentially expressed genes at the 24 hour time point (C24 vs P24) showed that 50.3% of the genes have known catalytic activity and 26.5% have known binding activity.

The significant differentially expressed alternatively spliced genes were similarly analyzed using PANTHER. For the 6 hour time point comparison (C06 vs P06), a total of 3763 genes were found to be significantly differentially expressed. 46.6% of these genes were found to have catalytic function and 33.9% of the genes have binding activity [Figure 17]. For the 24 hour time point comparison (C24 vs P24) a total of 2797 alternatively spliced genes were found to be significantly differentially expressed with 46.4% having catalytic activity and 33.7% having binding activity [Figure 18].

Throughout the following section gene expression will be discussed, given the nature of the treatment of choice; Pladienolide B, which is an inhibitor of alternative splicing. Genes that are upregulated as a result of exposure to the inhibitory nature of the treatment, would under normal conditions be inhibited by alternative splicing and as such the treatment is able to prevent the inhibitory action of alternative splicing and subsequently allow the expression of the genes which are potentially not involved in stress response. While the opposite holds true for genes that are down-regulated, under normal conditions these genes are either not affected by alternative splicing or are positively affected. The introduction of pladienolide B, an inhibitor of alternative splicing causes the genes to be negatively affected and for their expression levels to drop. This lends itself to the thinking that genes that are downregulated as a result of a lack of the action of alternative splicing are potentially essential in plant stress response and may play a role in the regulation of stress in plants.

3.7.2 Functional Analysis of the top 100 alternatively spliced differentially expressed genes.

Given that the administrated treatment (Pladienolide B) has been shown to affect alternative splicing by inhibiting the splicing machinery, the following section will focus on investigating the significant differentially expressed genes that were alternatively

spliced. Owing to the large number of alternatively spliced genes that were differentially expressed; 3763 genes for the 6 hour time point comparison & 2797 genes for the 24 hour time point comparison, only the top 100 genes expressed will be discussed. This is true for both time points of the comparison.

3.7.2.1 6 hour time point comparison

This section will address the genes that were differentially expressed in the 6 hour time point comparison (between control and treatment, C06vsP06) and not expressed in the 24 hour time point comparison. A total of 48 genes were uniquely expressed in the 6 hour treatment time point when compared to the control. Of which 44 have been previously identified and functionally characterized [Table 2] and the remaining 4 genes have unreported functional classification as of this date.

MAC3A is a pre mRNA processing factor 19 gene involved in mRNA splicing, expression levels were suppressed from 15.149 FPKM to 8.32993 as a result of the treatment. It has been suggested that MAC3A is a member of the MOS4-Associated Complex (MAC) which functions redundantly in the regulation of the plant immune system and response to plant pathogens a fact that has been corroborated by *Kourmpetis et al* (Kourmpetis et al., 2011; Monaghan et al., 2009). It is also homologous with the yeast and human E3 ubiquitin ligase PRP19 (Monaghan et al., 2009) which is similar in nature to PRP 39 which shows a high level of expression across the time points and will be discussed in detail later on.

ADF11 is a hypothesized actin-depolymerizing factor protein belonging to the ADF/cofilins family of actin-binding proteins that through multiple in-vitro experiments and in-vivo experiments has been shown to play a role during the development of organisms; these f-actin proteins have been shown to be key regulators in flowering and cell and organ expansion in *Arabidopsis thaliana* (Dong et al., 2001; Maciver et al., 2002). Genes belong to the ADF/cofilins family have been known to be expressed in multiple in various organisms (Maciver et al., 2002). BT3 is a non-motor actin binding protein belonging to the BTB AND TAZ DOMAIN proteins which has been hypothesized to

regulate telomerase activity in mature organs upon exposure to auxins. BT3 is also expressed when the predominate gene responsible for the gametophyte development is lost (Berr et al., 2010; Mahé et al., 2007; Ren et al., 2007; Robert et al., 2009).

CKL13 has been classified as a non-receptor serine/threonine protein kinase which potentially could cause the downstream lack of functional modification and as such could adversely affect the enzymatic activity and location of other proteins. (Delhaize & Randall, 1995; Menges et al., 2005). RPL2, RPS19, RPS11, RPS7, RPL16, RPS11, RPS15, RPL2, RPL14, RPL20, RPL32, RPS19, ORF110A and RPL16 are all ribosomal proteins which have been shown to possess different isoforms as well as potentially being post translationally modified. These proteins could possibly be the result of the altered expression of the previously discussed genes, which are all involved in the developmental process of *Arabidopsis thaliana*. AT1G05730, AT1G04790, AT1G01810, AT1G01730 are all expressed transcripts that are functionally uncharacterized to date and pose a new avenue for investigation on the effects of alternative splicing.

YCF10 a gene belonging to a family of algal YCF 27 proteins and was upregulated following the administration of Pladienolide B, along with several others of the YCF family including (YCF 1, 2,3,4,5 and 6). These genes exhibited an increase in the expression level at the 6 hour time point, this family has been shown to have a role in chloroplast evolution and encode for transcriptional regulators (Ashby et al., 2002). All 7 of the expressed genes hold a role in chloroplast evolution ranging from chloroplast envelope membrane protein to photosystem assembly protein and center protein (Sato et al., 1999). YCF3 is a chaperone involved in PSI (Photosystem I) assembly and its splicing has been known to lead to defective PSI assembly and degradation (Landau et al., 2009), given the nature of the treatment and that it effects splicing it points to a potential negative impact on the growth and development of the plant. PSBJ is another photosystem II reaction center protein that possess similar function as some of the YCF family genes that through investigation has been linked to developmental control of the PSII core and OEC (Oxygen Evolving Complex) proteins, having a vital role in correct assembly (Suorsa et al., 2003). PSBJ levels show an almost two fold increase in expression following the

treatment which points towards the over activity of the genes necessary for survival as a means to adapt to the effects of the treatment.

GATL5 is a galacturonosyltransferase like 5 protein that is involved in the production of *Arabidopsis thaliana* seed coat mucilage which due to the specialized epidermis, ruptures upon exposure to water and is able to encapsulate the seed in order to aid in the protection, hydration and distribution of said seed (Kong et al., 2013; Ralet et al., 2016; Saez-Aguayo et al., 2013). Its slight increase in expression levels indicate that the seed potentially requires more protection or hydration in order to reach maturity.

ATAF1 is an NAC transcription factor that has a major role in negatively regulating the plant response to stress and development by directly regulating the abscisic acid biosynthesis genes in *Arabidopsis thaliana.* Abscisic acid possess a regulatory role in abiotic stress responses by directing the plant towards desiccation tolerance and enabling the adaptation to water stress (de Torres-Zabala et al., 2007; Fan et al., 2009; Jensen et al., 2013; Wang et al., 2009). This is the second differentially expressed gene to possess action towards plant-pathogen interactions.

It is clear that the uniquely expressed alternatively spliced genes in the 6 hour time point after the administration for the treatment; Pladienolide B, have correlating functions all linked to the modulation of the development of *Arabidopsis thaliana* ranging from ribosomal proteins, protein kinases that will affect the functional modification of downstream binding proteins to actin binding proteins that alter the telomerase activity and regulate the maturation of developmental organs in plants. The various ribosomal proteins possess the highest upregulation with a fold change ranging from 1.5 increase for RPL32 and RPL32, a 0.9 fold change for RPL2, and a 0.6 fold change for RPL16, RPL14, RPS11. RPS15 has the highest upregulation with a fourfold increase in gene expression between control and treatment. ADF11 is a hypothesized actin-depolymerizing factor protein that shows the largest down regulation with a threefold decrease between the control and the treatment. Another down regulated gene is the CKL13 which is classified as a non-receptor serine/threonine protein kinase, it shows a slight decrease in expression with a -0.6 fold

change. MAC3A a pre mRNA processing factor 19 gene involved in mRNA splicing is similarly down regulated with a -0.8 fold change in gene expression.

As such the majority of the genes that are involved in the developmental process of plants are all significantly affected by the action of Pladienolide B, from this it can be deduced that inhibiting a specific splicing factor in the spliceosome will ultimately lead to the altered expression of genes directly involved in the maturation and development of plants. Despite the highly investigated nature of *Arabidopsis thaliana* a significant number of genes remain functionally uncharacterized to date and present a new area for further investigation. These uncharacterized genes (AT1G05730, AT1G04790, AT1G01810, and AT1G01730) are upregulated and show an increase of 1.5 times in gene expression and are a point of interest for future studies.

3.7.2.2 24 hour time point Comparison

This section will address the genes that were uniquely differentially expressed in the 24 hour treatment time point comparison (between control and treatment, C24vsP24) and not expressed in the 6 hour treatment time point. A total of 51 genes were uniquely expressed in the 24 hour time point comparison between the control and treatment. Of the 51 genes, 40 have been previously identified and functionally characterized [Table 3], and the remaining 11 genes have unreported functional classification as of this date.

A few of the expressed transcripts such as AT1G05140, AT1G07170, AT1G05960, AT1G06640, AT1G06650, AT1G03200, AT1G03210, AT1G06645, AT1G06870, AT1G06135, and AT1G07170 have been hypothetically functionally characterized but remain unnamed as they are denoted by their annotation, these transcripts possess a variety of functions ranging from PHD finger-like domain containing proteins, chloroplast related membrane metalloprotease ARASP to Oxidase homolog related genes to name a few.

ABCI19: ATP-binding cassette (ABC) transporter. It belongs to a superfamily of oligopeptide permease proteins responsible for transporting a wide range of substrates across membranes, ABC is considered one of the largest protein transporter families that

are driven by ATP hydrolysis. ABCI19 is one of the 22 functionally characterized *Arabidopsis* transporters that has been shown to translate to cytosolic proteins and is required for a multitude of processes including plant development, response to stress and pathogenic resistance (Marin et al., 2006). ABC transporters have an overall involvement in plant development and survival (Kang et al., 2011). HSP17 is a heat shock protein induced by heat and osmotic stress (Sun et al., 2001), it shows the highest level of upregulation with an almost fivefold increase in expression. The versatile response to stress has not been fully explored in *Arabidopsis* and the involvement of proteins that are joined with particular types of stress such as salt, drought, and osmotic stress is something that is still under investigation.

Similar to the 6 hour time point comparison of the expressed genes, a few are characterized as ribosomal proteins such as AT1G06380, and RPL4. This differs from the considerable number of ribosomal proteins that were differentially expressed at the 6 hour time point comparison, suggesting that with the extended exposure of the plant to the treatment some of the genes are no longer effected and the plant is able to adapt to the effects of the treatment. Several of the genes that are differentially expressed are characterized as transcription factors such as BHLH128; that could be a potential component of a regulatory network that controls root nitrate response which is essential for proper plant development (Vidal et al., 2013).

A considerable number of genes in enzymatic actions are differentially expressed following the 24 hour exposure to the treatment including but not limited to ATNDI1 a dehydrogenase that belongs to a family of enzymes with yet unexplored function, location and specificity (Moore et al., 2003). CER1 is a decarbonylase gene that possess similarity to integral membrane enzymes and has been shown to be involved in wax biosynthesis and pollen fertility (Aarts et al. 1995). ATPI, GTE4, ATNDI1, ORF315, AT1G01300, COX1, NAD3, ATPI, RCI3, UGT71CA, FRO2, UGT71C5, and PETD are all genes with enzymatic actions ranging from glycosyltransferase to ferric reduction oxidase. CHR9 (Chromatin remodeling 9) is a DNA helicase and NIH is identified as an ATP-Dependant RNA helicase that is downregulated following the 24 hour time point of the treatment.

AXR3 and IAAI2 are auxin responsive proteins that are a part of the AUX/IAA gene family that are primarily transcriptional repressors that upon exposure to auxins even at low concentrations are able to prevent auxin induced gene expression (Leyser et al., 1996; Nanao et al., 2014; Rouse et al., 1998). The interaction between the genes is vital for auxin signaling and could potentially lead to new information about auxin induced gene regulation.

There is noticeable similarity in the functions of the genes differentially expressed but with an expected decrease in the number of the genes affected, this suggests that over time the treatment exerts a greater effect and considerably alters the expression levels of a wider variety of genes or that the plant is able to adapt to the effects of the treatment. Genes with the highest level of upregulation in gene expression have varying functions ranging from HSP17; a heat shock protein that shows the highest increase in expression with an almost fivefold increase, to ABCI19 a member of the ABC transporter family that shows a 1.5 increase in gene expression between the control and treatment at the 24 hour time point. Unlike the 6 hour time point, there is a considerably larger number of downregulated genes at the 24 hour time point; almost double the amount, genes with the largest decrease in gene expression are those with various enzymatic functions such as ATNDI1, UGT71CA, and FRO2.

At the 24 hour time point genes show an increased reaction to the prolonged exposure to Pladienolide B, with a greater number of genes becoming down-regulated as a result. While the number of genes affected may have decreased overall, the change in gene expression levels is much higher when compared to the gene expression levels at the 6 hour time point. The decrease in the number of affected genes indicates that the plant adapts to the treatment, as the length of exposure increases.

3.7.2.3 Continuously Expressed Genes

This section will address the change in the expression levels of some genes; genes that were found to be differentially expressed; when comparing control and treatment gene, across both time point but with varying expression levels. A total of 54 genes were jointly expressed in both the 6 hour and 24 hour time point comparison with 50 showing previously reported functional classification [Table 4] and 4 genes having unreported functional classification as of this date.

PRP39 is a pre mRNA processing factor 39 gene involved in mRNA splicing, previously shown to affect flowering time in *Arabidopsis thaliana*. Following administration of the treatment *(*Pladienolide B*)* expression of the gene increased by 1.5 folds, this level of expression was maintained and increased from the 6 hour exposure to the 24 hour exposure; expression levels went from 46.3784 FPKM at the 6 hour time point to 55.0048 FPKM at the 24 hour time point. PRP39 has been shown to be conserved in yeast, humans and plants owing to the tetratricopeptide repeat proteins present within the structure (Bayne et al., 2008; Beggs, 1993; Karpov & Blume, 2008; C. Wang et al., 2007)**.** Similarly IAA10 and ATG04090 are two genes that were simultaneously expressed with PRP39, potentially due to close proximity within the transcript. IAA10 is classified as an auxin responsive protein, auxins have been known to regulate transcription for early response genes and are highly involved in cell type specification (Chandler & Werr, 2015; Hagen & Guilfoyle, 2002). AT1G04090 has no clear functional classification to date but has been marked in TAIR database as an unknown product, a study conducted to investigate the potential conservation of microtubule-associated proteins (MAPs); which play a role in increasing the rate of polymerization, assembly and stabilization of microtubules and microtubule functions, between animal & plant tubulins has been able to positively map a universal MAP1 motif to the area identified as AT1G04090 (Karpov & Blume, 2008). It can be deduced that functions that are involved in the early stages of

development and have a high level of importance are similarly over expressed as a result of the Pladienolide B treatment potentially to counteract its effect.

A variety of functions were continuously expressed following the exposure to the treatment where there are several genes involved in ATP synthase such as ATP1,6,9 where ATP 9 expression causes mitochondrial dysfunction in the form of the sterility of the male *Arabidopsis thaliana* (Gómez-Casati et al., 2002). Again there are several genes continuously expressed with enzymatic activity, these expressed genes have a wider variety of enzymatic activity than those uniquely expressed at the 6 hour time point and the 24 hour. These genes are AT1G01350, AT1G01630, AT1G01210, AT1G03220, AT1G03230, AT1G04430, AT1G05000, AT1G05700, FKGP, FZL, GAE2, MCCA, NAD4, NAD7, ORFX, PME7, COX2 COX 3 BXL2, and BGLU11. The functions vary from ligase, dehydrogenase, dehydratase, reductase, methyltransferase, and to aspartic protease.

A series of proteins expressed after the treatment exposure are all linked to the previously discussed photosystem I reaction and in particular the center subunit, these genes are PSAI, PSAJ, PSBI, PSBK, PSBL, and PSBT; as previously mentioned these genes are all involved in chloroplast evolution and are needed for mature plant development (Sato et al., 1999). These photosystem 1 reaction genes are upregulated with an almost 1.5 fold change in expression between the 6 hour and 24 hour time point condition. Similarly to ABCI19, PGP11 belongs to the ABC transporter B family. This family is responsible for a transporting substrates across membranes (Kang et al., 2011; Marin et al., 2006). It is also the gene with the highest fold change in expression between the 6 hour and 24 hour time point, PGP11 is one of the highest down-regulated genes with a fivefold decrease in expression.

Similarly out of the 50 continuously expressed genes 24 are downregulated (between 6 hour and 24 hour time point) and the remaining genes are up-regulated (between the 6 hour and 24 hour). The genes with the largest positive fold change i.e. up regulated are the genes with enzymatic function such as COX 3 & 2 with an almost 3 fold

increase in gene expression, while the genes with the largest negative fold change (down-regulated) are the Photosystem I reaction genes such as PSAI, PSAJ, PSBI, PSBK, PSBT, and PPSBT. They exhibit almost twofold decrease in gene expression across the 24 hour time point. AT1G05320 and AT1G04090; both uncharacterized genes, show a constant rate of expression with only a slight increase in gene expression (0.06 and 0.08 fold change respectively), ORF294 another uncharacterized gene shows a 15 fold decrease in expression when comparing the 6 hour and 26 hour time points.

Genes that are continuously expressed at both the 6 and 24 hour time points have similar functions to those uniquely expressed at the 6 or 24 hour time point. A larger number of genes that are continuously affected have enzymatic function of various actions and to a smaller extent genes involved in the photosystem reaction being the most effected with a fivefold decrease in expression between the 6 hour and 24 hour time point.

Chapter 4: Conclusion and Future Recommendations

Understanding the response of the model organism *Arabidopsis thaliana* to various stressors is an important step in deepening the understanding of the highly adaptive nature of plants and the way various treatments can affect the transcriptome. The RNA seq profile of the pladienolide B exposed *Arabidopsis* prepared by KAUST was analyzed in an investigatory fashion of identifying which genes were effected and the various biological processes affected by the response to the stress. The differentially expressed genes were functionally analyzed to view the affected functional processes.

Throughout the exposure of the plant to pladienolide B, the number of significant differentially expressed genes decreased from 12646 genes at the 6 hour time point to 10649 genes at the 24 hour time point. This is true for the alternatively spliced genes where at the 6 hour time point a total of 3763 genes were significantly differentially expressed and at the 24 hour time point only 2797 genes were found to be significantly expressed. Genes uniquely expressed at both the 6 & 24 hour time point show similar functions involved in developmental regulation, stress response proteins and response to pathogens, all of which are essential for the maturation and development of the plant.

Further investigation of the role of the expressed genes with uncharacterized functions or hypothetical functions is needed as it may lead to further understanding the way plants react to stress. As well as investigating the different forms of alternative splicing that are affected by the treatment. This research offers a deeper understanding of how plants are effected by stress and the interplay between stress and alternative splicing through the characterization of the differentially expressed genes.

TABLES AND FIGURES

Technology	**Microarray**	RNA-seq
Technology specifications		
Principle	Hybridization	High throughput sequencing
Resolution	Several-100bp	Single base
Throughout	High	High
Reliance on genomic sequence	Yes	In certain cases.
Background noise	High	Low
Application		
Simultaneous mapping of transcribed genes and gene expression	Yes	Yes
Ability to distinguish different isoforms	Limited	Yes
Practical issues		
Required amount of RNA	High	Low
Cost of mapping transcriptomes of large genomes	High	Low

Table 1: Comparison of the differences between microarray and RNA-seq. (Wang et al., 2009)

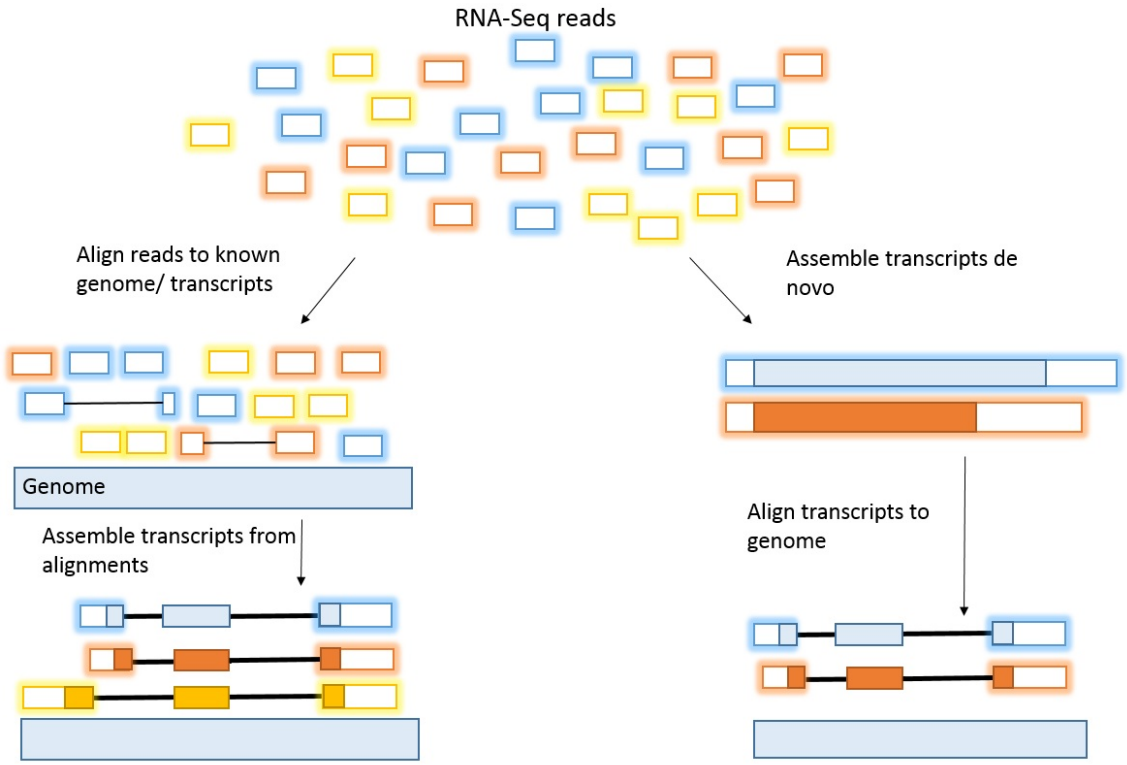

Figure 1: Approaches for transcript assembly from RNA-seq reads. The approach on the left, describes genome assembly using a reference genome (Trapnell et al., 2009a); this approach first aligns the RNA-seq reads to the reference genome followed by transcript reconstruction from the alignment (Haas & Zody, 2010). The approach on the right describes de novo transcript assembly, where the transcript sequence is assembled directly from the RNA-seq reads; they are then aligned to the genome. RNA-seq reads are colored according to the transcript isoform from which they are derived. Protein-coding regions of the constructed isoforms are depicted in darker colors (Haas & Zody, 2010).

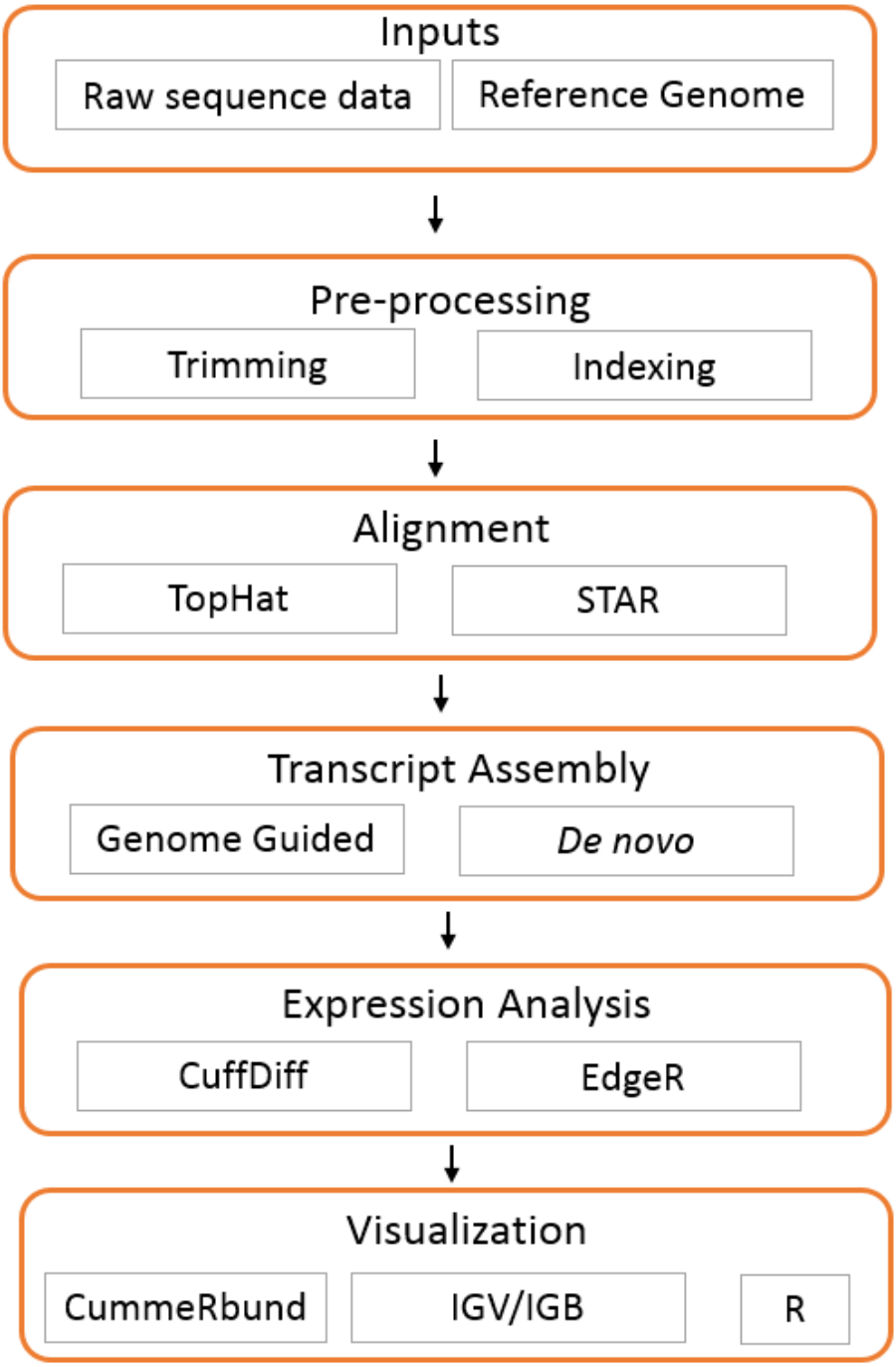

Figure 2: RNA-seq analysis flow chart.

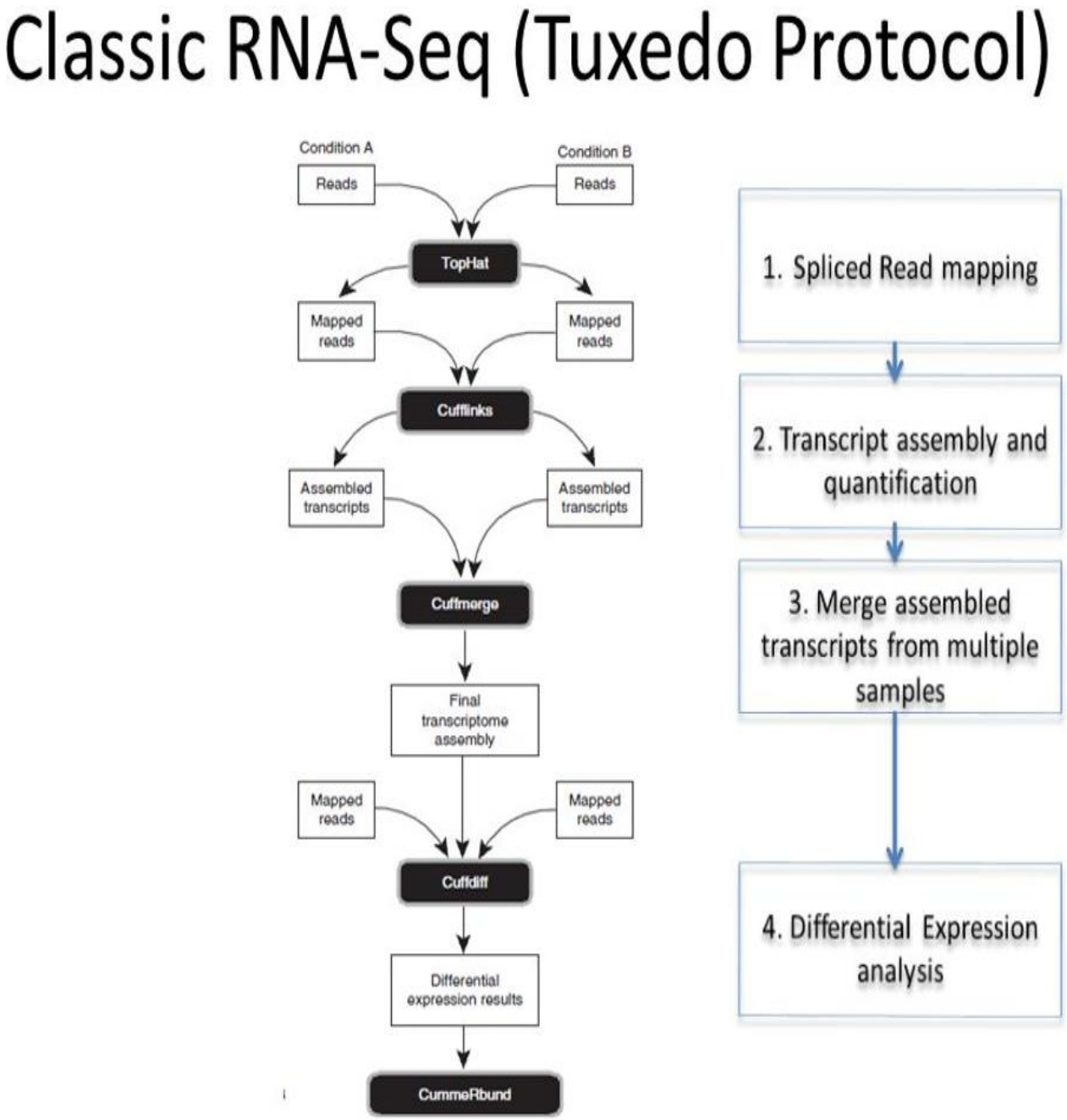

Figure 3: Flowchart of the classic RNA-seq analysis protocol "Tuxedo Protocol" adapted from (Trapnell et al., 2012)

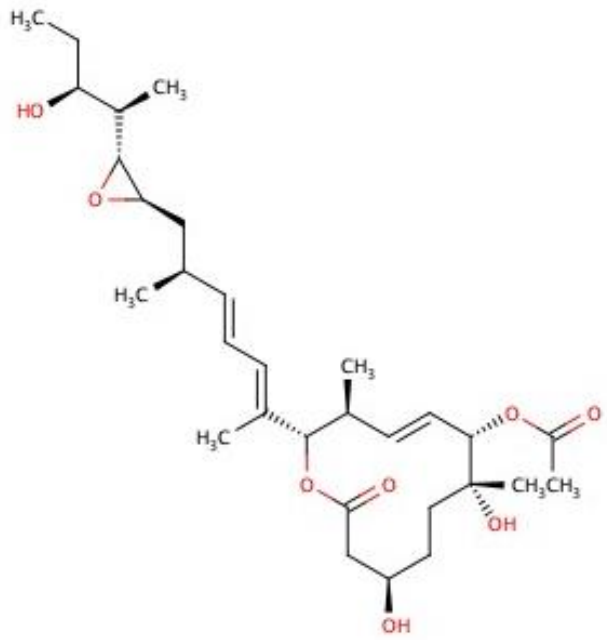

Figure 4: Chemical structure of pladienolide B. this compound is produced by *Streptomyces platensis* Mer-11107 and has been shown to have a molecular formula of $C_{30}H_{48}O_8$ and a molecular weight of 536 (Substances et al., 2004).

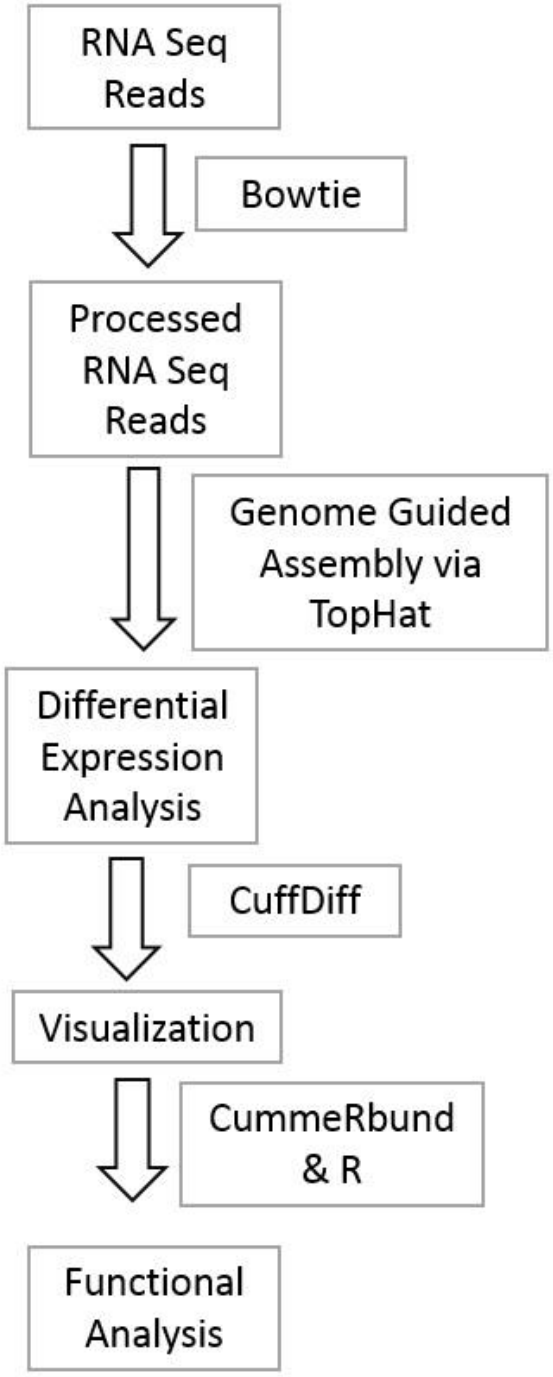

Figure 5: Workflow overview.

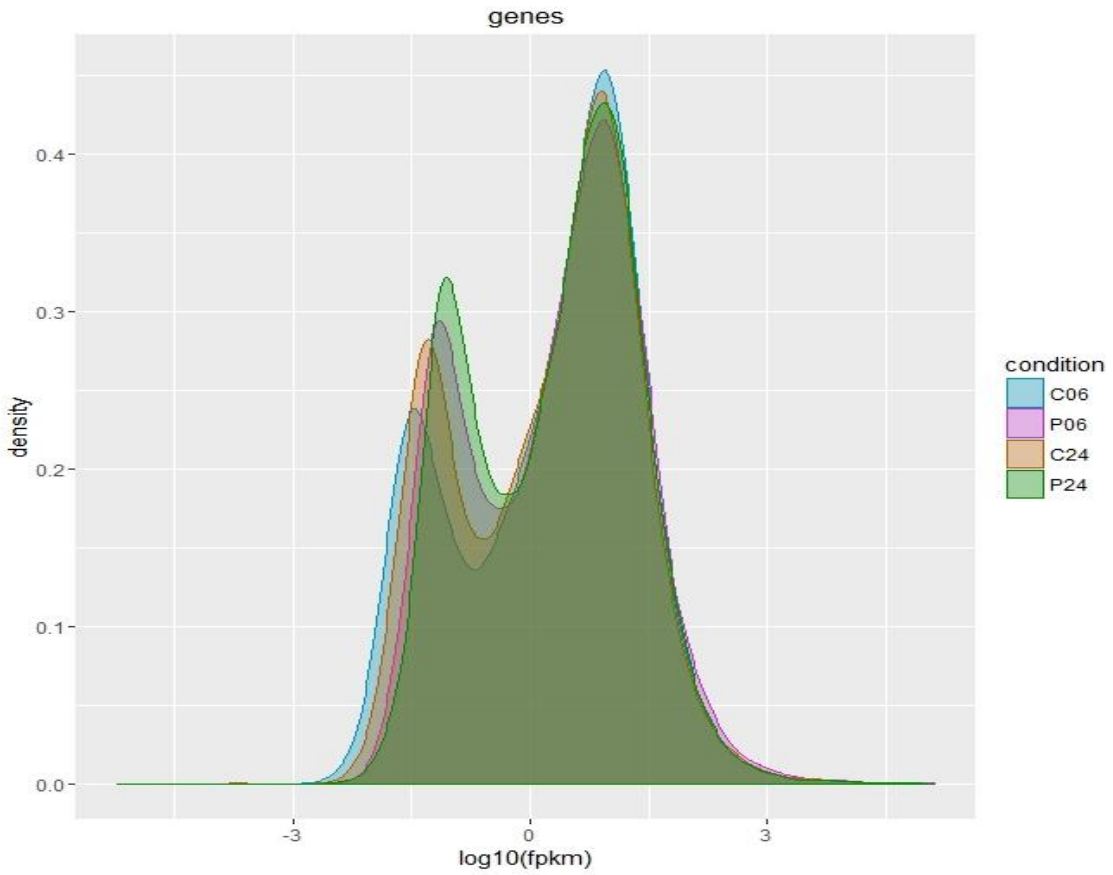

Figure 6: Density plot of Control (C) vs. Treatment (P). Density plot showing the Frequency of FPKM of each sample after 6 and 24 hours. A higher portion of genes at the 24 time point of the treatment condition (P24) are differentially expressed at a log10(FPKM) of around -2, when compared to the remaining conditions. A second peak of differentially expressed genes can be seen at a log10(FPKM) of around 2, these genes belong to the 6 hour time point of the control condition.

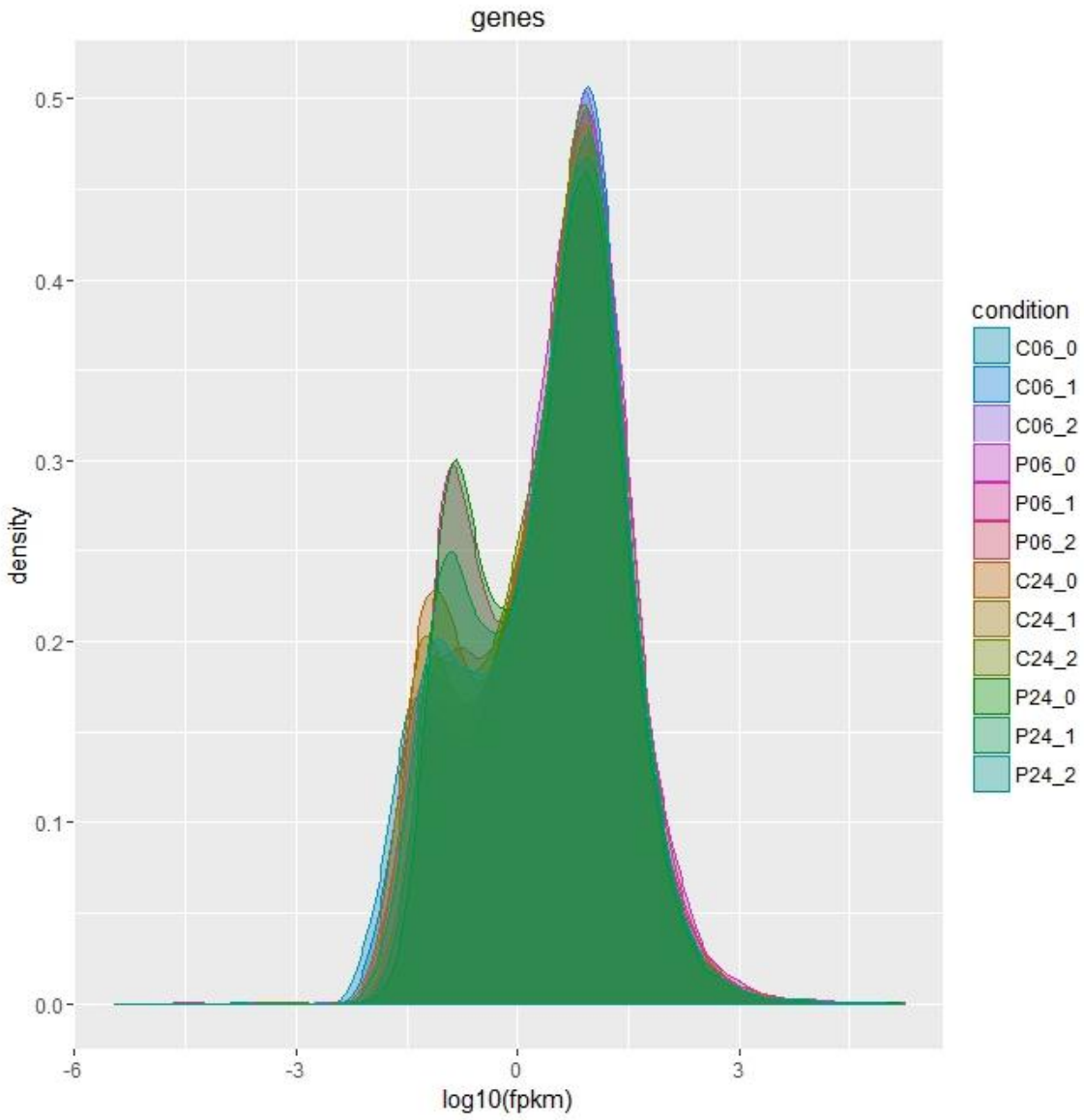

Figure 7: Density plot of Control vs. Treatment (all replicates). Density plot showing the FPKM of each sample in all replicates of each condition. Two prominent peaks can be seen, one at around log10(FPKM) of -1 where the differentially expressed genes belong to the 24 hour treatment time point. The second peak appears at around log10(FPKM) of 1 and the differentially expressed genes belong to the 6 hour time point condition.

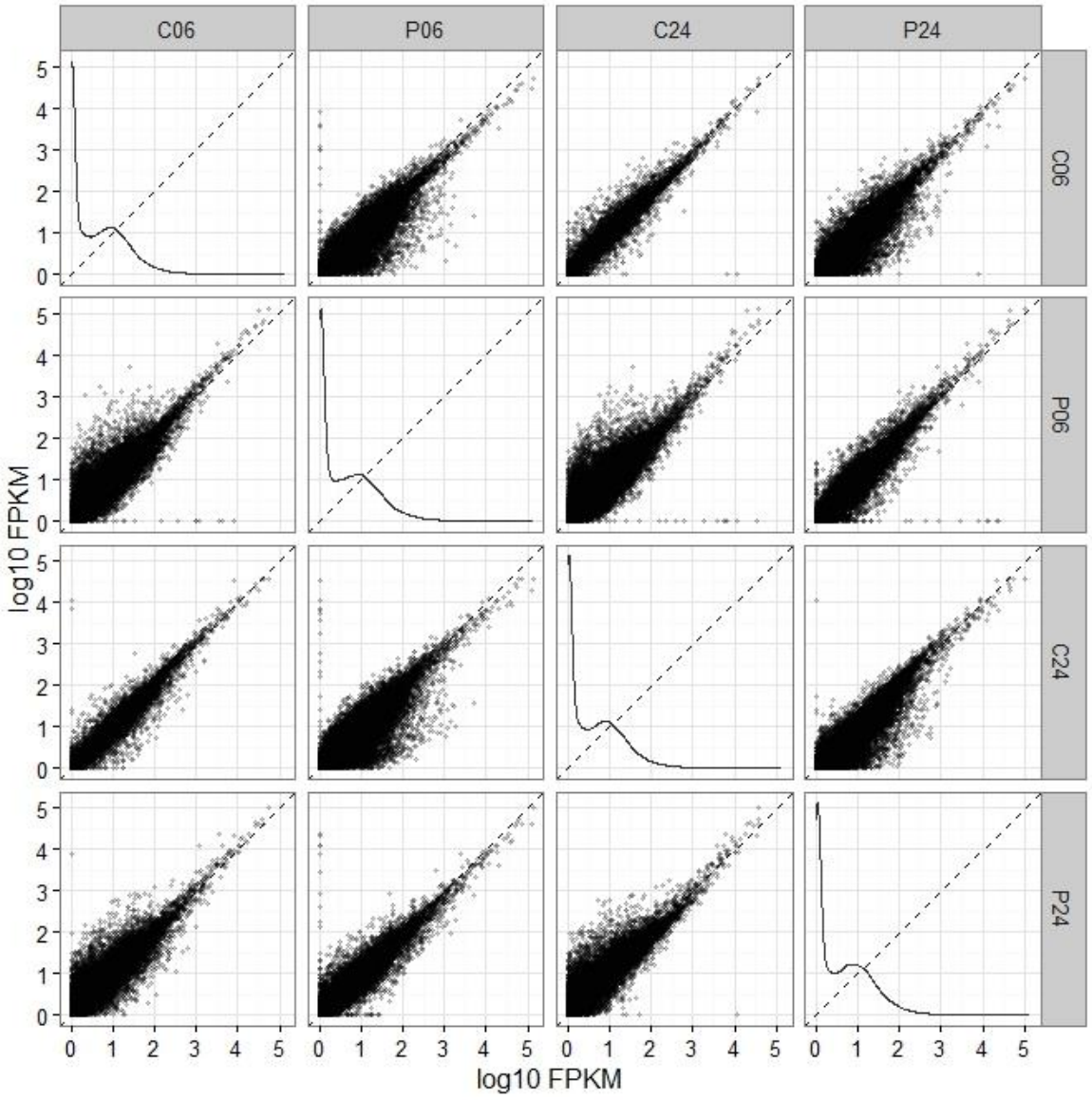

Figure 8: Scatter plot illustrating global trends in the gene expression between pairs of conditions using normalized expression values (Log10FPKM). Scatter plots generated from a comparison between the different time points of the control condition show a similar expression; while the plots generated from a comparison of treatment vs. control show a largely dissimilar expression pattern with a wide range of expression regardless of the time point.

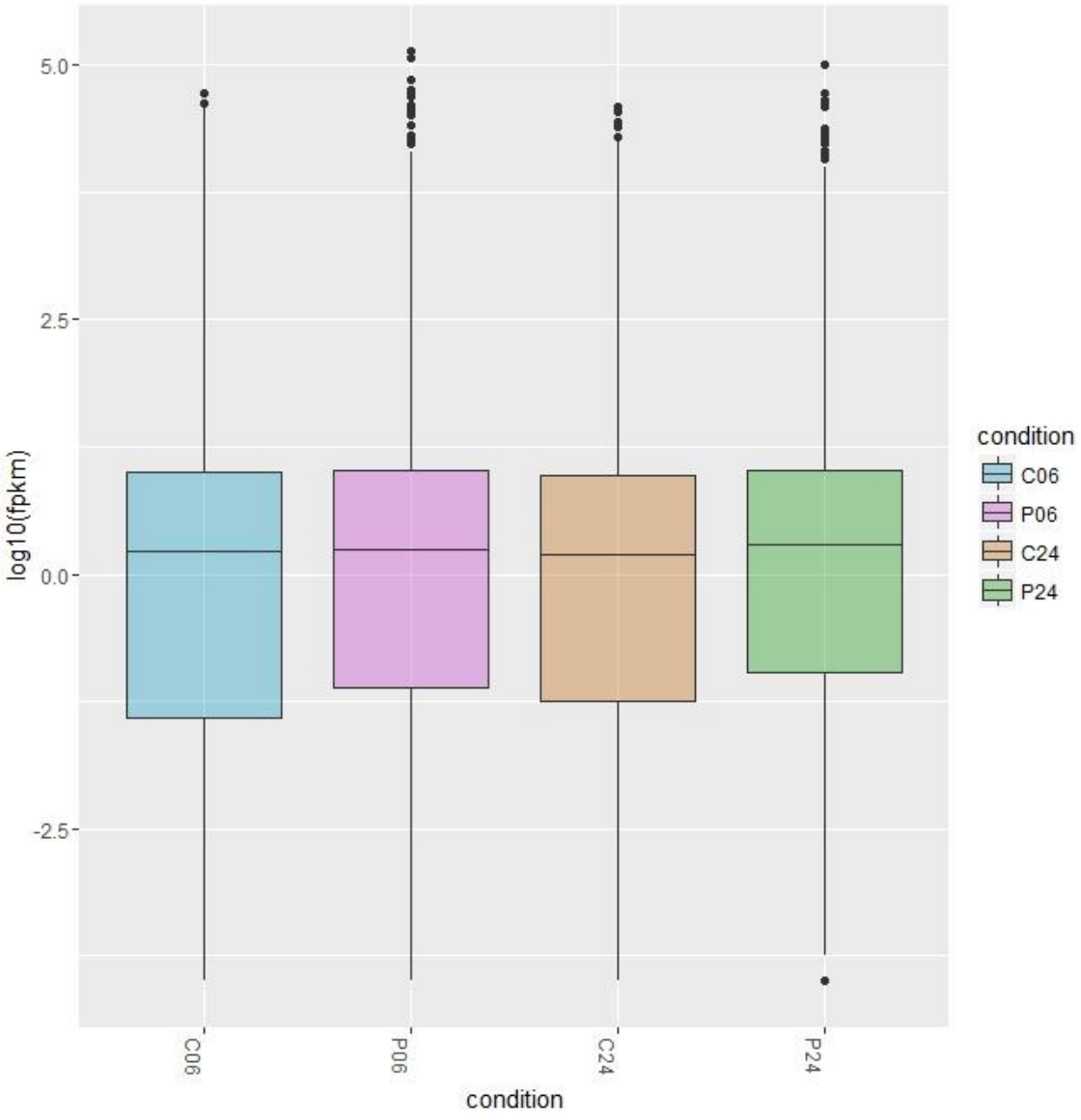

Figure 9: Boxplot of different time points for the 2 conditions, showing the distributional characteristics of a group of values and illustrating an overall pattern of the response of the genes in each condition. All samples show a similar pattern of expression as indicated by the position of the middle quartile and the median line. The slightly larger size of the control conditions suggests a wide range of differential expression.

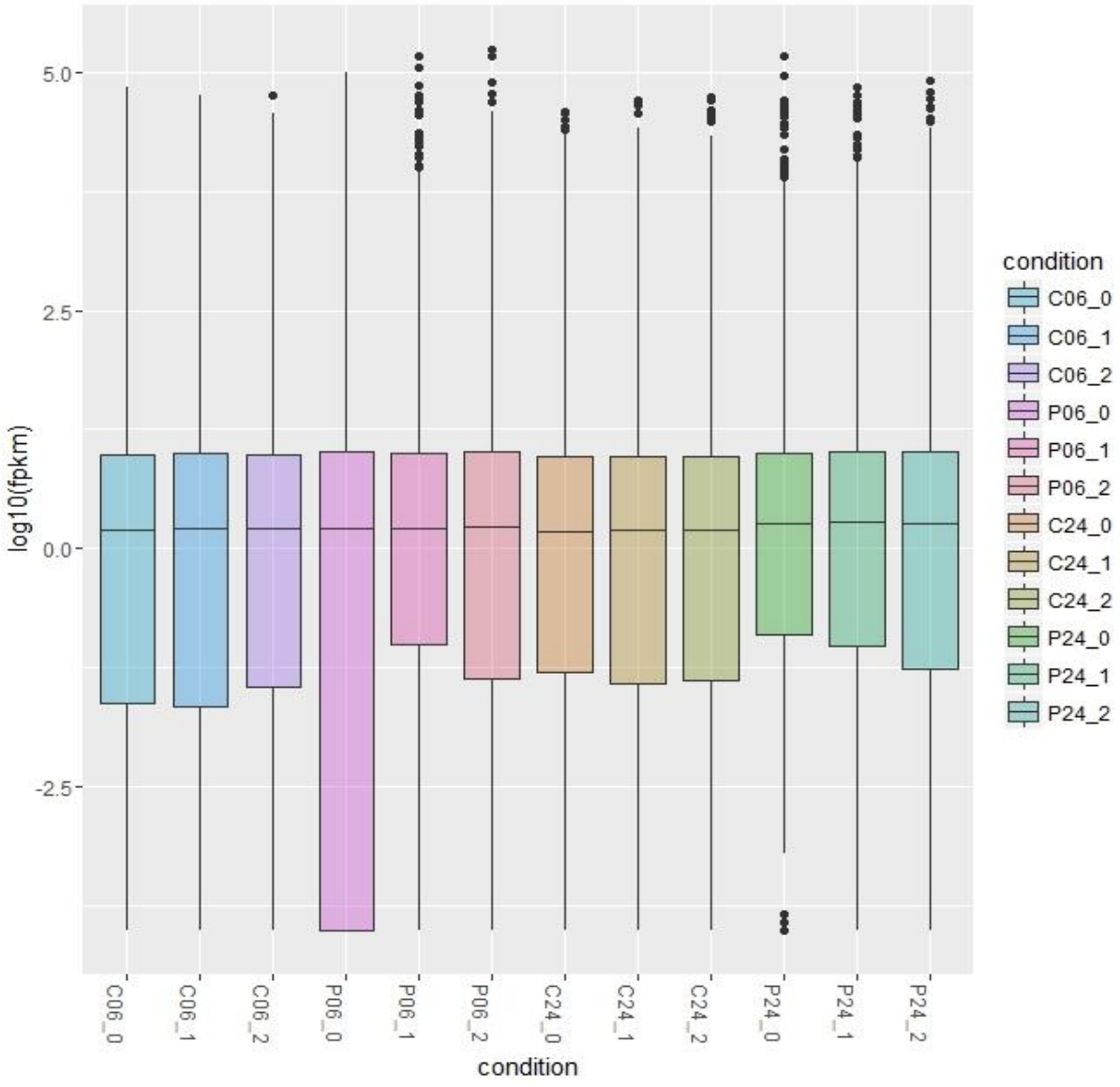

Figure 10: Boxplot with all replicates for both conditions. The 6 hour time point show a wider range of differential expression values with P06_0 (the zero replicate for the 6 hour time point of the treatment condition) having the tallest box indicating the widest range of differential expression throughout all the samples. The 24 hour time point samples show a similar distribution pattern of expression with the P24_0 (zero replicate of the 24 hour time point of the treatment condition) having the smallest variation in expression. Across all samples the median line is located at a similar location and intersecting the middle quartile into two parts.

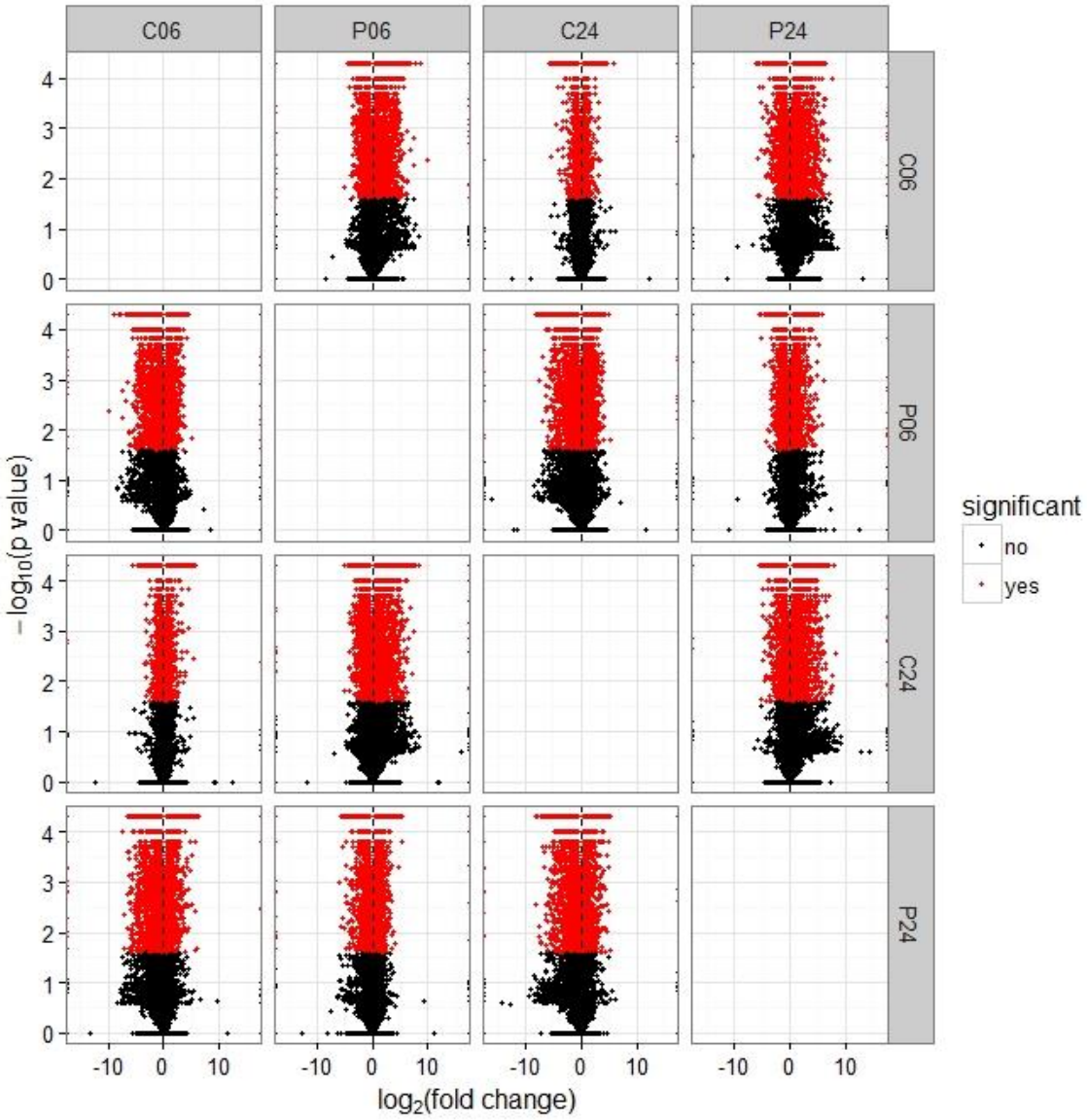

Figure 11: Volcano plots of –log10 p value vs. log fold change. The plots represent the statistical and biologically significant genes, represented by red dots. The biologically and statistically significant differentially expressed transcripts are located in the upper middle section of each comparison.

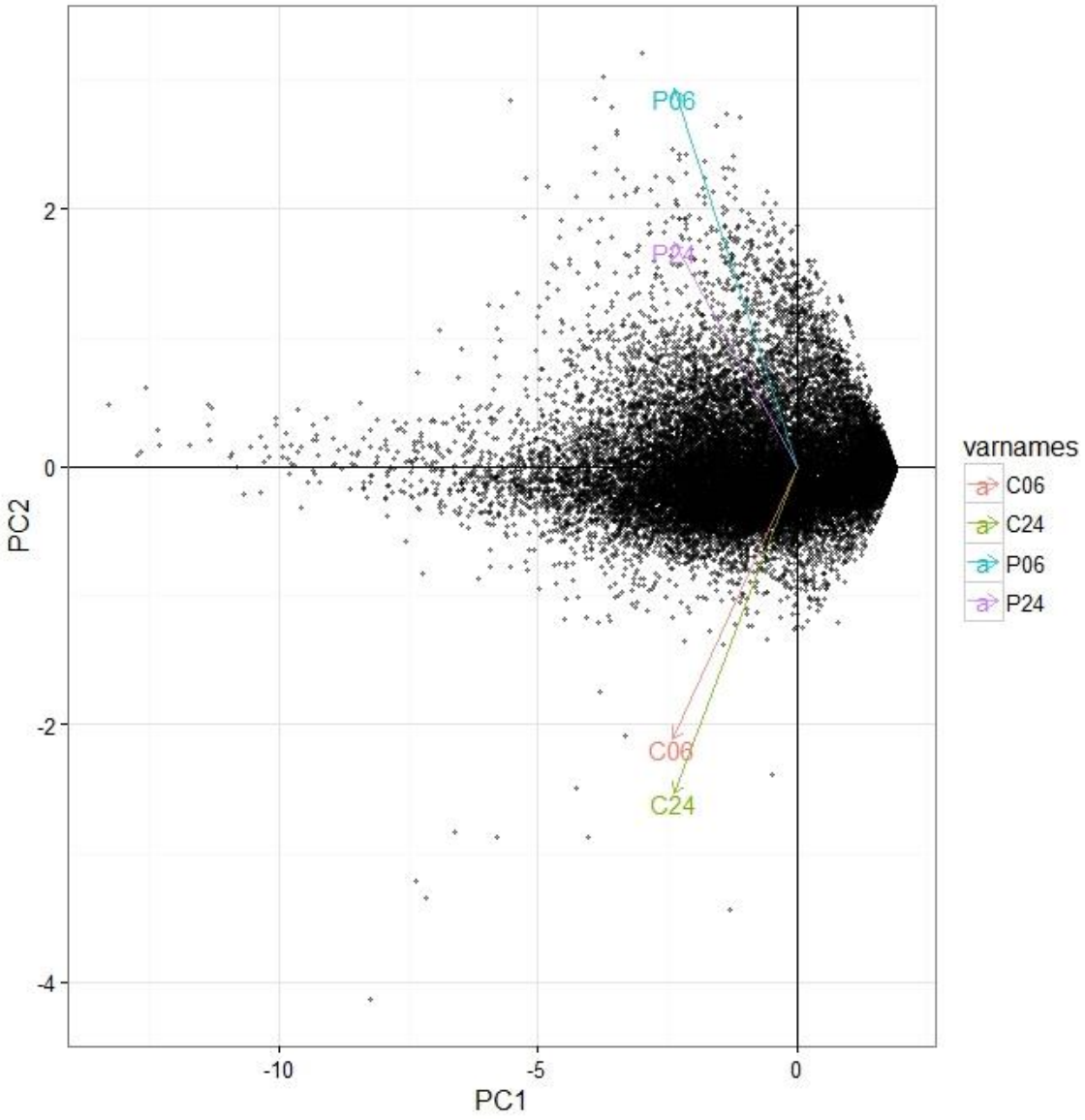

Figure 12: PCA (Principle Component Analysis) of PC1 vs. PC2. PC1 (Principle Component 1) the largest component that accounts for the most variation in the data is unable to differentiate between the different conditions. PC2 (Principle Component 2) is the second component having the highest variance possible under the condition that it is uncorrelated to the preceding components; is able to differentiate between the conditions and separating the control condition (located in the lower left quartile of the graph) from the treatment condition (located in the upper left quartile in the graph)

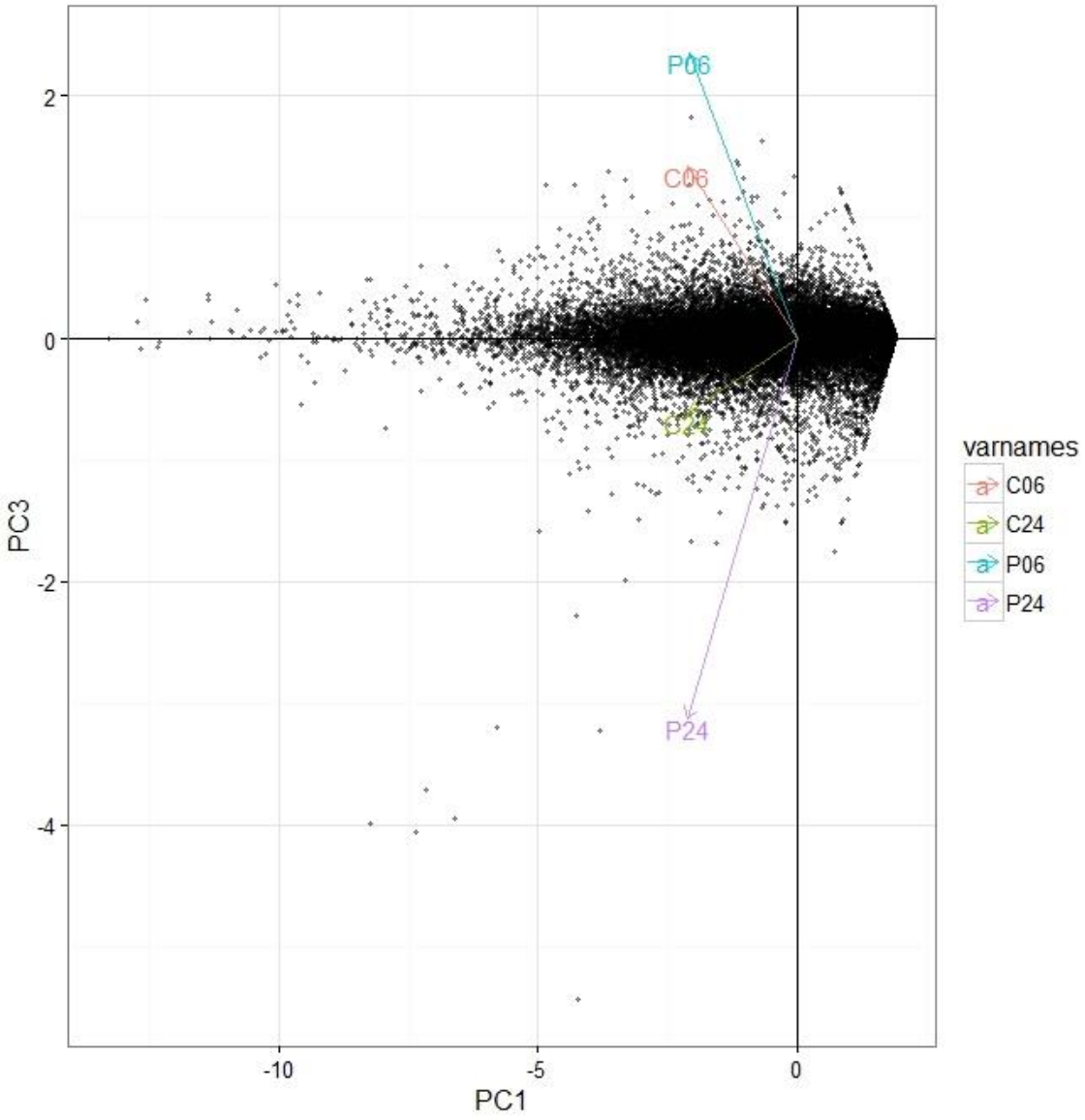

Figure 13: PCA (Principle Component Analysis) of PC1 vs. PC3. In this comparison PC3, the third principle component is able to further differentiate between the conditions by separating them according to the time point. The 24 hour time point samples are located in the lower left quartile, both the control and the treatment condition. The 6 hour sample from both the control and the treatment is located in the upper left quartile.

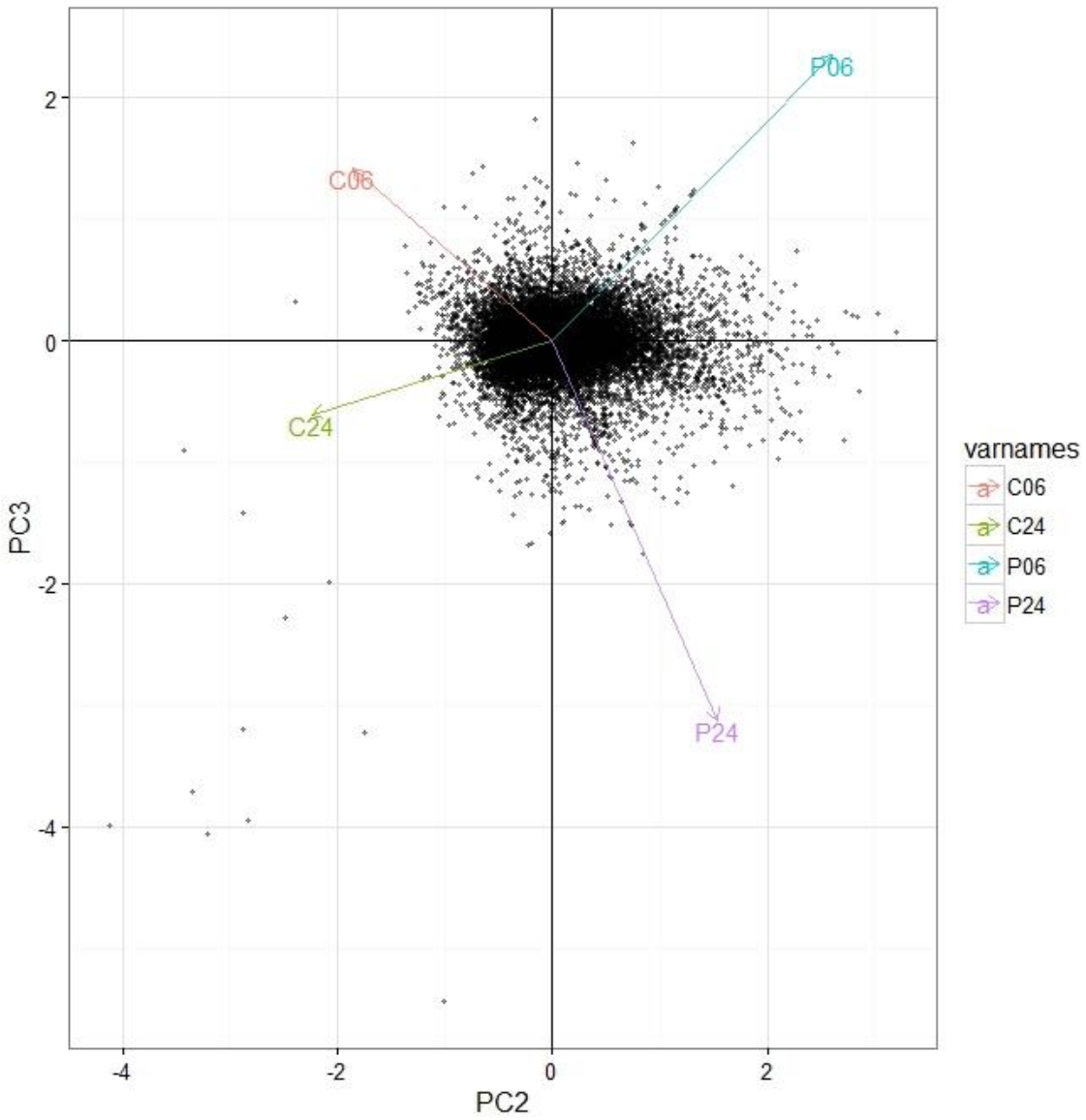

Figure 14: PCA (Principle Component Analysis) of PC2 vs. PC3. A comparison between the two principle components that were able to differentiate between the conditions is able to place each of the 4 samples (2 conditions with 2 time points each for a total of 4 samples) into an individual quartile with the 24 hour time point samples being placed in the lower half and the 6 hour time points located in the upper half.

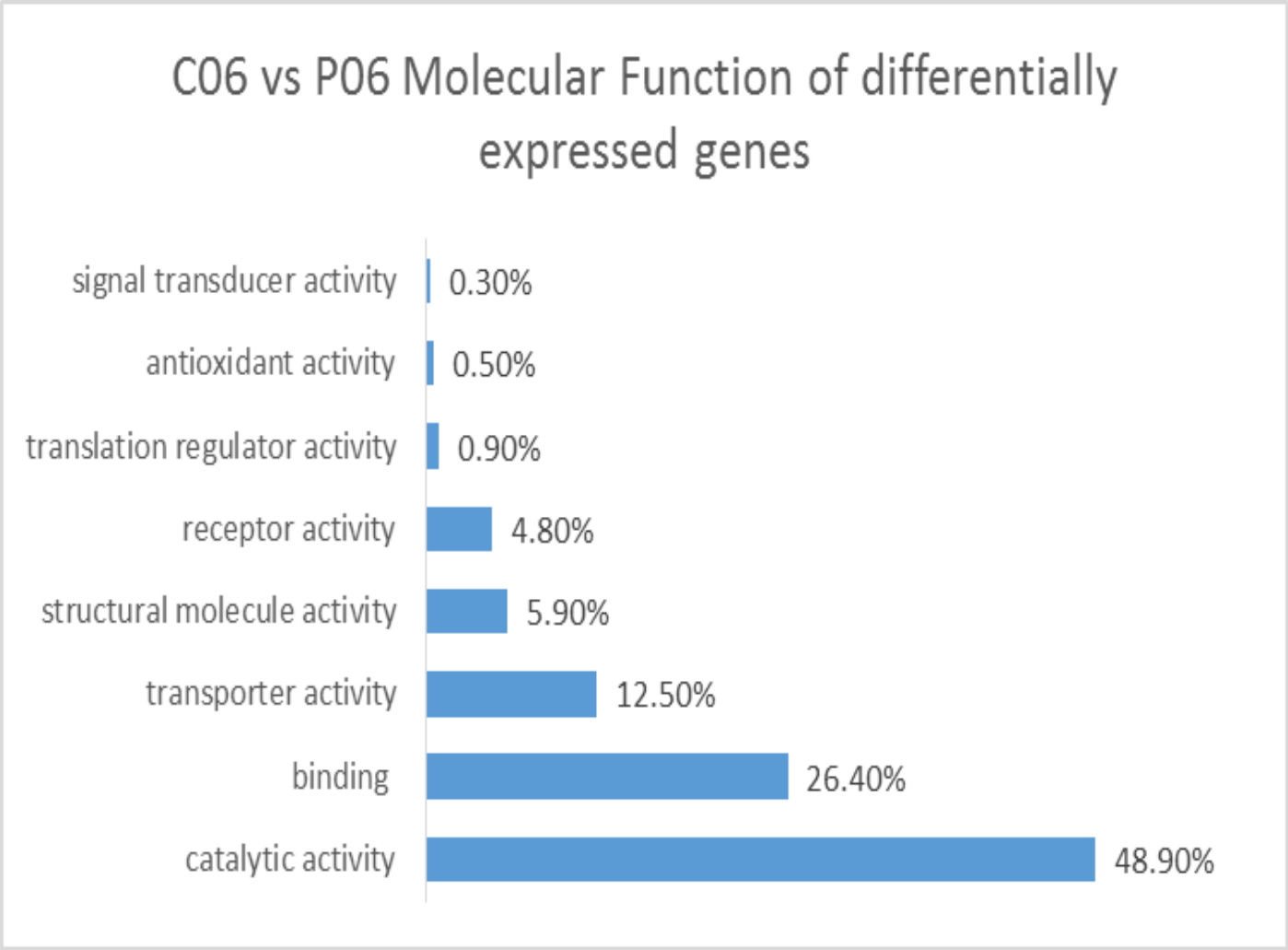

Figure 15: CO6 vs P06 Molecular function of differentially expressed genes. Molecular Function of the 12646 significant differentially expressed genes from the 6 hour time point (Control vs. Treatment)

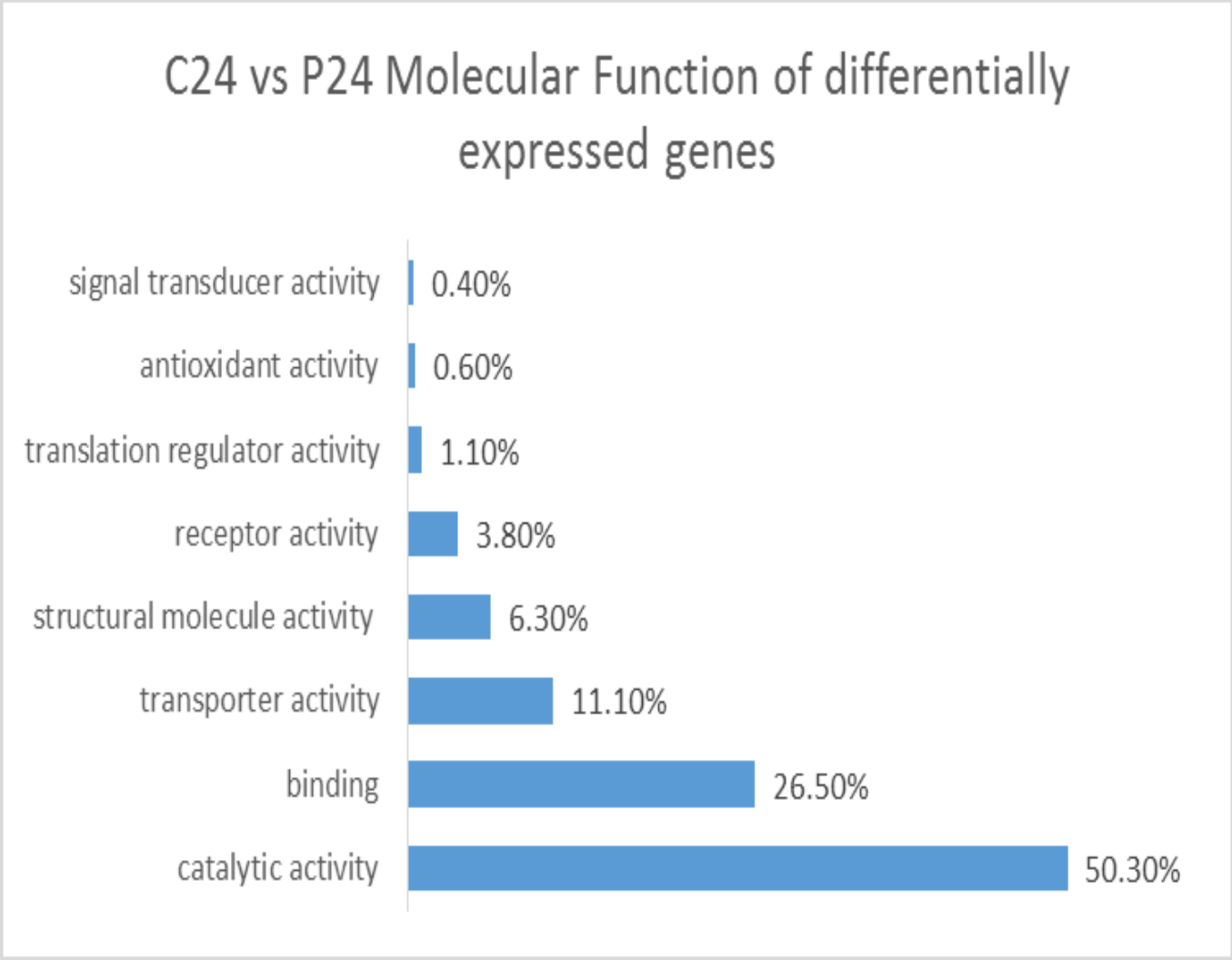

Figure 16: C24 vs P24 Molecular function of differentially expressed genes. Molecular Function of the 10649 significant differentially expressed genes from the 24 hour time point (Control vs. Treatment)

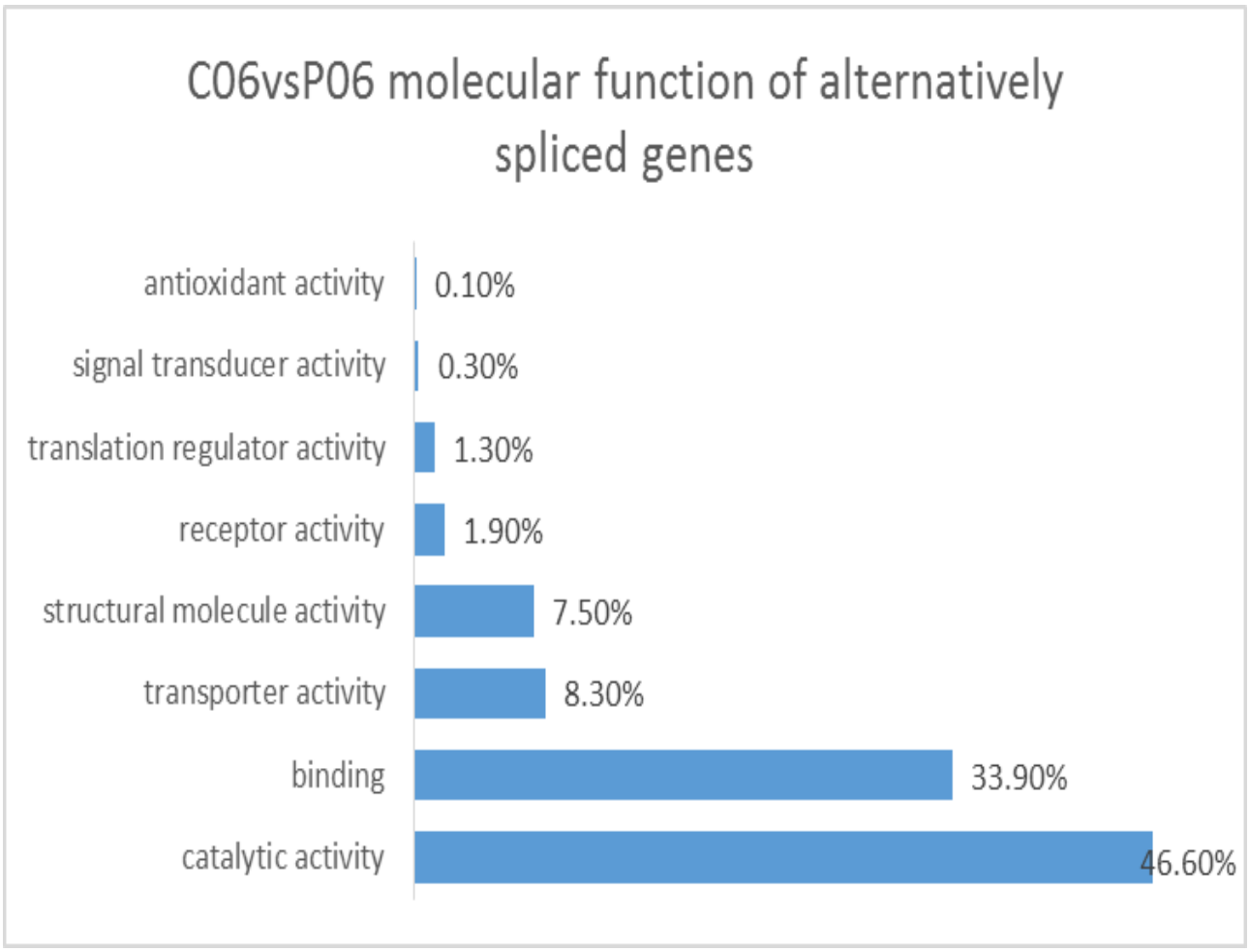

Figure 17: CO6 vs P06 Molecular function of alternatively spliced differentially expressed genes. Molecular Function of the 3763 significant differentially expressed alternatively spliced genes in the 6 hour time point.

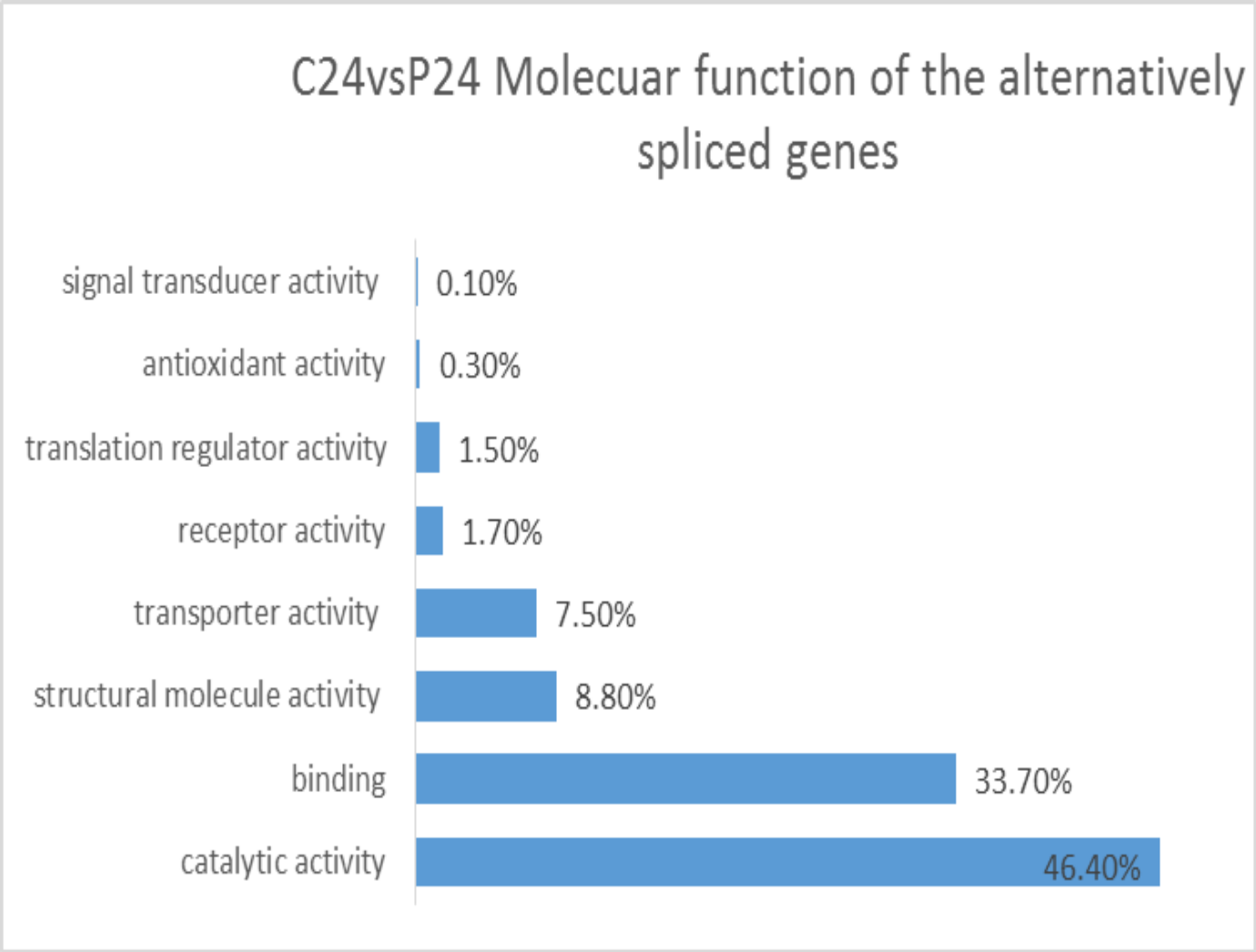

Figure 18: C24 vs P24 Molecular function of alternatively spliced differentially expressed genes. Molecular Function of the 2797 significant differentially expressed alternatively spliced genes in the 24 hour time point.

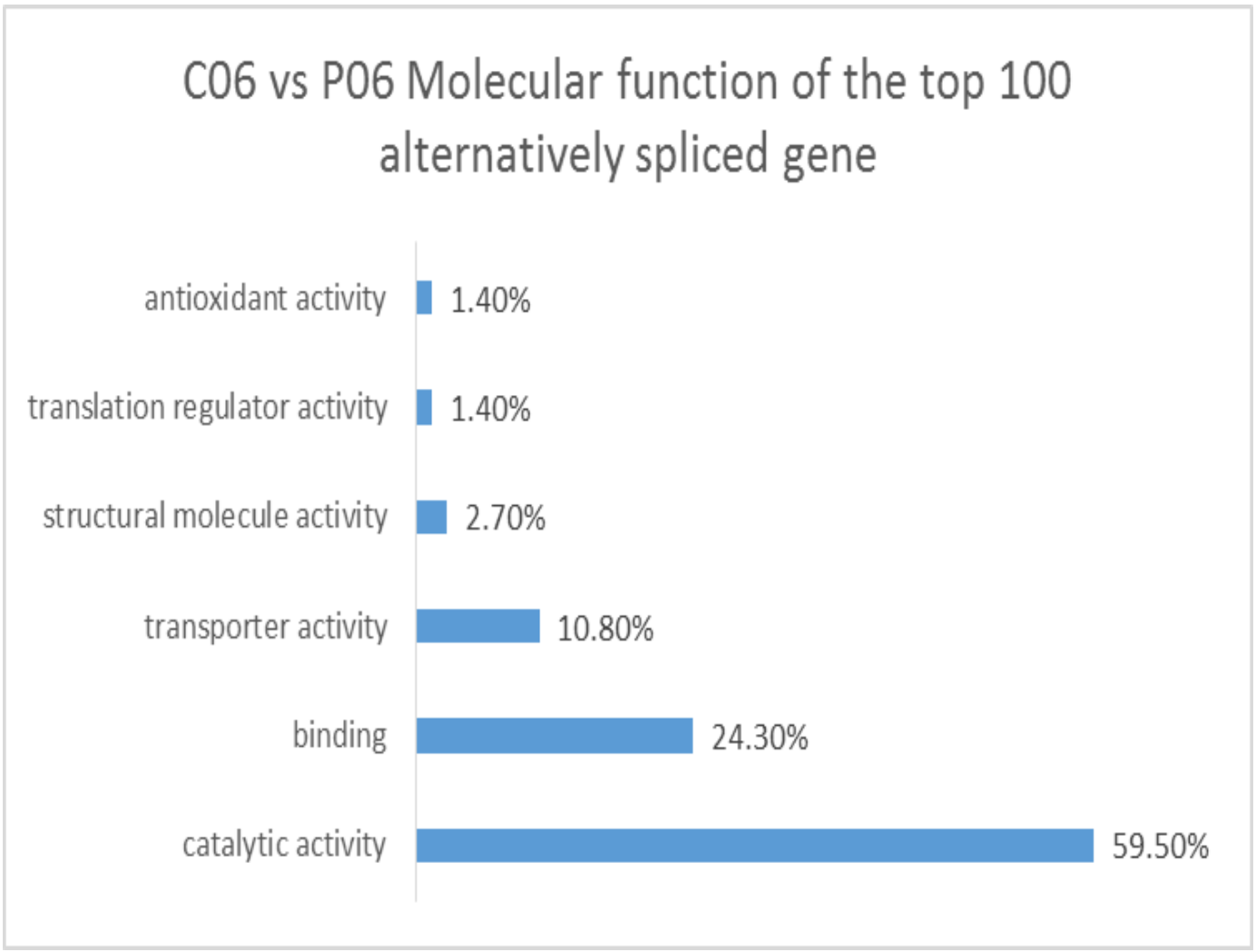

Figure 19: CO6 vs P06 molecular function of the top 100 alternatively spliced genes.

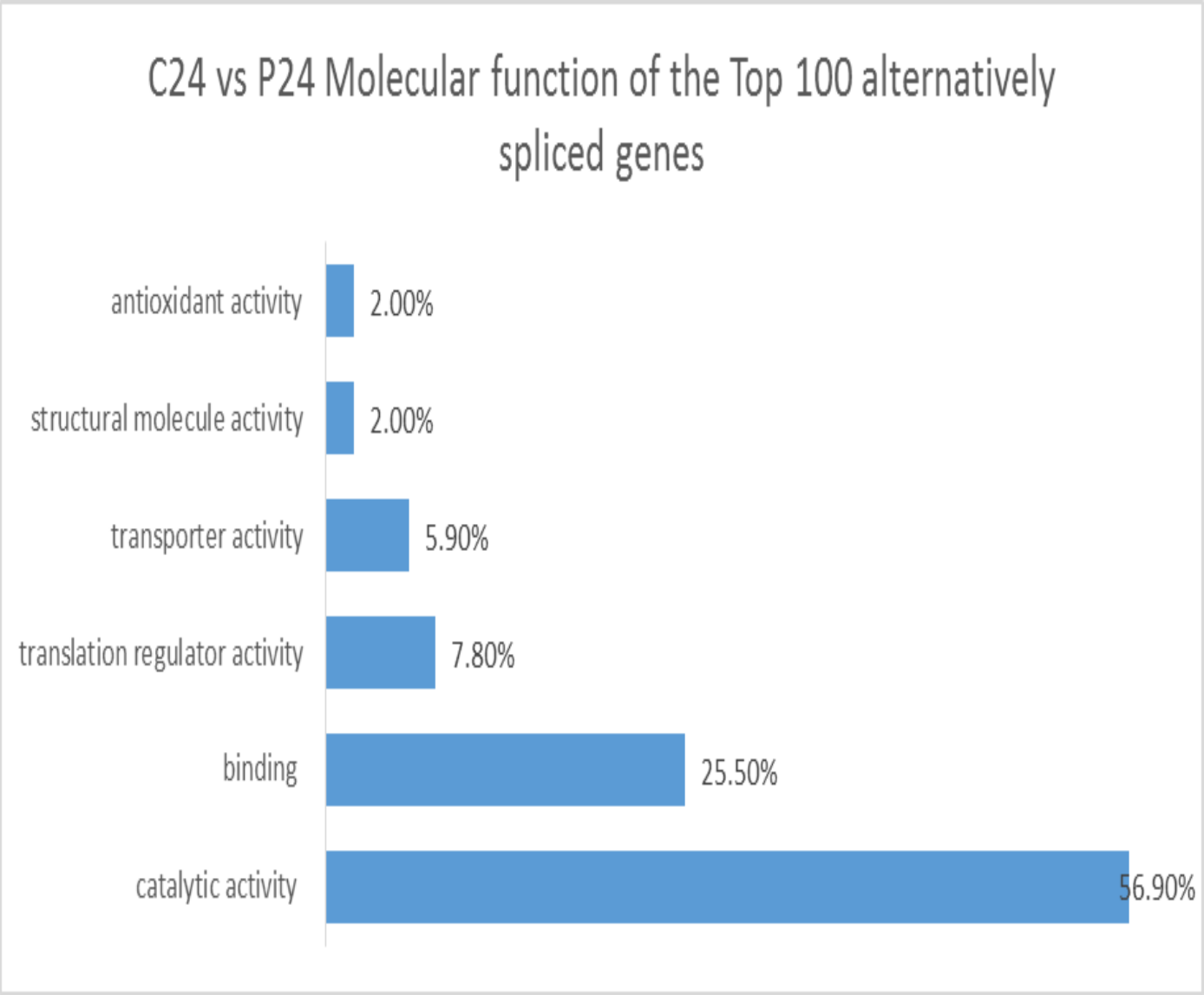

Figure 20: C24 vs P24 molecular function of the top 100 alternatively spliced genes.

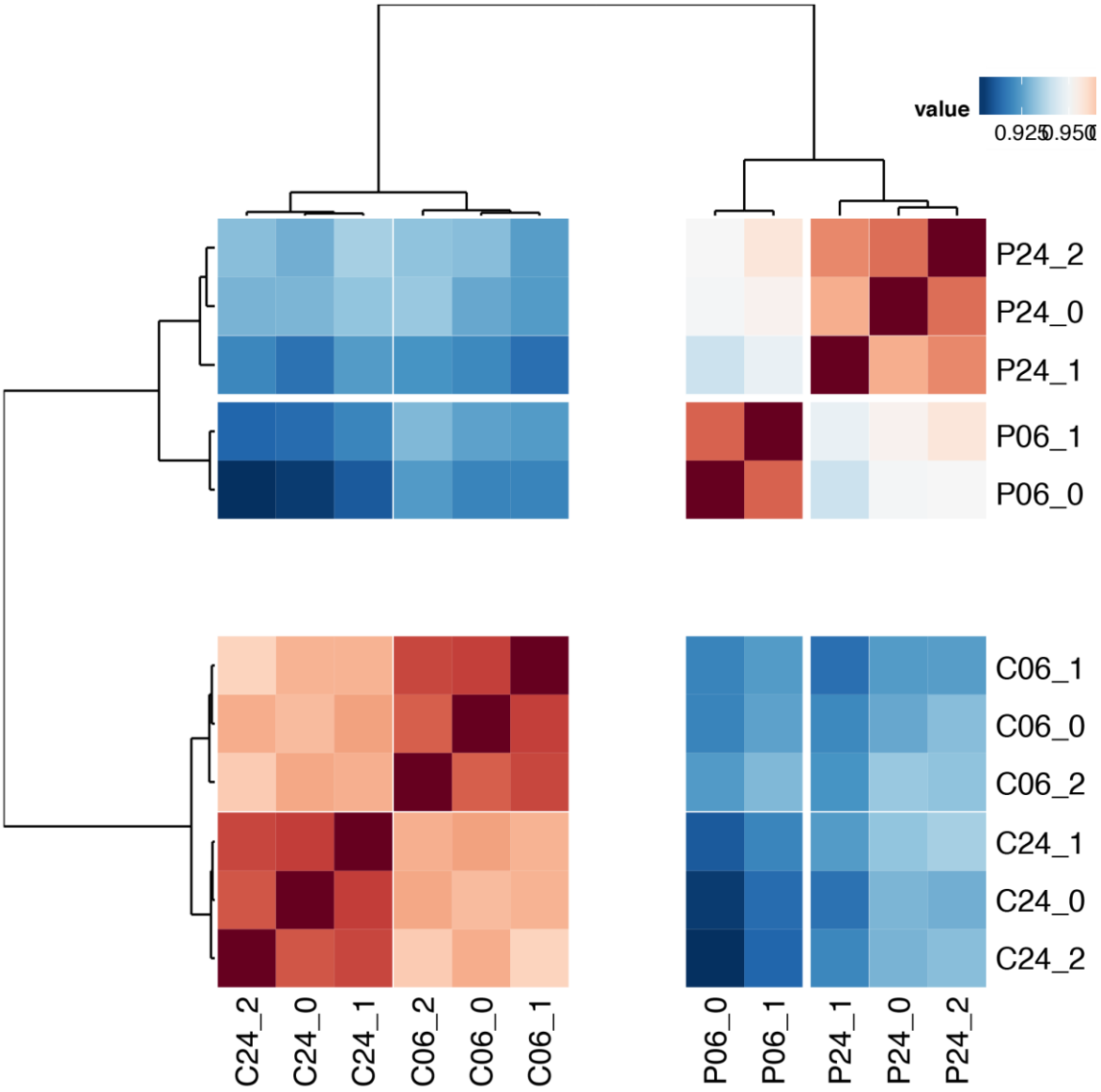

Figure 21: Heat map showing the correlation of gene expression between samples and replicates.

Mapped ID	Gene Name/Gene Symbol	PANTHER Family/Subfamily	PANTHER Protein Class
AT1G05730	Uncharacterized protein;At1g05730;ortholog	PROTEIN FAM136A (PTHR21096:SF1)	
RPL2	60S ribosomal protein L8-1;RPL8A;ortholog	60S RIBOSOMAL PROTEIN L8 (PTHR13691:SF28)	ribosomal protein(PC00171)
ASG4	Protein REVEILLE 3;RVE3;ortholog	MYB-LIKE PROTEIN G (PTHR12802:SF71)	chromatin/chromatin-binding protein(PC00171);transcription cofactor(PC00009)
AT1G01500	Uncharacterized protein At1g01500;At1g01500;ortholog	SUBFAMILY NOT NAMED (PTHR42938:SF4)	dehydrogenase(PC00176)
ADF11	Putative actin-depolymerizing factor 11;ADF11;ortholog	ACTIN-DEPOLYMERIZING FACTOR 1-RELATED (PTHR11913:SF34)	non-motor actin binding protein(PC00085)
RPS19	40S ribosomal protein S19, mitochondrial;RPS19;ortholog	40S RIBOSOMAL PROTEIN S19, MITOCHONDRIAL (PTHR11880:SF13)	ribosomal protein(PC00171)
RPS11	30S ribosomal protein S11, chloroplastic;rps11;ortholog	28S RIBOSOMAL PROTEIN S11, MITOCHONDRIAL (PTHR11759:SF18)	ribosomal protein(PC00171)
AN	C-terminal binding protein AN;AN;ortholog	C-TERMINAL BINDING PROTEIN AN (PTHR43254:SF2)	dehydrogenase(PC00176)

BT3	BTB/POZ and TAZ domain-containing protein 3;BT3;ortholog	BTB/POZ AND TAZ DOMAIN-CONTAINING PROTEIN 3 (PTHR24413:SF156)	non-motor actin binding protein(PC00085);serine protease(PC00041);transcription cofactor(PC00165)
SYTE	Synaptotagmin-5;SYT5;ortholog	C2 DOMAIN-CONTAINING PROTEIN (PTHR10774:SF119)	
AT1G01240	At1g01240/F6F3_11;F633.5;ortholog	EXPRESSED PROTEIN (PTHR33868:SF5)	
YCF10	Chloroplast envelope membrane protein;cemA;ortholog	CHLOROPLAST ENVELOPE MEMBRANE PROTEIN (PTHR33650:SF5)	
RPS7	Ribosomal protein S7, mitochondrial;RPS7;ortholog	RIBOSOMAL PROTEIN S7, MITOCHONDRIAL (PTHR11205:SF28)	ribosomal protein(PC00171)
AT1G04790	At1g04790;F13M7.22;ortholog	SUBFAMILY NOT NAMED (PTHR22763:SF134)	
PSBJ	Photosystem II reaction center protein J;psbJ;ortholog	PHOTOSYSTEM II REACTION CENTER PROTEIN J (PTHR34812:SF3)	
AT1G01810	T1N6.23;T1N6.23;ortholog		
NDHI	NAD(P)H-quinone oxidoreductase subunit I, chloroplastic;ndhI;ortholog	NAD(P)H-QUINONE OXIDOREDUCTASE SUBUNIT I, CHLOROPLASTIC (PTHR10849:SF24)	dehydrogenase(PC00176);reductase(PC00092)
RPL16	60S ribosomal protein L16, mitochondrial;RPL16;ortholog	39S RIBOSOMAL PROTEIN L16, MITOCHONDRIAL (PTHR12220:SF16)	ribosomal protein(PC00171)

RPS11	40S ribosomal protein S11-1;RPS11A;ortholog	40S RIBOSOMAL PROTEIN S11 (PTHR10744:SF15)	ribosomal protein(PC00171)
RPS15	30S ribosomal protein S15, chloroplastic;rps15;ortholog	30S RIBOSOMAL PROTEIN S15, CHLOROPLASTIC (PTHR23321:SF38)	ribosomal protein(PC00171)
RPL32	50S ribosomal protein L32, chloroplastic;rpl32;ortholog	50S RIBOSOMAL PROTEIN L32, CHLOROPLASTIC (PTHR36083:SF3)	
RPS15	40S ribosomal protein S15-1;RPS15A;ortholog	40S RIBOSOMAL PROTEIN S15 (PTHR11880:SF15)	ribosomal protein(PC00171)
YCF3	Photosystem I assembly protein Ycf3;ycf3;ortholog	PHOTOSYSTEM I ASSEMBLY PROTEIN YCF3 (PTHR26312:SF111)	
RPL2	60S ribosomal protein L2, mitochondrial;RPL2;ortholog	60S RIBOSOMAL PROTEIN L2-RELATED (PTHR13691:SF30)	ribosomal protein(PC00171)
GATL5	Probable galacturonosyltransferase-like 5;GATL5;ortholog	GALACTURONOSYLTRANSFERASE-LIKE 5-RELATED (PTHR13778:SF20)	
RPL14	50S ribosomal protein L14, chloroplastic;rpl14;ortholog	50S RIBOSOMAL PROTEIN L14, CHLOROPLASTIC (PTHR11761:SF22)	ribosomal protein(PC00171)
RPL20	50S ribosomal protein L20, chloroplastic;rpl20;ortholog	50S RIBOSOMAL PROTEIN L20, CHLOROPLASTIC (PTHR10986:SF17)	ribosomal protein(PC00171)

NDHE	NAD(P)H-quinone oxidoreductase subunit 4L, chloroplastic;ndhE;ortholog	NAD(P)H-QUINONE OXIDOREDUCTASE SUBUNIT 4L, CHLOROPLASTIC (PTHR11434:SF7)	dehydrogenase(PC00176);reductase(PC00092)
YCF9	Photosystem II reaction center protein Z;psbZ;ortholog	PHOTOSYSTEM II REACTION CENTER PROTEIN Z (PTHR34971:SF3)	
ATAF1	NAC domain-containing protein 2;NAC002;ortholog	NAC DOMAIN-CONTAINING PROTEIN 2 (PTHR31719:SF49)	
RPS19	30S ribosomal protein S19, chloroplastic;rps19;ortholog	30S RIBOSOMAL PROTEIN S19, CHLOROPLASTIC (PTHR11880:SF16)	ribosomal protein(PC00171)
ORF110A	Putative uncharacterized mitochondrial protein AtMg00280;AtMg00280;ortholog	RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN, CATALYTIC DOMAIN (PTHR23321:SF36)	ribosomal protein(PC00171)
PSAC	Photosystem I iron-sulfur center;psaC;ortholog	PHOTOSYSTEM I IRON-SULFUR CENTER (PTHR24960:SF59)	
AT1G04210	At1g04210;F20D22.2;ortholog	INHIBITOR OF NUCLEAR FACTOR KAPPA-B KINASE EPSILON SUBUNIT HOMOLOG 1-RELATED (PTHR24359:SF27)	protein kinase(PC00220)
MAC3A	Pre-mRNA-processing factor 19 homolog 1;PRP19A;ortholog	PRE-MRNA-PROCESSING FACTOR 19 (PTHR22840:SF14)	mRNA splicing factor(PC00171)
YCF6	Cytochrome b6-f complex subunit 8;petN;ortholog	CYTOCHROME B6-F COMPLEX SUBUNIT 8 (PTHR35773:SF3)	

RPL16	50S ribosomal protein L16, chloroplastic;rpl16;ortholog	50S RIBOSOMAL PROTEIN L16, CHLOROPLASTIC-RELATED (PTHR12220:SF17)	ribosomal protein(PC00171)
KAB1	Probable voltage-gated potassium channel subunit beta;KAB1;ortholog	HYPERKINETIC, ISOFORM M (PTHR43150:SF5)	reductase(PC00176);voltage-gated potassium channel(PC00198)
ATGLR3	Glutamate receptor 2.1;GLR2.1;ortholog	GLUTAMATE RECEPTOR 2.1-RELATED (PTHR18966:SF317)	
CKL13	At1g04440;CKL13;ortholog	CASEIN KINASE 1-LIKE PROTEIN 8 (PTHR11909:SF260)	non-receptor serine/threonine protein kinase(PC00220)
PSBF	Cytochrome b559 subunit beta;psbF;ortholog	CYTOCHROME B559 SUBUNIT BETA (PTHR33391:SF13)	
AT1G01730	Putative uncharacterized protein At1g01730;T1N6.14;ortholog	SUBFAMILY NOT NAMED (PTHR35459:SF3)	
AT1G01800	AT1G01800 protein;At1g01800;ortholog	(+)-NEOMENTHOL DEHYDROGENASE-RELATED (PTHR43490:SF23)	dehydrogenase(PC00176);reductase(PC00092)
NDHJ	NAD(P)H-quinone oxidoreductase subunit J, chloroplastic;ndhJ;ortholog	NAD(P)H-QUINONE OXIDOREDUCTASE SUBUNIT J, CHLOROPLASTIC (PTHR10884:SF12)	

Table 2: Functional classification of the genes uniquely expressed at the 6 hour time point.

Mapped ID	Gene Name/Gene Symbol	PANTHER Family/Subfamily	PANTHER Protein Class
GTE4	Transcription factor GTE4;GTE4;ortholog	TRANSCRIPTION FACTOR GTE4 (PTHR22880:SF203)	acetyltransferase(PC00220);chromatin/chromatin-binding protein(PC00038)
NAD3	NADH-ubiquinone oxidoreductase chain 3;ND3;ortholog	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 3 (PTHR11058:SF16)	
MATR	Maturase (Fragment);matR;ortholog	SUBFAMILY NOT NAMED (PTHR33642:SF5)	
AT1G05140	Probable membrane metalloprotease ARASP2, chloroplastic;ARASP2;ortholog	MEMBRANE METALLOPROTEASE ARASP, CHLOROPLASTIC-RELATED (PTHR42837:SF2)	
COX1	Cytochrome c oxidase subunit 1;COX1;ortholog	CYTOCHROME C OXIDASE SUBUNIT 1 (PTHR10422:SF30)	oxidase(PC00176)
PETG	Cytochrome b6-f complex subunit 5;petG;ortholog	CYTOCHROME B6-F COMPLEX SUBUNIT 5 (PTHR35516:SF3)	
AT1G07170	PHD finger-like domain-containing protein 5B;At1g07170;ortholog	PHD FINGER-LIKE DOMAIN-CONTAINING PROTEIN 5A-RELATED (PTHR13120:SF4)	
AT1G06240	Putative uncharacterized protein At1g06240;At1g06240;ortholog	SUBFAMILY NOT NAMED (PTHR42782:SF1)	oxidoreductase(PC00176)
PCNA1	Proliferating cellular nuclear antigen 1;PCNA;ortholog	PROLIFERATING CELL NUCLEAR ANTIGEN (PTHR11352:SF7)	DNA polymerase processivity factor(PC00171)
ATPI	ATP synthase subunit a, chloroplastic;atpI;ortholog	ATP SYNTHASE SUBUNIT A, CHLOROPLASTIC (PTHR42823:SF2)	

RCI3	Peroxidase 3;PER3;ortholog	PEROXIDASE 3-RELATED (PTHR31235:SF88)	
CHR9	Switch 2;SWI2;ortholog	SWITCH 2 (PTHR10799:SF900)	DNA helicase(PC00171)
AT1G05960	ARM repeat superfamily protein;At1g05960;ortholog	ARM REPEAT SUPERFAMILY PROTEIN (PTHR12444:SF13)	
RPL4	50S ribosomal protein L4, chloroplastic;RPL4;ortholog	39S RIBOSOMAL PROTEIN L4, MITOCHONDRIAL (PTHR10746:SF13)	
CER1	Protein ECERIFERUM 1;CER1;ortholog	PROTEIN CER1-LIKE 1-RELATED (PTHR11863:SF82)	hydroxylase(PC00176);oxidase(PC00122)
AT1G01300	Aspartyl protease family protein;F6F3.10;ortholog	ASPARTYL PROTEASE-RELATED (PTHR13683:SF459)	aspartic protease(PC00121)
RPL4	60S ribosomal protein L4-2;RPL4D;ortholog	60S RIBOSOMAL PROTEIN L4 (PTHR19431:SF2)	
UGT71C4	UDP-glycosyltransferase 71C4;UGT71C4;ortholog	UDP-GLYCOSYLTRANSFERASE 71C3-RELATED (PTHR11926:SF500)	
ATNDI1	Internal alternative NAD(P)H-ubiquinone oxidoreductase A1, mitochondrial;NDA1;ortholog	NADH DEHYDROGENASE-RELATED (PTHR43706:SF3)	dehydrogenase(PC00176);oxidase(PC00092);reductase(PC00175)
AT1G06640	1-aminocyclopropane-1-carboxylate oxidase homolog 2;At1g06640;ortholog	1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE HOMOLOG 1-RELATED (PTHR10209:SF322)	
AT1G06470	Probable sugar phosphate/phosphate translocator At1g06470;At1g06470;ortholog	NUCLEOTIDE-SUGAR TRANSPORTER YMD8-RELATED (PTHR11132:SF174)	transporter(PC00227)
AXR3	Auxin-responsive protein IAA17;IAA17;ortholog	AUXIN-RESPONSIVE PROTEIN IAA14-RELATED (PTHR31734:SF81)	

NIH	Nuclear DEIH-boxhelicase;NIH;ortholog	ATP-DEPENDENT RNA HELICASE YTHDC2-RELATED (PTHR18934:SF184)	RNA helicase(PC00171)
AT1G06650	1-aminocyclopropane-1-carboxylate oxidase homolog 3;At1g06650;ortholog	1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE HOMOLOG 1-RELATED (PTHR10209:SF322)	
BHLH128	Transcription factor bHLH128;BHLH128;ortholog	TRANSCRIPTION FACTOR BHLH128-RELATED (PTHR16223:SF90)	
PDE247	Pentatricopeptide repeat-containing protein At1g05750, chloroplastic;PDE247;ortholog	SUBFAMILY NOT NAMED (PTHR24015:SF1333)	RNA binding protein(PC00171);serine/threonine protein kinase receptor(PC00031);transporter(PC00197)
CAF1-1	Probable CCR4-associated factor 1 homolog 1;CAF1-1;ortholog	CCR4-ASSOCIATED FACTOR 1 HOMOLOG 1-RELATED (PTHR10797:SF16)	transcription factor(PC00218)
AT1G03200	At1g03200;At1g03200;ortholog		
AT1G06380	Ribosomal protein L1p/L10e family;T2D23.8;ortholog	RIBOSOMAL L1 DOMAIN-CONTAINING PROTEIN 1 (PTHR23105:SF95)	ribosomal protein(PC00171)
AT1G03210	F15K9.19 protein;F15K9.19;ortholog	PHENAZINE BIOSYNTHESIS-LIKE DOMAIN-CONTAINING PROTEIN (PTHR13774:SF23)	
AT1G06645	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein;At1g06645;ortholog	1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE HOMOLOG 1-RELATED (PTHR10209:SF322)	

FRO2	Ferric reduction oxidase 2;FRO2;ortholog	FERRIC REDUCTION OXIDASE 2-RELATED (PTHR11972:SF98)	
ABCI19	ABC transporter I family member 19;ABCI19;ortholog	ABC TRANSPORTER I FAMILY MEMBER 19-RELATED (PTHR12847:SF12)	ATP-binding cassette (ABC) transporter(PC00227)
UGT71C5	UDP-glycosyltransferase 71C5;UGT71C5;ortholog	UDP-GLYCOSYLTRANSFERASE 71C1-RELATED (PTHR11926:SF442)	
PETD	Cytochrome b6-f complex subunit 4;petD;ortholog	CYTOCHROME B6-F COMPLEX SUBUNIT 4 (PTHR19271:SF11)	
IAA12	Auxin-responsive protein IAA12;IAA12;ortholog	AUXIN-RESPONSIVE PROTEIN IAA12-RELATED (PTHR31734:SF75)	
ORF315	Uncharacterized ATP synthase C chain-like protein;AtMg00040;ortholog	ATP SYNTHASE F(0) COMPLEX SUBUNIT C3, MITOCHONDRIAL (PTHR10031:SF16)	ATP synthase(PC00227)
ORF118	Uncharacterized mitochondrial protein AtMg01010;AtMg01010;ortholog	SUBFAMILY NOT NAMED (PTHR35289:SF3)	
AT1G06870	Probable thylakoidal processing peptidase 2, chloroplastic;TPP2;ortholog	THYLAKOIDAL PROCESSING PEPTIDASE 1, CHLOROPLASTIC-RELATED (PTHR43390:SF6)	
AT1G06135	Uncharacterized protein;At1g06135;ortholog	SUBFAMILY NOT NAMED (PTHR33592:SF3)	
HSP17	17.4 kDa class I heat shock protein;HSP17.4A;ortholog	17.4 KDA CLASS I HEAT SHOCK PROTEIN-RELATED (PTHR11527:SF225)	chaperone(PC00072)

Table 3: Functional classification of the genes uniquely expressed at 24 hour time point.

Mapped ID	Gene Name/Gene Symbol	PANTHER Family/Subfamily	PANTHER Protein Class
IAA10	Auxin-responsive protein IAA10;IAA10;ortholog	AUXIN-RESPONSIVE PROTEIN IAA10 (PTHR31734:SF61)	
AT1G03230	Aspartyl protease-like protein;F15K9.16;ortholog	ASPARTYL PROTEASE-LIKE PROTEIN (PTHR13683:SF474)	aspartic protease(PC00121)
COX2	Cytochrome c oxidase subunit 2;COX2;ortholog	CYTOCHROME C OXIDASE SUBUNIT 2 (PTHR22888:SF14)	oxidoreductase(PC00176)
AT1G05320	Uncharacterized protein;At1g05320;ortholog		
ORF294	Uncharacterized mitochondrial protein AtMg01200;AtMg01200;ortholog	SUBFAMILY NOT NAMED (PTHR35289:SF3)	
PME7	Probable pectinesterase/pectinesterase inhibitor 7;PME7;ortholog	PECTINESTERASE/PECTINESTERASE INHIBITOR 20-RELATED (PTHR31707:SF84)	
DI19-2	Protein DEHYDRATION-INDUCED 19 homolog 2;DI19-2;ortholog	PROTEIN DEHYDRATION-INDUCED 19 HOMOLOG 2-RELATED (PTHR31875:SF12)	
AT1G01630	Polyphosphoinositide binding protein, putative;T1N6.1;ortholog	BINDING PROTEIN, PUTATIVE-RELATED (PTHR10174:SF190)	dehydrogenase(PC00176)
AT1G02420	Putative pentatricopeptide repeat-containing protein At1g02420;At1g02420;ortholog	SUBFAMILY NOT NAMED (PTHR24015:SF1413)	RNA binding protein(PC00171);serine/threonine protein kinase receptor(PC00031);transporter(PC00197)

BXL2	Probable beta-D-xylosidase 2;BXL2;ortholog	BETA-D-XYLOSIDASE 2-RELATED (PTHR42721:SF7)	glucosidase(PC00121)
AT1G01210	DNA-directed RNA polymerase subunit;F6F3.2;ortholog	DNA-DIRECTED RNA POLYMERASE III SUBUNIT RPC10 (PTHR11239:SF21)	
PRP39	Pre-mRNA-processing factor 39;PRP39;ortholog	PRE-MRNA-PROCESSING FACTOR 39 (PTHR17204:SF32)	mRNA splicing factor(PC00171)
AT1G01350	Zinc finger CCCH domain-containing protein 1;At1g01350;ortholog	PROTEIN RNF113A1-RELATED (PTHR12930:SF4)	nucleic acid binding(PC00171);ubiquitin-protein ligase(PC00142)
NAD7	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2;NAD7;ortholog	NADH DEHYDROGENASE [UBIQUINONE] IRON-SULFUR PROTEIN 2, MITOCHONDRIAL (PTHR11993:SF32)	dehydrogenase(PC00176);reductase (PC00092)
FKGP	Bifunctional fucokinase/fucose pyrophosphorylase;FKGP;ortholog	L-FUCOSE KINASE (PTHR32463:SF1)	
BGLU11	Beta-glucosidase 11;BGLU11;ortholog	BETA-GLUCOSIDASE 1-RELATED (PTHR10353:SF98)	
GBF4	G-box-binding factor 4;GBF4;ortholog	BASIC LEUCINE ZIPPER TRANSCRIPTION FACTOR-RELATED (PTHR22952:SF222)	
PSBN	Protein PsbN;psbN;ortholog	PROTEIN PSBN (PTHR35326:SF4)	
FZL	Probable transmembrane GTPase FZO-like, chloroplastic;FZL;ortholog	TRANSMEMBRANE GTPASE FZO-LIKE, CHLOROPLASTIC-RELATED (PTHR43681:SF1)	small GTPase(PC00095)
AT1G03910	Cactin;CTN;ortholog	CACTIN (PTHR21737:SF11)	
AT1G02350	Protoporphyrinogen oxidase-like protein;At1g02350;ortholog		

URH2	Probable uridine nucleosidase 2;URH2;ortholog	URIDINE NUCLEOSIDASE 2-RELATED (PTHR12304:SF27)	
BPS1	Protein BPS1, chloroplastic;BPS1;ortholog	BYPASS1-RELATED PROTEIN-RELATED (PTHR31509:SF30)	
ATP1	ATP synthase subunit alpha, mitochondrial;ATPA;ortholog	ATP SYNTHASE SUBUNIT ALPHA, MITOCHONDRIAL (PTHR43089:SF1)	ATP synthase(PC00227);DNA binding protein(PC00068);anion channel(PC00002);ligand-gated ion channel(PC00171)
ATP6-1	ATP synthase subunit a-1;ATP6-1;ortholog	ATP SYNTHASE SUBUNIT A (PTHR11410:SF8)	
AT1G04090	At1g04090;At1g04090;ortholog	SUBFAMILY NOT NAMED (PTHR42656:SF4)	
PDLP2	Cysteine-rich repeat secretory protein 3;CRRSP3;ortholog	CYSTEINE-RICH REPEAT SECRETORY PROTEIN 11-RELATED (PTHR32080:SF13)	
BHLH13	Transcription factor bHLH13;BHLH13;ortholog	TRANSCRIPTION FACTOR BHLH13 (PTHR11514:SF75)	
ORF204	Uncharacterized mitochondrial protein AtMg01410;AtMg01410;ortholog	SUBFAMILY NOT NAMED (PTHR34456:SF4)	
PSAJ	Photosystem I reaction center subunit IX;psaJ;ortholog	PHOTOSYSTEM I REACTION CENTER SUBUNIT IX (PTHR36082:SF4)	
NAD4	NADH-ubiquinone oxidoreductase chain 4;ND4;ortholog	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (PTHR43507:SF3)	dehydrogenase(PC00176);reductase (PC00092)
PSBK	Photosystem II reaction center protein K;psbK;ortholog	PHOTOSYSTEM II REACTION CENTER PROTEIN K (PTHR35325:SF2)	

GAE2	UDP-glucuronate 4-epimerase 2;GAE2;ortholog	UDP-GLUCURONATE 4-EPIMERASE 2-RELATED (PTHR43574:SF13)	dehydratase(PC00144);epimerase/racemase(PC00091);oxidoreductase(PC00135)
COX3	Cytochrome c oxidase subunit 3;COX3;ortholog	CYTOCHROME C OXIDASE SUBUNIT 3 (PTHR11403:SF13)	oxidase(PC00176)
PGP11	ABC transporter B family member 11;ABCB11;ortholog	ABC TRANSPORTER B FAMILY MEMBER 11-RELATED (PTHR24221:SF299)	
ATMPK11	Mitogen-activated protein kinase 11;MPK11;ortholog	MITOGEN-ACTIVATED PROTEIN KINASE 11-RELATED (PTHR24055:SF279)	non-receptor serine/threonine protein kinase(PC00220)
AT1G05700	Probable LRR receptor-like serine/threonine-protein kinase At1g05700;At1g05700;ortholog	LRR RECEPTOR-LIKE SERINE/THREONINE-PROTEIN KINASE MEE39-RELATED (PTHR27003:SF122)	
ORFX	Uncharacterized tatC-like protein ymf16;YMF16;ortholog	SEC-INDEPENDENT PROTEIN TRANSLOCASE PROTEIN TATC (PTHR30371:SF5)	
XIB	Myosin-8;XI-B;ortholog	MYOSIN-13-RELATED (PTHR13140:SF533)	G-protein modulator(PC00095);actin binding motor protein(PC00022);cell junction protein(PC00085)
PSBL	Photosystem II reaction center protein L;psbL;ortholog	PHOTOSYSTEM II REACTION CENTER PROTEIN L (PTHR33391:SF9)	
AT1G04430	Probable methyltransferase PMT8;At1g04430;ortholog	METHYLTRANSFERASE PMT8-RELATED (PTHR10108:SF1009)	methyltransferase(PC00220)
PSBT	Photosystem II reaction center protein T;psbT;ortholog	PHOTOSYSTEM II REACTION CENTER PROTEIN T (PTHR36411:SF4)	

AT1G03740	F21B7.34;At1g03740;ortholog	F21B7.34-RELATED (PTHR24056:SF258)	non-receptor serine/threonine protein kinase(PC00220);non-receptor tyrosine protein kinase(PC00137)
MCCA	Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial;MCCA;ortholog	METHYLCROTONOYL-COA CARBOXYLASE SUBUNIT ALPHA, MITOCHONDRIAL (PTHR18866:SF114)	ligase(PC00142)
AT1G05000	Probable tyrosine-protein phosphatase At1g05000;At1g05000;ortholog	TYROSINE-PROTEIN PHOSPHATASE-LIKE PROTEIN OCA2 (PTHR31126:SF24)	
PSBI	Photosystem II reaction center protein I;psbI;ortholog	PHOTOSYSTEM II REACTION CENTER PROTEIN I (PTHR35772:SF3)	
AT1G03220	Aspartyl protease-like protein;F15K9.17;ortholog	ASPARTYL PROTEASE-LIKE PROTEIN (PTHR13683:SF474)	aspartic protease(PC00121)
ATXPD	DNA repair helicase UVH6;UVH6;ortholog	TFIIH BASAL TRANSCRIPTION FACTOR COMPLEX HELICASE XPD SUBUNIT (PTHR11472:SF50)	DNA helicase(PC00171)
PSAI	Photosystem I reaction center subunit VIII;psaI;ortholog	PHOTOSYSTEM I REACTION CENTER SUBUNIT VIII (PTHR35775:SF4)	
PSBT	Photosystem II 5 kDa protein, chloroplastic;PSBT;ortholog	PHOTOSYSTEM II 5 KDA PROTEIN, CHLOROPLASTIC (PTHR34940:SF2)	
ATP9	ATP synthase subunit 9, mitochondrial;ATP9;ortholog	ATP SYNTHASE F(0) COMPLEX SUBUNIT C3, MITOCHONDRIAL (PTHR10031:SF16)	ATP synthase(PC00227)

Table 4: Functional classification of the genes expressed at both 6 hour & 24 hour time point.

	Control (C)	Treatment (P)
6 hour time point		
0 replicate	34092876	24410550
1 replicate	27470481	30688444
2 replicate	28909860	29625260
24 hour time point		
0 replicate	27474891	28370379
1 replicate	30154454	29331131
2 replicate	22128277	33418085

Table 5: Number of transcripts assembled from each file.

GLOSSARY

Sequencing depth

The total number of all sequences, reads or base pairs represented in a single sequencing experiment or series of experiments.

miRNA

microRNA (miRNA) are small noncoding RNA molecules that function in RNA silencing and post transcriptional regulation of gene expression (Ambros, 2004; Bartel, 2004)

tRNA

transfer RNA (tRNA) is an adaptor molecule composed of RNA that fictions as the link between mRNA & the amino acid sequence of proteins (Berg, Tymoczko, Stryer, & Stryer, 2002; Sharp, Schaack, Cooley, Burke, & Söll, 1985).

Contigs

A set of overlapping DNA segments resulting from the assembly of small DNA fragments that represent a consensus region of DNA (Gibson & Muse, 2009)

Trimming

A quality control process involving removal of low quality sequences or bases, adapters or contaminations to decrease errors.

Alignment

First step of RNA-seq analysis where the sequenced reads are aligned to a reference genome.

Assembly

Process where the reads aligned against a reference genome are assembled into transcripts.

FPKM

Fragments per kilobase of transcript per million mapped reads. A unit of measurement referring to read length generated from the expression of a transcript through RNA-seq. (Trapnell et al., 2010)

P-value

Is the statistical probability that the results generated from a given statistical model are not by chance when the null hypothesis is true. The lower the p-value the more statistically significant the result. (Hung, O'Neill, Bauer, & Kohne, 1997; Nuzzo, 2014).

Principle Component Analysis (PCA)

A statistical procedure that converts a set of potentially correlated variables into a smaller set of linearly uncorrelated variables called principle components, allowing for the identification of the most important variables contributing to the variability in the data through dimension reduction.

PANTHER (Protein Analysis THrough Evolutionary Relationships)

An online resource for comprehensive functional classification and data analysis. It is a comprehensive database of evolutionary and functional information about protein coding genes for 104 complete genomes (Mi et al., 2016).

Homology

Defined as the similarity between sequences based on shared ancestry

Orthologs

The similarity between sequences that is attributed to shared ancestry because of a speciation event.

References:

Aarts, M. G., Keijzer, C. J., Stiekema, W. J., & Pereira, A. (1995). Molecular characterization of the CER1 gene of arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. *The Plant Cell*, *7*(12), 2115–27. http://doi.org/10.1105/TPC.7.12.2115

Alamancos, G. P., Agirre, E., & Eyras, E. (2012). Chapter 26 Methods to Study Splicing from High-Throughput RNA Sequencing Data, *1126*(Chen 2011), 1–31. http://doi.org/10.1007/978-1-62703-980-2

Ambros, V. (2004). The functions of animal microRNAs. *Nature*, *431*(7006), 350–5. http://doi.org/10.1038/nature02871

Antibiotics, O. F. (2004). obtained . EcoR I fragments from one of the clones were inserted into the EcoR I site of pUC18 . Sequencing of the genomic DNA inserts was performed with a Dye Terminator Cycle Sequencing kit (Applied Biosystems , U . S . A .) and an ABI Prism 377 DNA Se, *57*(3), 173–179.

Arabidopsis Genome Initiative, T. A. G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, *408*(6814), 796–815. http://doi.org/10.1038/35048692

Arabidopsis thaliana %93 Overview. (n.d.). Encyclopedia of Life. Retrieved from http://eol.org/pages/583954/overview

Arabidopsis thaliana (thale cress). (n.d.). Kew Gardens. Retrieved from http://www.kew.org/science-conservation/plants-fungi/arabidopsis-thaliana-thale-cress

Arai, K., Buonamici, S., Chan, B., Corson, L., Endo, A., Gerard, B., … Keaney, G. F. (2014). Total Synthesis of 6‐Deoxypladienolide D and Assessment of Splicing Inhibitory Activity in a Mutant SF3B1 Cancer Cell Line, 5560–5563.

Ashby, M. K., Houmard, J., & Mullineaux, C. W. (2002). The ycf 27 genes from cyanobacteria and eukaryotic algae: distribution and implications for chloroplast evolution. *FEMS Microbiology Letters*, *214*(1).

Barah, P., Jayavelu, N. D., Rasmussen, S., Nielsen, H. B., Mundy, J., & Bones, A. M. (2013). Genome-scale cold stress response regulatory networks in ten Arabidopsis thaliana ecotypes. *BMC Genomics*, *14*(1), 722. http://doi.org/10.1186/1471-2164-14-

722

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, *116*(2), 281–97. http://doi.org/10.1016/S0092-8674(04)00045-5

Bayne, E. H., Portoso, M., Kagansky, A., Kos-Braun, I. C., Urano, T., Ekwall, K., … Allshire, R. C. (2008). Splicing Factors Facilitate RNAi-Directed Silencing in Fission Yeast. *Science*, *322*(5901).

Beck, E. H., Heim, R., & Hansen, J. (2004). Plant resistance to cold stress: mechanisms and environmental signals triggering frost hardening and dehardening. *Journal of Biosciences*, *29*(4), 449–459. http://doi.org/10.1007/BF02712118

Beggs, J. D. (1993). Yeast protein splicing factors involved in nuclear pre-mRNA splicing. *Molecular Biology Reports*, *18*(2), 99–103. http://doi.org/10.1007/BF00986763

Berg, J. M. (Jeremy M., Tymoczko, J. L., Stryer, L., & Stryer, L. (2002). *Biochemistry* (5th ed.). San Francisco: W.H. Freeman. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK21154/

Berr, A., McCallum, E. J., Ménard, R., Meyer, D., Fuchs, J., Dong, A., & Shen, W.-H. (2010). Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development. *The Plant Cell*, *22*(10), 3232–48. http://doi.org/10.1105/tpc.110.079962

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, *72*(1), 291–336. http://doi.org/10.1146/annurev.biochem.72.121801.161720

Boyko, A., & Kovalchuk, I. (2008). Epigenetic control of plant stress response. *Environmental and Molecular Mutagenesis*, *49*(1), 61–72. http://doi.org/10.1002/em.20347

Chandler, J. W., & Werr, W. (2015). Cytokinin-auxin crosstalk in cell type specification. *Trends in Plant Science*, *20*(5), 291–300. http://doi.org/10.1016/j.tplants.2015.02.003

Chen, W., Provart, N. J., Glazebrook, J., Katagiri, F., Chang, H., Mauch, F., … Wang, X. (2013). Expression Profile Matrix of Arabidopsis Transcription Factor Genes Suggests Their Putative Functions in Response to Environmental Stresses © American Society of Plant Biologists Expression Profile Matrix of Arabidopsis Transcription Factor Genes Suggests. *The Plant Cell*, *14*(3), 559–574.

http://doi.org/10.1105/tpc.010410

Cleveland, W. S. (1993). *Visualizing data*. Murray Hill, N.J. Summit, N.J: At & T Bell Laboratories.

Coelho, S. M., Peters, A. F., Charrier, B., Roze, D., Destombe, C., Valero, M., & Cock, J. M. (2007). Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. *Gene*, *406*(1-2), 152–70. http://doi.org/10.1016/j.gene.2007.07.025

Cs, M., E, M., Mj-f, S., & W, G. (2003). Occurrence and fate of macrolide antibiotics in wastewater treatment plants and. *In the Glatt Valley Watershed, Switzerland. Environ Sci Technol*, *37*(24), 5479–86.

Cui, X., & Churchill, G. A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology*, *4*(4), 210. http://doi.org/10.1186/gb-2003-4-4-210

Čulić, O., Eraković, V., & Parnham, M. J. (2001). Anti-inflammatory effects of macrolide antibiotics. *European Journal of Pharmacology*, *429*(1-3), 209–29. http://doi.org/10.1016/S0014-2999(01)01321-8

CummeRbund - An R package for persistent storage, analysis, and visualization of RNA-Seq from cufflinks output. (n.d.). Retrieved from http://compbio.mit.edu/cummeRbund

de Klerk, E., 't Hoen, P. A. C., Djebali, S., al., et, Neph, S., al., et, … Baranov, P. V. (2015). Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends in Genetics : TIG*, *31*(3), 128–39. http://doi.org/10.1016/j.tig.2015.01.001

de Torres-Zabala, M., Truman, W., Bennett, M. H., Lafforgue, G., Mansfield, J. W., Rodriguez Egea, P., … Grant, M. (2007). Pseudomonas syringae pv. tomato hijacks the Arabidopsis abscisic acid signalling pathway to cause disease. *The EMBO Journal*, *26*(5), 1434–1443. http://doi.org/10.1038/sj.emboj.7601575

Delhaize, E., & Randall, P. J. (1995). Characterization of a Phosphate-Accumulator Mutant of Arabidopsis thaliana. *Plant Physiology*, *107*(1), 207–213. http://doi.org/10.1104/PP.107.1.207

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.

http://doi.org/10.1093/bioinformatics/bts635

Dong, C. H., Xia, G. X., Hong, Y., Ramachandran, S., Kost, B., & Chua, N. H. (2001). ADF proteins are involved in the control of flowering and regulate F-actin organization, cell expansion, and organ growth in Arabidopsis. *The Plant Cell*, *13*(6), 1333–46. http://doi.org/10.1105/TPC.010051

Durian, G., Rahikainen, M., Alegre, S., Brosché, M., & Kangasjärvi, S. (2016). Protein Phosphatase 2A in the Regulatory Network Underlying Biotic Stress Resistance in Plants. *Frontiers in Plant Science*, *7*(June), 1–17. http://doi.org/10.3389/fpls.2016.00812

Engström, P. G., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., Rätsch, G., … Bertone, P. (2013). Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods*, *10*(12), 1185–91. http://doi.org/10.1038/nmeth.2722

Fan, J., Hill, L., Crooks, C., Doerner, P., & Lamb, C. (2009). Abscisic acid has a key role in modulating diverse plant-pathogen interactions. *Plant Physiology*, *150*(4), 1750–61. http://doi.org/10.1104/pp.109.137943

Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., … Mockler, T. C. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana, 45–58. http://doi.org/10.1101/gr.093302.109.2008

Friendly, M., & Denis, D. (2005). The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, *41*(2), 103–130. http://doi.org/10.1002/jhbs.20078

Frigge, M., Hoaglin, D. C., & Iglewicz, B. (1989). Some Implementations of the Boxplot. *The American Statistician*, *43*(1), 50. http://doi.org/10.2307/2685173

Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. Academic Press. Retrieved from https://books.google.com/books?visbn=0-12-269851-7

Gan, X., Stegle, O., Behr, J., Steffen, J. G., Drewe, P., Hildebrand, K. L., … Ra, G. (2011). Multiple reference genomes and transcriptomes for Arabidopsis thaliana, 1–5. http://doi.org/10.1038/nature10414

Garber, M., Grabherr, M. G., Guttman, M., & Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods*, *8*(6), 469–477. http://doi.org/10.1038/nmeth.1613

Genome Assembly. (n.d.). The Arabidopsis Information Resource. Retrieved from http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/agicomplete.jsp

Gibson, G., & Muse, S. V. (2009). *A primer of genome science* (3rd ed.). Sinauer Associates.

Gómez-Casati, D. F., Busi, M. V, Gonzalez-Schain, N., Mouras, A., Zabaleta, E. J., & Araya, A. (2002). A mitochondrial dysfunction induces the expression of nuclear-encoded complex I genes in engineered male sterile *Arabidopsis thaliana*. *FEBS Letters*, *532*(1-2), 70–74. http://doi.org/10.1016/S0014-5793(02)03631-1

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–52. http://doi.org/10.1038/nbt.1883

Griffith, M., Griffith, O. L., Mwenifumbo, J., Goya, R., Morrissy, a S., Morin, R. D., … Marra, M. A. (2010). Alternative expression analysis by RNA sequencing. *Nature Methods*, *7*(10), 843–7. http://doi.org/10.1038/nmeth.1503

Griffith, M., Walker, J. R., Spies, N. C., Ainscough, B. J., & Griffith, O. L. (2015). Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. *PLoS Computational Biology*, *11*(8), 1–20. http://doi.org/10.1371/journal.pcbi.1004393

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. http://doi.org/10.1038/nprot.2013.084

Haas, B. J., & Zody, M. C. (2010). news and views Advancing RNA-Seq analysis. *Nature Publishing Group*, *28*(5), 421–423. http://doi.org/10.1038/nbt0510-421

Hagen, G., & Guilfoyle, T. (2002). Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Molecular Biology*, *49*(3/4), 373–385. http://doi.org/10.1023/A:1015207114117

HanumanthaRao, B., Nair, R. M., & Nayyar, H. (2016). Salinity and High Temperature Tolerance in Mungbean [Vigna radiata (L.) Wilczek] from a Physiological Perspective. *Frontiers in Plant Science*, *7*(June), 1–20.

http://doi.org/10.3389/fpls.2016.00957

Helt, G. A., Nicol, J. W., Erwin, E., Blossom, E., Blanchard, S. G., Chervitz, S. A., … Loraine, A. E. (2009). Genoviz Software Development Kit: Java tool kit for building genomics visualization applications. *BMC Bioinformatics*, *10*(20), 266. http://doi.org/10.1186/1471-2105-10-266

Honeyman, J. N., Simon, E. P., Robine, N., Chiaroni-Clarke, R., Darcy, D. G., Lim, I. I. P., … Haussler, D. (2014). Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. *Science (New York, N.Y.)*, *343*(6174), 1010–4. http://doi.org/10.1126/science.1249484

Hung, H. M. J., O'Neill, R. T., Bauer, P., & Kohne, K. (1997). The Behavior of the P-Value When the Alternative Hypothesis is True. *Biometrics*, *53*(1), 11. http://doi.org/10.2307/2533093

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, *7*(8), 1534–50. http://doi.org/10.1038/nprot.2012.086

Jarrell, S. B. (1994). *Basic statistics*. Dubuque, Iowa: Wm. C. Brown Pub.

Jensen, M. K., Lindemose, S., de Masi, F., Reimer, J. J., Nielsen, M., Perera, V., … Skriver, K. (2013). ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene *NCED3* in *Arabidopsis thaliana*. *FEBS Open Bio*, *3*(1), 321–327. http://doi.org/10.1016/j.fob.2013.07.006

Jolliffe, I. T. (2002). *Principal Component Analysis*. New York: Springer-Verlag. http://doi.org/10.1007/b98835

Kalari, K. R., Nair, A. A., Bhavsar, J. D., O'Brien, D. R., Davila, J. I., Bockol, M. A., … Salzberg, S. (2014). MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. *BMC Bioinformatics*, *15*(1), 224. http://doi.org/10.1186/1471-2105-15-224

Kanada, R. M., Itoh, D., Nagai, M., Niijima, J., Asai, N., Mizui, Y., … Kotake, Y. (2007). Total Synthesis of the Potent Antitumor Macrolides Pladienolide B and D **, 4350–4355. http://doi.org/10.1002/anie.200604997

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., & Martinoia, E. (2011). Plant ABC Transporters. *The Arabidopsis Book*, *9*, e0153.

http://doi.org/10.1199/tab.0153

Kanoh, S., & Rubin, B. K. (2010). Mechanisms of action and clinical application of macrolides as immunomodulatory medications. *Clinical Microbiology Reviews*, *23*(3), 590–615. http://doi.org/10.1128/CMR.00078-09

Karpov, P. A., & Blume, Y. B. B. B. (2008). Bioinformatic Search For Plant Homologues Of Animal Structural Maps In The Arabidopsis Thaliana Genome (pp. 373–394). Springer Netherlands. http://doi.org/10.1007/978-1-4020-8843-8_18

Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation, *7*(12). http://doi.org/10.1038/nmeth.1528

Kong, Y., Zhou, G., Abdeen, A. A., Schafhauser, J., Richardson, B., Atmodjo, M. A., … Hahn, M. G. (2013). GALACTURONOSYLTRANSFERASE-LIKE5 is involved in the production of Arabidopsis seed coat mucilage. *Plant Physiology*, *163*(3), 1203–17. http://doi.org/10.1104/pp.113.227041

Kotake, Y., Sagane, K., Owa, T., Mimori-kiyosue, Y., Shimizu, H., Uesugi, M., … Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide, *3*(9), 570–575. http://doi.org/10.1038/nchembio.2007.16

Kourmpetis, Y. A. I., van Dijk, A. D. J., van Ham, R. C. H. J., & ter Braak, C. J. F. (2011). Genome-wide computational function prediction of Arabidopsis proteins by integration of multiple data sources. *Plant Physiology*, *155*(1), 271–81. http://doi.org/10.1104/pp.110.162164

Kranner, I., Minibayeva, F. V., Beckett, R. P., & Seal, C. E. (2010). What is stress? Concepts, definitions and applications in seed science. *New Phytologist*, *188*(3), 655–673. http://doi.org/10.1111/j.1469-8137.2010.03461.x

Kranner, I., Minibayeva, F. V, Beckett, R. P., Seal, C. E., The, S., Phytologist, N., … Seal, C. E. (2016). What is stress ? Concepts , definitions and applications in seed science Published by : Wiley on behalf of the New Phytologist Trust Linked references are available on JSTOR for this article : New mmi ^ H Tansley review What is stress ? Concepts , definit, *188*(3), 655–673.

Kreps, J. A., Wu, Y., Chang, H., Zhu, T., Wang, X., Harper, J. F., … Diego, S. (2002). Transcriptome Changes for Arabidopsis in Response to Salt , Osmotic , and Cold

Stress 1 [w], *130*(December), 2129–2141. http://doi.org/10.1104/pp.008532.with

Landau, A. M., Lokstein, H., Scheller, H. V., Lainez, V., Maldonado, S., & Prina, A. R. (2009). A cytoplasmically inherited barley mutant is defective in photosystem I assembly due to a temperature-sensitive defect in ycf3 splicing. *Plant Physiology*, *151*(4), 1802–11. http://doi.org/10.1104/pp.109.147843

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. http://doi.org/10.1186/gb-2009-10-3-r25

Lewis, B. P., Green, R. E., & Brenner, S. E. (2002). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans.

Leyser, H. M. O., Pickett, F. B., Dharmasiri, S., & Estelle, M. (1996). Mutations in the AXR3 gene of Arabidopsis result in altered auxin response including ectopic expression from the SAUR-AC1 promoter. *The Plant Journal*, *10*(3), 403–413. http://doi.org/10.1046/j.1365-313x.1996.10030403.x

Li, S., Tighe, S. W., Nicolet, C. M., Grove, D., Levy, S., Viale, A., … Kim, D. (2014). Multi-platform and cross-methodological reproducibility of transcriptome profiling by RNA-seq in the ABRF Next- Generation Sequencing Study. *Nat Biotechnol.*, *32*(9), 915–925. http://doi.org/10.1038/nbt.2972.Multi-platform

Li, W. (2012). Volcano plots in analyzing differential expressions with mRNA microarrays. *Journal of Bioinformatics and Computational Biology*, *10*(6), 1231003. http://doi.org/10.1142/S0219720012310038

Liu, F., Ying, G.-G., Tao, R., Zhao, J.-L., Yang, J.-F., & Zhao, L.-F. (2009). Effects of six selected antibiotics on plant growth and soil microbial and enzymatic activities. *Environmental Pollution (Barking, Essex : 1987)*, *157*(5), 1636–42. http://doi.org/10.1016/j.envpol.2008.12.021

Liu, H., Tian, X. I. N., Li, Y., Wu, C., & Zheng, C. (2008). Microarray-based analysis of stress-regulated microRNAs in Arabidopsis thaliana, 836–843. http://doi.org/10.1261/rna.895308.)

López-Boado, Y. S., & Rubin, B. K. (2008). Macrolides as immunomodulatory medications for the therapy of chronic lung diseases. *Current Opinion in Pharmacology*, *8*(3), 286–291. http://doi.org/10.1016/j.coph.2008.01.010

López-Bucio, J., Campos-Cuevas, J. C., Hernández-Calderón, E., Velásquez-Becerra, C., Farías-Rodríguez, R., Macías-Rodríguez, L. I., & Valencia-Cantero, E. (2007). Bacillus megaterium rhizobacteria promote growth and alter root-system architecture through an auxin- and ethylene-independent signaling mechanism in Arabidopsis thaliana. *Molecular Plant-Microbe Interactions : MPMI*, *20*(2), 207–17. http://doi.org/10.1094/MPMI-20-2-0207

Lu, B., Zeng, Z., & Shi, T. (2013). Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. *Science China. Life Sciences*, *56*(2), 143–55. http://doi.org/10.1007/s11427-013-4442-z

Maciver, S. K., Hussey, P. J., Bamburg, J., Pollard, T., Blanchoin, L., Mullins, R., … Kiehart, D. (2002). The ADF/cofilin family: actin-remodeling proteins. *Genome Biology*, *3*(5), reviews3007.1. http://doi.org/10.1186/gb-2002-3-5-reviews3007

Mahé, L., Combes, M.-C., & Lashermes, P. (2007). Comparison between a coffee single copy chromosomal region and Arabidopsis duplicated counterparts evidenced high level synteny between the coffee genome and the ancestral Arabidopsis genome. *Plant Molecular Biology*, *64*(6), 699–711. http://doi.org/10.1007/s11103-007-9191-6

Maher, C. a, Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., … Chinnaiyan, A. M. (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature*, *458*(7234), 97–101. http://doi.org/10.1038/nature07638

Maniatis, T., & Reed, R. (2002). Expression Machines. *Nature*, *416*(6880), 499–506. http://doi.org/10.1038/416499a

Marin, E., Divol, F., Bechtold, N., Vavasseur, A., Nussaume, L., & Forestier, C. (2006). Molecular characterization of three Arabidopsis soluble ABC proteins which expression is induced by sugars. *Plant Science*, *171*(1), 84–90. http://doi.org/10.1016/j.plantsci.2006.02.014

Marquez, Y., Brown, J. W. S., Simpson, C., Barta, A., & Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis, 1184–1195. http://doi.org/10.1101/gr.134106.111.Freely

Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A., & Kalyna, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in

proteome plasticity. *Genome Research*, *25*(7), 995–1007. http://doi.org/10.1101/gr.186585.114

Martin, J. a., & Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, *12*(10), 671–682. http://doi.org/10.1038/nrg3068

Matlin, A. J., Clark, F., & Smith, C. W. J. (2005). Understanding alternative splicing: towards a cellular code. *Nature Reviews. Molecular Cell Biology*, *6*(5), 386–98. http://doi.org/10.1038/nrm1645

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician*, *32*(1), 12. http://doi.org/10.2307/2683468

Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., & Koornneef, M. (1998). Arabidopsis thaliana: a model plant for genome analysis. *Science (New York, N.Y.)*, *282*(5389), 662, 679–82. http://doi.org/10.1126/science.282.5389.662

Menges, M., De Jager, S. M., Gruissem, W., & Murray, J. A. H. (2005). Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *The Plant Journal*, *41*(4), 546–566. http://doi.org/10.1111/j.1365-313X.2004.02319.x

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, *8*(8), 1551–1566. http://doi.org/10.1038/nprot.2013.092

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Research*, *44*(D1), D336–D342. http://doi.org/10.1093/nar/gkv1194

Mi, H., & Thomas, P. (2009). PANTHER Pathway: An Ontology-Based Pathway Database Coupled with Data Analysis Tools (pp. 123–140). http://doi.org/10.1007/978-1-60761-175-2_7

Mittler, R. (2002). Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science*, *7*(9), 405–410. http://doi.org/10.1016/S1360-1385(02)02312-9

Monaghan, J., Xu, F., Gao, M., Zhao, Q., Palma, K., Long, C., … Li, X. (2009). Two Prp19-Like U-Box Proteins in the MOS4-Associated Complex Play Redundant Roles in Plant Innate Immunity. *PLoS Pathogens*, *5*(7), e1000526.

http://doi.org/10.1371/journal.ppat.1000526

Moore, C. S., Cook-Johnson, R. J., Rudhe, C., Whelan, J., Day, D. A., Wiskich, J. T., & Soole, K. L. (2003). Identification of AtNDI1, an internal non-phosphorylating NAD(P)H dehydrogenase in Arabidopsis mitochondria. *Plant Physiology*, *133*(4), 1968–78. http://doi.org/10.1104/pp.103.029363

Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., … Marra, M. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques*, *45*(1), 81–94. http://doi.org/10.2144/000112900

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq, *5*(7), 1–8. http://doi.org/10.1038/NMETH.1226

Nagalakshmi, U., Waern, K., & Snyder, M. (2010). RNA-Seq : A Method for Comprehensive Transcriptome Analysis, (January), 1–13. http://doi.org/10.1002/0471142727.mb0411s89

Nanao, M. H., Vinos-Poyo, T., Brunoud, G., Thévenon, E., Mazzoleni, M., Mast, D., … Metoz, F. (2014). Structural basis for oligomerization of auxin transcriptional regulators. *Nature Communications*, *5*, 265–285. http://doi.org/10.1038/ncomms4617

Network component analysis: Reconstruction of regulatory signals in biological systems. (n.d.). Retrieved from http://www.pnas.org/content/100/26/15522.full.pdf

Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*, *506*(7487), 150–152. http://doi.org/10.1038/506150a

O'Neil, S. T., & Emrich, S. J. (2013). Assessing De Novo transcriptome assembly metrics for consistency and utility. *BMC Genomics*, *14*(1), 465. http://doi.org/10.1186/1471-2164-14-465

Osakabe, Y., Osakabe, K., Shinozaki, K., & Tran, L.-S. P. (2014). Response of plants to water stress. *Frontiers in Plant Science*, *5*(March), 86. http://doi.org/10.3389/fpls.2014.00086

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature Reviews. Genetics*, *12*(2), 87–98. http://doi.org/10.1038/nrg2934

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–5. http://doi.org/10.1038/ng.259

Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell*, *108*(4), 501–512. http://doi.org/10.1016/S0092-8674(02)00617-7

Pugnaire, F. I., & Luque, M. T. (2001). Changes in plant interactions along a gradient of environmental stress. *Oikos*, *93*, 42–49. http://doi.org/10.1034/j.1600-0706.2001.930104.x

Racine, J. S., Hall, P., & Li, Q. (2004). Cross-Validation and the Estimation of Conditional Probability Densities. *Journal of The American Statistical Association*, *99*(468), 1015–1026. Retrieved from http://econpapers.repec.org/article/besjnlasa/v_3a99_3ay_3a2004_3ap_3a1015-1026.htm

Ralet, M.-C., Crépeau, M.-J., Vigouroux, J., Tran, J., Berger, A., Sallé, C., … North, H. M. (2016). Xylans Provide the Structural Driving Force for Mucilage Adhesion to the Arabidopsis Seed Coat. *Plant Physiology*, *171*(1), 165–78. http://doi.org/10.1104/pp.16.00211

Reddy, A. S. N. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annual Review of Plant Biology*, *58*, 267–94. http://doi.org/10.1146/annurev.arplant.58.032806.103754

Ren, S., Mandadi, K. K., Boedeker, A. L., Rathore, K. S., & McKnight, T. D. (2007). Regulation of telomerase in Arabidopsis by BT2, an apparent target of TELOMERASE ACTIVATOR1. *The Plant Cell*, *19*(1), 23–31. http://doi.org/10.1105/tpc.106.044321

Rensink, W. A., & Buell, C. R. (2004). Arabidopsis to rice. Applying knowledge from a weed to enhance our understanding of a crop species. *Plant Physiology*, *135*(2), 622–9. http://doi.org/10.1104/pp.104.040170

Richardson, D. N., Rogers, M. F., Labadorf, A., Ben-Hur, A., Guo, H., Paterson, A. H., & Reddy, A. S. N. (2011). Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: Insights into sub-family classification and extent of alternative splicing. *PLoS ONE*, *6*(9). http://doi.org/10.1371/journal.pone.0024542

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., Mittler, R., … R. (2004). When Defense Pathways Collide. The Response of Arabidopsis to a Combination of Drought and Heat Stress. *Plant Physiology*, *134*(April), 1683–1696. http://doi.org/10.1104/pp.103.033431.1

Robert, H. S., Quint, A., Brand, D., Vivian-Smith, A., & Offringa, R. (2009). BTB and TAZ domain scaffold proteins perform a crucial function in Arabidopsis development. *The Plant Journal*, *58*(1), 109–121. http://doi.org/10.1111/j.1365-313X.2008.03764.x

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–40. http://doi.org/10.1093/bioinformatics/btp616

Rouse, D., Mackay, P., Stirnberg, P., Estelle, M., & Leyser, O. (1998). Changes in Auxin Response from Mutations in an AUX/IAA Gene. *Science*, *279*(5355).

Rühl, C., Stauffer, E., Kahles, A., Wagner, G., Drechsel, G., Rätsch, G., & Wachter, A. (2012). Polypyrimidine tract binding protein homologs from Arabidopsis are key regulators of alternative splicing with implications in fundamental developmental processes. *The Plant Cell*, *24*(11), 4360–75. http://doi.org/10.1105/tpc.112.103622

Saez-Aguayo, S., Ralet, M.-C., Berger, A., Botran, L., Ropartz, D., Marion-Poll, A., & North, H. M. (2013). PECTIN METHYLESTERASE INHIBITOR6 promotes Arabidopsis mucilage release by limiting methylesterification of homogalacturonan in seed coat epidermal cells. *The Plant Cell*, *25*(1), 308–23. http://doi.org/10.1105/tpc.112.106575

Sammeth, M., Foissac, S., & Guigó, R. (2008). A general definition and nomenclature for alternative splicing events. *PLoS Computational Biology*, *4*(8), e1000147. http://doi.org/10.1371/journal.pcbi.1000147

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., & Tabata, S. (1999). Complete structure of the chloroplast genome of Arabidopsis thaliana. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, *6*(5), 283–90. http://doi.org/10.1093/DNARES/6.5.283

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of

gene expression patterns with a complementary DNA microarray. *Science (New York, N.Y.)*, *270*(5235), 467–70. http://doi.org/10.1126/science.270.5235.467

Sharp, S. J., Schaack, J., Cooley, L., Burke, D. J., & Söll, D. (1985). Structure and transcription of eukaryotic tRNA genes. *CRC Critical Reviews in Biochemistry*, *19*(2), 107–44. http://doi.org/10.3109/10409238509082541

Shinkai, M., Henke, M. O., & Rubin, B. K. (2008). Macrolide antibiotics as immunomodulatory medications: Proposed mechanisms of action. *Pharmacology and Therapeutics*, *117*(3), 393–405. http://doi.org/10.1016/j.pharmthera.2007.11.001

Singh, D., Chan, J. M., Zoppoli, P., Niola, F., Sullivan, R., Castano, A., … Minna, J. D. (2012). Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science (New York, N.Y.)*, *337*(6099), 1231–5. http://doi.org/10.1126/science.1220834

Staiger, D. (2015). Shaping the Arabidopsis Transcriptome through, *2015*.

Staiger, D., & Simpson, G. G. (2015). Enter exitrons. *Genome Biology*, *16*(1), 136. http://doi.org/10.1186/s13059-015-0704-3

Substances, N., & Antitumor, V. (2004). We have discovered seven novel 12-membered macrolides , pladienolides A to G , from Streptomyces platensis Mer-11107 , with pladienolide B the most potently inhibiting hypoxia induced-VEGF expression and proliferation of the U251 cancer cell line . A grow, *57*(3).

Substances, N., Properties, P., Okuda, A., & Kawamura, N. (2004). In the course of our screening using fermented broth from soil microorganisms , novel metabolites (pladienolides), possessing inhibitory activity against vascular endothelial growth factor (VEGF) expression and cancer cell proliferation , were isolate, *57*(3).

Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., … Yaspo, M.-L. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science (New York, N.Y.)*, *321*(5891), 956–60. http://doi.org/10.1126/science.1160342

Sun, W., Bernard, C., Van De Cotte, B., Van Montagu, M., & Verbruggen, N. (2001). At-HSP17.6A, encoding a small heat-shock protein in Arabidopsis, can enhance osmotolerance upon overexpression. *The Plant Journal*, *27*(5), 407–415. http://doi.org/10.1046/j.1365-313X.2001.01107.x

Sunkar, R., Ã, D. B., & Kirch, H. (2003). Overexpression of a stress-inducible aldehyde

dehydrogenase gene from Arabidopsis thaliana in transgenic plants improves stress tolerance, 452–464. http://doi.org/10.1046/j.1365-313X.2003.01819.x

Suorsa, M., Regel, R. E., Paakkarinen, V., Battchikova, N., Herrmann, R. G., & Aro, E.-M. (2003). Protein assembly of photosystem II and accumulation of subcomplexes in the absence of low molecular mass subunits PsbL and PsbJ. *European Journal of Biochemistry*, *271*(1), 96–107. http://doi.org/10.1046/j.1432-1033.2003.03906.x

TAIR - About Arabidopsis. (n.d.). *Www.arabidopsis.org*. Retrieved from https://www.arabidopsis.org/portals/education/aboutarabidopsis.jsp

Tang, H., & Thomas, P. D. (2016). PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*, *32*(14), 2230–2232. http://doi.org/10.1093/bioinformatics/btw222

Tenson, T., Lovmar, M., & Ehrenberg, M. (2003). The mechanism of action of macrolides, lincosamides and streptogramin B reveals the nascent peptide exit path in the ribosome. *Journal of Molecular Biology*, *330*(5), 1005–14. http://doi.org/10.1016/S0022-2836(03)00662-4

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. http://doi.org/10.1093/bib/bbs017

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. http://doi.org/10.1038/nbt.2450

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009a). TopHat : discovering splice junctions with RNA-Seq, *25*(9), 1105–1111. http://doi.org/10.1093/bioinformatics/btp120

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009b). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, *25*(9), 1105–11. http://doi.org/10.1093/bioinformatics/btp120

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–78. http://doi.org/10.1038/nprot.2012.016

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., …

Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–5. http://doi.org/10.1038/nbt.1621

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

Van Keuren-Jensen, K., Keats, J. J., & Craig, D. W. (2014). Bringing RNA-seq closer to the clinic. *Nature Biotechnology*, *32*(9), 884–885. http://doi.org/10.1038/nbt.3017

Vidal, E. A., Moyano, T. C., Riveras, E., Contreras-López, O., & Gutiérrez, R. A. (2013). Systems approaches map regulatory networks downstream of the auxin receptor AFB3 in the nitrate response of Arabidopsis thaliana roots. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(31), 12840–5. http://doi.org/10.1073/pnas.1310937110

Wang, B., & Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants, *103*(18), 7175–7180.

Wang, C., Tian, Q., Hou, Z., Mucha, M., Aukerman, M., & Olsen, O.-A. (2007). The Arabidopsis thaliana AT PRP39-1 gene, encoding a tetratricopeptide repeat protein with similarity to the yeast pre-mRNA processing protein PRP39, affects flowering time. *Plant Cell Reports*, *26*(8), 1357–1366. http://doi.org/10.1007/s00299-007-0336-5

Wang, W., Vinocur, B., & Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta*, *218*(1), 1–14. http://doi.org/10.1007/s00425-003-1105-5

Wang, X., Basnayake, B. M. V. S., Zhang, H., Li, G., Li, W., Virk, N., … Song, F. (2009). The *Arabidopsis* ATAF1, a NAC Transcription Factor, Is a Negative Regulator of Defense Responses Against Necrotrophic Fungal and Bacterial Pathogens. *Molecular Plant-Microbe Interactions*, *22*(10), 1227–1238. http://doi.org/10.1094/MPMI-22-10-1227

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., … Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, *453*(7199), 1239–1243.

http://doi.org/10.1038/nature07002

Will, C. L., & Luhrmann, R. (2011). Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, *3*(7), a003707–a003707. http://doi.org/10.1101/cshperspect.a003707

Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., Iwata, M. (2011). Biological validation that SF3b is a target of the antitumor macrolide pladienolide, *278*, 4870–4880. http://doi.org/10.1111/j.1742-4658.2011.08387.x

Zalewska-Kaszubska, J., & Górska, D. (2001). Anti-inflammatory capabilities of macrolides. *Pharmacological Research : The Official Journal of the Italian Pharmacological Society*, *44*(6), 451–4. http://doi.org/10.1006/phrs.2001.0884