

American University in Cairo

## AUC Knowledge Fountain

---

Theses and Dissertations

Student Research

---

6-1-2016

### Genome-wide transcriptomic analysis of alternative splicing modulation in *Arabidopsis thaliana*

Moustafa Ahmed Abohawya

Follow this and additional works at: <https://fount.aucegypt.edu/etds>

---

#### Recommended Citation

##### APA Citation

Abohawya, M. (2016). *Genome-wide transcriptomic analysis of alternative splicing modulation in Arabidopsis thaliana* [Master's Thesis, the American University in Cairo]. AUC Knowledge Fountain. <https://fount.aucegypt.edu/etds/630>

##### MLA Citation

Abohawya, Moustafa Ahmed. *Genome-wide transcriptomic analysis of alternative splicing modulation in Arabidopsis thaliana*. 2016. American University in Cairo, Master's Thesis. AUC Knowledge Fountain. <https://fount.aucegypt.edu/etds/630>

This Master's Thesis is brought to you for free and open access by the Student Research at AUC Knowledge Fountain. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AUC Knowledge Fountain. For more information, please contact [thesisadmin@aucegypt.edu](mailto:thesisadmin@aucegypt.edu).



THE AMERICAN UNIVERSITY IN CAIRO  
الجامعة الأمريكية بالقاهرة

School of Sciences and Engineering

Genome-wide transcriptomic analysis of  
alternative splicing modulation in  
*Arabidopsis thaliana*

A Thesis Submitted to

Biotechnology Graduate Program

in partial fulfillment of the requirements for  
the degree in Master of Science

by

Moustafa Ahmed Abohawya

Bachelor of Pharmaceutical sciences, Tanta University 2012

Under the supervision of:

Main Adviser

Associate. Prof. Ahmed Moustafa

Biology Department

The American University in Cairo

Co-adviser

Assistant Prof. Magdy Mahfouz

Biological and Environmental

Sciences & Engineering Division

King Abdullah University of

Sciences and Technology

**May /2016**

## ABSTRACT

Eukaryotic genes architecture is characterized by incorporation of non-coding sequences interrupting the coding sequences. Through splicing, excision of those non-coding sequences, known as introns, is taking place co-transcriptionally, i.e. synchronically with the synthesis of precursor messenger m-RNA from the template DNA, in a precisely-regulated manner. Alternative splicing is a phenomena where not all splice sets are used in each transcript processed. Thus, this process enhances the proteome and transcriptome complexity through generating different transcripts, with different functionality or stability, from a single gene. Aberrant splicing have been associated with a lot of disease including cancer as well as Parkinsonism. Chemical compounds or drugs that target splicing regulation have also been developed and characterized. Indole derivatives have proven before their potency in targeting serine-arginine rich splicing regulators of alternative splicing. In this study, alternative splicing in *Arabidopsis* have been targeted through three different Indole derivatives, 6-Methylindole, 2, 5-Dimethylindole and 5-Bromoindole-2-Carboxylic Acid. Full transcriptomic analysis of the RNA-seq data generated from treated samples, at both 24 and 72 hrs time points, was conducted to reveal the alternative splicing modulation and gene expression patterns changes under those splicing modulators treatment. Interestingly, observation of splicing pattern alteration under the influence of Indole derivatives used in this study indicate that most treatments, 4 out of 6, and specifically after certain onset of time 72 hrs, showed significant enhancement of AS, mostly through intron retention splicing events. Photorespiration related genes were altered frequently, on AS and/or expression level, under different Indole derivatives treatments and time points. More interestingly, RNA binding proteins and splicing factors were differentially spliced, a finding that is consistent with the prevalence of auto-regulatory nature of those genes. Furthermore, various potential functional outcomes may originate from differential splicing exhibited under Indole treatments. In hydrolase-like protein (AT5G17670), an alternative 5' splice site skipped a pre-mature termination codon. In SR34a SR splicing factor, exon skipping event at the 5' untranslated region might help the transcript escape regulatory miRNAs. Overall, this study shall introduce a transcriptomic snapshot of Genome-wide gene expression and alternative splicing after targeting serine-arginine splicing factors by Indole derivatives.

# TABLE OF CONTENTS

Abstract.....	II
Table of contents.....	III
List of figures.....	V
List of abbreviations.....	VIII
Acknowledgements.....	IX
Chapter One: Introduction and study objectives.....	1
1.1. Precursor m-RNA splicing in eukaryotes.....	1
1.2. Mechanism of splicing.....	2
1.3. Mechanism of Alternative splicing.....	4
1.4. Regulation of Alternative splicing.....	5
1.5. Alternative splicing role in plant adaptation to environmental stresses .....	7
1.6. Signals coupling to alternative splicing.....	9
1.7. Drugs modulating alternative splicing .....	10
1.8. Selective alteration of alternative splicing through targeting serine arginine rich splicing factors by Indole derivatives.....	11
1.9. Study design and objectives.....	13
Chapter Two: Materials and Methods.....	17
2.1. Samples preparations and treatments.....	17
2.2 RNA extraction, library preparation and sequencing.....	17
2.3 RNA-seq reads quality control and quality assessment.....	17
2.4. RNA-seq reads alignment and junction prediction.....	18

2.5 Differential gene expression analysis .....	19
2.6 Alternative splice events annotation from confident splice junctions .....	19
2.7 Differential alternative splicing analysis .....	21
2.8 Visualization of differentially expressed alternative splice events .....	22
<b>Chapter Three: Results and discussion.....</b>	<b>24</b>
3.1 Global alterations of alternative splicing.....	24
3.2 Differentially spliced genes associated with Indole derivatives treatment .....	26
3.3 Differential gene expression analysis .....	28
3.4 Functional consequences of specific alternative splicing events promoted by different Indole derivative .....	29
<b>Chapter Four: Conclusion and future perspectives.....</b>	<b>61</b>
<b>References .....</b>	<b>63</b>

## LIST OF FIGURES

Fig 1.1 Mechanism of Splicing and splicesomal assembly on splice site...	14
Fig 1.2 <i>Cis</i> acting factors and types of alternative splice events.....	15
Fig 1.3 Alternative splicing regulation through trans-acting splicing factors.....	16
Fig2.1 Density distribution bar plot of known and annotated splice junctions in replicate 1 of control sample at 24 hours' time point.....	22
Fig2.2 Flow chart of the study design.....	23
Fig3.1 Total AS events distribution.....	31
Fig3.2 Annotated and Novel AS events distribution at 24 hrs time point...	32
Fig3.3 Annotated and Novel AS events distribution at 72 hrs time point...	33
Fig3.4 Different AS events distribution.....	34
Fig3.5 Distribution of statistically significant differentially expressed splicing events.....	35
Fig3.6 Distribution of statistically significant differentially spliced genes...	36
Fig 3.7 Genes distribution according to event splice type affecting it.....	37
Fig 3.8 Functional categorization of differentially spliced genes between control and 6-Metyindole treated samples at 24 hrs time point.....	38
Fig 3.9 Functional categorization of differentially spliced genes between control and 6-Metyindole treated samples at 72 hrs time point.....	39
Fig 3.10 Functional categorization of differentially spliced genes between control and 2, 5-Dimethylindolelyindole treated samples at 24 hrs time point.....	40

Fig 3.11 Functional categorization of differentially spliced genes between control and 2, 5-Dimethylindolelyindole treated samples at 72 hrs time point.....41

Fig 3.12 Functional categorization of differentially spliced genes between control and 5-Bromoindole-2-Carboxylic Acid treated samples at 24 hrs time point.....42

Fig 3.13 Functional categorization of differentially spliced genes between control and 5-Bromoindole-2-Carboxylic Acid treated samples at 72 hrs time point.....43

Fig 3.14 Volcano plots of expression alterations under .....44

Fig 3.15 Heatmap of top differentially expressed genes between control and 6-Methylindole treated samples at 24 hrs time point .....45

Fig 3.16 Heatmap of top differentially expressed genes between control and 2, 5-Dimethylindole treated samples at 24 hrs time point.....46

Fig 3.17 Heatmap of top differentially expressed genes between control and 5-Bromoindole-2-Carboxylic Acid treated samples at 24 hrs time point...47

Fig 3.18. Heatmap of top differentially expressed genes between control and 6-Methylindole treated samples at 72 hrs time point.....48

Fig 3.19. Heatmap of top differentially expressed genes between control and 2, 5-Dimethylindole treated samples at 72 hrs time point.....49

Fig 3.20 Heatmap of top differentially expressed genes between control and 5-Bromoindole-2-Carboxylic Acid treated samples at 72 hrs time point.....50

Fig 3.21 Differential usage of alternative 5' splice site that skips a premature termination codon between control and 2, 5-Dimethylindole of hydrolase-like protein (AT5G17670).....51

Fig 3.22 Differential exon skipping event in control vs. both 6-MethlyIndole and 2, 5-Dimethylindole treated samples at 24 hrs of SR34A serine arginine rich protein splicing factor.....52

Fig 3.23 Differential Intron retention event between control and 2,5-Dimethylindole at 72 hrs time point of CPK28 calcium dependent protein kina.....53



## LIST OF ABBREVIATIONS

AS	Alternative splicing
CS	Constitutive splicing
Pre-m-RNA	Precursor messenger RNA
At	<i>Arabidopsis Thaliana</i>
SRE	Splicing regulatory elements
snRNPs	small nuclear Ribonucleoproteins
SR	serine arginine rich proteins
hnRNPs	heterogeneous nucleoriboncleoproteins
PTC	pre-mature termination codon
NMD	non-sense mediated decay
GRP	Glycine rich proteins
PTB	Polyprimindine tract binding proteins.
HsfA	Heat shock factor A
SMN	Survival Motor Neuron
SMA	Spinal Muscle Atrophy
ASO	Antisense Oligonucleotides
ICV	Intracerebro-ventricular Injection
FDA	Food and Drug Administration
SFS2/ASF	splicing factor 2
ESE	exon splicing enhancer
IDC	Indole derivative compound
JuncBASE	Junction based analysis of alternative splicing
IGV	Integrative genomic viewer

# ACKNOWLEDGEMENTS

I'd like to acknowledge Dr.Ahmed Moustafa, associate professor of genomics and bioinformatics, biology department, the American University in Cairo for his infinite generous support, help and encouragement. He wasn't only a thesis adviser or mentor, but also a role model scientist to follow. I'd be utterly grateful for him, forever.

I'd like to acknowledge Dr.Magdy Mahfouz and Dr. Haroon Butt from King Abdallah University for science and technology for conducting this study (most specifically; samples treatment, RNA extraction, Library preparation and RNA-sequencing).

I'd like to thank Angela Brooks from University of California Berkeley, who developed JuncBASE platform, for her generous help to utilize it to analyze alternative splicing in *Arabidopsis*. I'd like to thank Dr.Xiong and Dr.Fing, from King Abdallah University for science and technology for their pioneering approach to use juncBASE in *Arabidopsis* and helping me following them in the context of my study.

I'd further acknowledge all my old and current lab mates (Mariam, Yasmeen, Hazem, Diana, Aya, Dina) and all AUC biotech grad program fellows and friends for making my time in the program more enjoyable. I'd like to acknowledge Dr.Reem El-Olaby, for being that supportive for me to join the program. Also I'd acknowledge Eman Elzeniny for allowing me to adapt her self-designed powerpoint theme in my thesis defense presentation.

I'd like to generously thank Mr.Moataz Alfi, Alfi foundation and University fellowship to support my studies in AUC for three consecutive years. I'd never be able to make it without their support and help.

I'd also to acknowledge all fellows from bioinformatics and programming forums, seqanswers, biostars and stackoverflow, for being that supportive for beginners to learn how to use bioinformatics tool. As a research who got a minimal background in programming, those forums provided me with a priceless information and knowledge.

Finally, I'd like to acknowledge my family, friends and parents for their faith in me throughout my educational and research journey. I'll always owe you.

# Chapter One: Introduction and study objectives

## *1.1. Precursor m-RNA splicing in eukaryotes:*

Once transcription starts inside the nucleus, multiple processes will cooperate to harmonize the flow of expressed information encoded in the genome to guarantee its accordance with the desired cellular functions. Major processes include messenger RNA (m-RNA) 5' end capping, splicing and 3' end polyadenylation (addition of poly (A) tail at the 3' end of the m-RNA. Further, m-RNA bases covalent modification, editing, degradation, localization, export to translation site in cytoplasm and translation initiation control are contributing to the processes of m-RNA expression tuning. [1] Together, those processes are known as post-transcriptional modifications. The recognition of the post-transcriptional modifications as critical key players in regulation almost all physiological process in plants is remarkably increased. [2-8]

The incorporation of non-coding sequences interrupting the coding sequences is a unique characteristic of eukaryotic genes architecture and a fundamental difference from prokaryotic genes.[9-11] In splicing, excision of those non-coding sequences, known as introns, is taking place co-transcriptionally, i.e. synchronically with the synthesis of precursor messenger m-RNA (pre m-RNA) from the template DNA, in a precisely-regulated manner.[11] A precursor m-RNA is processed into mature m-RNA by removal of introns and ligation of the remaining flanking coding sequences, known as exons, through two trans-esterification reactions into mature m-RNA molecule. This molecule then becomes ready to be transported from the transcription and processing site, nucleus, to the translation site, the cytoplasm, where ribosomes will translate it into a functional entity, a protein.

Precursor m-RNA splicing is performed by one of the largest complexes of the eukaryotic cell, the spliceosome. Splicing machinery of the spliceosome is a multifunctional tool that is responsible and capable of selection of splice sites (i.e. introns and exons sequences and junction's definition) as well as exon-exon ligation and introns removal. Based on whether a single set or multiple sets of splice sites can be identified for a certain gene, any gene can undergo either a constitutive splicing (CS) or an alternative splicing (AS). In constitutive splicing, a single set of splice sites is utilized by the splicing machinery to produce a single mature m-RNA and a single protein from a single gene.[12,

13] On the other hand, in alternative splicing, splicing machinery has the potential to process the pre m-RNA in different ways by utilizing different combinations of splice sites, and as a result, generate different sets of mature m-RNAs, called splice variants or transcripts. Those different transcripts, consequently, are then translated and utilized to generate different proteins, which originated basically from the same single gene.[13, 14] Those different proteins generated from the same gene through AS can exhibit altered or modified function(s) (activity/interaction/binding properties), subcellular localization and stability.[15] Accordingly, alternative splicing is considered a major mechanism through which eukaryotes can enhance the coding potential of their genomes, increase transcriptome plasticity and proteome diversity. More interestingly, AS is believed to play a role in modulation of genes expression through modifying the relative abundances of functional and non-functional splice variants by non-sense mediated decay (NMD) along with controlling their recruitment to ribosomes and translation efficiency.[16-18] Thus, AS shall be regarded as a critical mechanism of post-transcriptional regulation of m-RNA.

Interestingly, AS was found to be extensively prevalent in intron-containing genes in human, as up to 95% of those genes shows an alternative splicing pattern of their pre m-RNA.[19, 20] In plants, almost over 40% of intron-containing genes undergo alternative splicing.[21, 22] Recently, full transcriptome coverage techniques and de novo transcriptome assembly software have estimated the percentage of alternatively spliced genes to be 61% of intro-containing genes in *Arabidopsis Thaliana*. [23, 24] However, this prevalence of alternative splicing may increase through the continuous discovery of novel splice variants recruited in different developmental stages, organs, cell types or states of response to abiotic and biotic stress. [24-30]

### *1.2. Mechanism of splicing*

In eukaryotes, short consensus sequences lie within the introns (mostly at the exon/intron borders) are responsible for guiding the recruitment of different components of splicing machinery to the splice site. (See Fig 1.1) At the beginning of an intron, 5' splice site is almost 9 nucleotide long with a consensus start GU dinucleotide. On the other end, 3' splice site of an intron has a consensus AG dinucleotide. Further, a pyrimidine rich region, called the polypyrimidine tract, is located upstream of the 3' splice site. Upstream from the polypyrimidine tract, an A sequence, called a branch point, is

recognized by spliceosome during the first step of splicing. Collectively, those sequences, known as splicing signals or splicing sites, are essential components of pre-mRNA for substrate recognition and catalysis by the core splicing machinery.[12] Other group of RNA motifs, known as splicing regulatory elements (SRE), are divided into two main types; splicing enhancers and silencers. Interestingly, they can lie within both introns and exons, thus result in complex context-dependent splicing regulation. (reviewed in [31]).

Structurally, the spliceosome, one of the largest macromolecular complexes of a eukaryotic cell, is composed of both non coding RNAs as well as a plethora of protein splicing factors. The assembly of different splicesomal components at splicing sites is precisely ordered to functionally regulate and accomplish splicing of the substrate pre m-RNA.[32-35]

The key structure of a spliceosome contains five small nuclear ribonucleoprotein complexes (snRNPs), designated U1,U2,U4,U5,U6. Each complex consists of a small nuclear RNA as well as definite set of protein factors. Those small nuclear RNAs are basically short, Uridine rich, non-coding, non-polyadenylated transcripts which also referred to as U1, U2, U4, U5, U6 snRNAs based on the Ribonucleoprotein complex they contribute to. The core proteins utilized in the structure of U1, U2, U4, U5 snRNPs belong mainly to the Sm protein family, for Smith, an autoimmune disease where patient develop antibodies against those proteins. However, the snRNP U6 is utilizing the related Sm-Like proteins (Lsm2 to Lsm8) in association with snRNAs in its core structure. [36]

The splicing process starts as U1 snRNP interacts with the 5' splice site of an intron. This interaction is mediated by base-pairing of the U1 snRNA to its complementary sequences within the pre-mRNA. The U2 two auxiliary factors subunits binds to the 3' splice site. While the U2AF 35 kDa bind to the intron-exon border, the U2AF 65 kD bind to the polyprimidine tract upstream of the 3' splice site. All the previous events lead to formation what's called complex E. Afterwards, U2 snRNP is interacting with the branch point via base-pairing of U2 snRNA to form complex A. Later, an assembly of the three complexes U4, U5, U6, is docking to the snRNP in order to form the pre-catalytic complex B that is made up of the whole five snRNPs. Subsequently, major rearrangements happen, leading to release of both U1 and U4 snRNPs as well as the formation of an activated complex B.

Once an activated complex B is formed, the splicing reaction is initiated through the first trans-esterification reaction that includes cleavage of the pre-m-RNA 5' splice site and ligate it to the branch point to form a lariat of the intron sequences. The second trans-esterification reaction takes place through cleavage of the 3' splice site, ligating the two exons and release the mature spliced mRNA. The lariat released is then degraded and all snRNPs are recycled and recruited to other splicing process.[32] Within the catalytic cycle, other enzymes may contribute to the process. For example, ATP-dependent RNA helicases can help reconfiguring RNA-RNA interactions between snRNPs RNA and pre-mRNA. [37]

This process of pre- m RNA splicing was mainly described and characterized in both mammalian and yeast through biochemical purification of different components of the spliceosome as well as in vitro splicing functional assays. However, the splicing process hasn't been studied extensively in plants, in part because the lack of an in vitro splicing assay that recapitulates the splicing reaction in plant systems. Although plant introns have been shown to be precisely spliced in HeLa nuclear extract, mammalian introns weren't efficiently spliced in plants, referring to a possible differences in the intron recognition mechanisms. [38-40] This may be, in part, due to the huge difference between the length of introns between animal, with intron length average 5 kb, and plants, where intron length average is only 160 bp.[41] Moreover, intron sequence composition is considered to be important, with high UA content of plant introns permits efficient splicing. [40-42]

In *Arabidopsis Thaliana's* genome, most of the mammalian and yeast splicing factors and U snRNA counterparts have been identified. That suggests that the main central mechanism and process of splicing reaction are conserved.[43, 44] Furthermore, the frequent presences of some paralogs is referring to a probable level of redundancy in function among splicing factors families.

### *1.3. Mechanism of Alternative Splicing*

The molecular mechanism of alternative splicing is mediated by differential use of some splice sites, so that not all splice sites are used to generate the mature mRNA. As a result, some exons can be skipped and removed with the flanking intron, a splicing event called exon skipping. In other events, whole introns can be retained, designated as intron

retention. (See Fig 1.2). More interestingly, in some events, different 5' or 3' splice sites can be utilized to incorporate variable parts of the exon or to remove variable parts of the intron. Those processes are known as alternative donor and alternative acceptor splicing events. Those splicing events contribute to the major differences between mRNA isoforms generated from the same gene. Encoded proteins from those different isoforms can entail different domain, as a sequence different biological functions. [45] Further, variations of the sequences of those alternative spliced isoforms can promote differential stability, translatability, regulation by microRNAs or subcellular localization. Some isoforms may be directed to RNA decay pathways as a results of alternative splicing. For example, one isoform that contain a retained intron can exhibit open reading frame pre-mature termination due to stop codon in this intron. Also, if alternative splicing add or remove a sequence not divided by three, that may lead to shift of the open reading frame and inclusion of a pre-mature termination codon (PTC). Those PTCs can be recognized through NMD that rapidly and efficiently filter the cell's transcriptome from PTC containing isoforms, thus saving the cell resources and energy waste on producing aberrant nonfunctional proteins [46-50]. Consequently, alternative splicing coupling to NMD may cause quantitative changes of overall and functional transcripts expression levels. [46, 51]

#### *1.4 Regulation of Alternative splicing*

However conserved splice sites are the signals that guide the spliceosome where to splice, pre m-RNA contain other sequence motifs that interact with the RNA-binding proteins to regulate the usage of splice sites. Motifs are divided according to the consequence of binding to them into splicing enhancers and silencers, which can be located within both exons and introns.[52] This interaction between those cis-acting motifs in the pre m-RNA and RNA binding proteins trans-acting splicing factors to determine the splicing reactions outcomes is designated as the splicing code.(see Fig 1.3) [53, 54]In metazoans as well as plants, there are two main classes of splicing factors; serine-arginine rich proteins and heterogeneous nuclear Ribonucleoproteins.[31] In *Arabidopsis Thaliana*, a lot of trans-splicing factors have been identified in the genome. Their role in regulating alternative splicing in the context of growth, development and response to biotic or abiotic stress is undoubtedly of huge importance to understand a lot

of the plant's physiological and pathological processes.

#### *1.4.1 Serine-arginine rich proteins*

Serine-arginine rich proteins are group of proteins that entails mainly two types of domain; (1) one or two RNA- recognition motifs (RRMs) that binds to cis-acting motifs on mRNA and (2) serine-arginine rich (RS) domain that mediates protein-protein interactions.[55, 56] Eighteen SR proteins have been identified in *Arabidopsis Thaliana* genome. However eight of them have counterparts in metazoans, the rest ten proteins are plant specific. [55, 57, 58] Moreover, plant genomes show expansion of SR proteins. While 18 SR genes have been found in *Arabidopsis*, only 12 have been found in humans. Ectopic expression of different SR proteins have led to various physiological and morphological phenotypes and affected alternative splicing of a set of genes.[59, 60]

Alternative splicing of SR proteins has been extensively studied. Interestingly, a lot of SR proteins were found to auto-regulate their own expression through alternative splicing to produce PTC containing transcripts. Those transcripts either got translated into aberrant proteins that lack one or more functional domains or got degraded through NMD pathway. As a consequence, those events lead to alterations in transcripts abundance and functionality. [60-64] Conservation of alternative splicing of SR proteins evolutionarily suggests a physiological relevance of splicing pattern.[65]

It has been found that alternative splicing of SR protein RS31 across different light regimes to be regulated through reduced pool of plastquinone in chloroplasts, suggesting that the retrograde signaling may use alternative splicing to orchestrate nuclear gene expression. [66]

#### *1.4.2 Heterogeneous Nuclear Ribonucleoproteins*

Second class of splicing regulators are called heterogeneous Nuclear Ribonucleoproteins (hnRNPs).[67] They entails two main group of splicing factors; poly-pyrimidine tract binding proteins and glycine rich proteins.

Poly-pyrimidine binding proteins are collection of factors that bind to polyprimidine rich motifs on introns. Extensively studies about those factors in mammals have revealed that they can repress or activate recognition of splice sites close to their binding sites. Further, it was found that the action of those factors can be mediated through combinatorial manner with other hnRNPs as well as other SR proteins. [68, 69]



More importantly, both poly-pyrimidine tract binding PTB1 and PTB2 shows negative auto-regulation through inclusion of PTC by a cassette exon splicing event.[70] RNA-seq analysis of the transcriptome of *Arabidopsis Thaliana* after over-expression or knock-down of PTB1 and PTB2 showed an altered splicing pattern of a plethora of transcripts.[24, 71]

Glycine rich proteins is a group of small RNA binding proteins found in higher plants and represents a simplified version of mammalian hnRNPs. Those proteins harbor RRM domain at N-terminal and a glycine stretch at the C-terminal. They have been found to be involved in plant responses to environmental stresses. Two family members AtGRP7 and AtGRP8 have a prevalent role in alternative splicing. Both proteins were found to play a regulatory role in circadian timing system. AtGRP7 undergoes negative auto-regulation through binding to its own m-RNA. This binding lead to usage of a cryptic splice site that generates a transcript subject to NMD. AtGRP8 uses the same negative feedback loop to control its own expression. Even more interestingly, they reciprocally cross regulate each other's expression. They exhibit the first posttranscriptional feedback loop in the circadian system of any organism. [72-76]

## *1.5 Alternative splicing role in plant adaptation to environmental stresses*

### *1.5.1 Splicing regulation in stress response*

Due to the sessile nature of plants, they can be easily affected by change in environmental conditions. Significant changes in ambient light, temperature, or soil's salt and water content, together referred to as abiotic stress, can greatly influence plant performance. In order to adapt to possible deviations in such environmental conditions, the plant was evolved to rapidly respond to those changes. Those responses include Abscic acid stress hormone induction as well as readjustment of the plant transcriptome to up regulate plant components that are able to promote tolerance of the plant to abiotic stress. However, it was found that both low and high temperature may not only modulate abundance of certain transcripts but also promote alterations in their alternative splicing patterns.[22, 77-79] Recently, alterations in pre-mRNA alternative splicing pattern was observed in *Arabidopsis* seedlings under salt stress.[80]

Heat shock transcription factors are key players of plant's response to an elevated

temperatures through binding to heat shock elements of heat shock proteins genes.[81] HsfA2 of *Arabidopsis* is of the most important transcription factors involve in plant's thermo tolerance. HsfA2 showed two important post-transcriptional modification through AS. First, at 37°C, a 31-bp mini exon in the single intron of HsfA2 is spliced into the transcript, generation HsfA2 II transcript. This exon retains a PTC that leads to direction of the transcript to NMD, thus regulation the level of HsfA2 active protein.[82] Second, at 42°C, another 5' cryptic splice site is activated and resulting in another transcript, HsfA2 III. This transcript codes for a truncated protein, S-HsfA2. This short protein retains a helix-turn-helix DNA binding motif and it was found to bind to HsfA2 promoter heat shock elements, resulting in enhancement of its own transcription. At 45°C, only HsfA2 III was detected rather than HsfA2 II transcript.[83] Such positive feedback loop through AS post-transcriptional modification is a great example of AS involvement in gene expression regulation during response to stress.

### *1.5.2 SR proteins Alternative splicing in plant stress responses*

Pre-m-RNA of *Arabidopsis* SR proteins has tend to show extensive AS under the influence of various environmental stresses and hormones. [22, 80, 84, 85] High salinity stress was found to introduce PTCs into the mRNA of several SR proteins through usage of alternative 5', 3' splice sites or intron retention events. 15 of the 18 *Arabidopsis* SR proteins tested showed altered pattern of alternative splicing when grown in high salinity environment. [80]

Further, a lot of SR proteins exhibit different AS under extreme temperatures. AtSR30 full intact ORF spanning transcript was enhanced at high temperatures and high light density. However, the unproductive PTC-containing transcript was suppressed under high temperatures.[22] Likewise, under dehydration stress, it was observed that SR45a, atypical SR-like protein, full functional transcript abundance increased relative to other splice variants.[86] All those findings were consistent with the suggested role of post-transcriptional AS as an efficient regulator of functional transcripts and proteins abundances under stress conditions.

## *1.6 Signals coupling to alternative splicing regulation*

Rapid reprogramming of the plant's transcriptome and proteome complexity upon stress signal is a necessity to provide a real time response to this stress. Pre-mRNA splicing was found to be regulated by external signals in various studies. [30, 87-89] The mechanisms through which pre-m-RNA splicing is altered by signals in plants are not fully identified.

Posttranslational modifications of splicing factors and splicesomal components, specifically through phosphorylation, was found to regulate splicing in animals. Interestingly, many protein kinases that phosphorylate splicing factors have already been discovered in plants. [90] Moreover, many SR proteins and other splicing regulators were identified as phosphoproteins through phosphoproteomic analysis. [91] Hence, it may be possible that second messengers, like calcium or reactive oxygen species, can modulate the phosphorylation status of splicing modulators through controlling their corresponding kinases and phosphatases' activity. [91] One example of complex regulatory loop that spans AS, NMD as well as phosphorylation is AFC2 gene in Arabidopsis, which codes for LAMMER kinase. This protein is able to phosphorylate SR proteins as well as other splicing factors. In previous research, it was revealed that exon skipping AS events in AFC2 resulted in alternative transcripts that are coupled to NMD, thus lead to alteration of AFC2 functional transcripts abundance. Consequently, any change in LAMMER kinase's abundance can modulate the phosphorylation status as well as activity of SR splicing factors, which in turn modulate alternative splicing.[92] This complex feedback loop of AFC2 gene exemplify the complex nature of coupling between posttranscriptional and posttranslational modification in shaping the plant's transcriptome, especially during response to stress signals.

As many RBPs including splicing factors was found to be sumoylated in global proteomic studies, it is possible that sumoylation is involved in splicing factors activity modulation and AS regulation. [93] Further, changes in cellular environmental conditions might impact RNA secondary structure and promote splicing pattern alteration. Evidence for RNA structure changes in response to ion and osmolyte concentration or temperature have already been provided in previous research. [94-97]

### *1.7 Alternative splicing modulating drugs*

Alternative splicing has been associated with different neurological and muscular diseases. [98-100]. Various studies showed that alternative splicing is contributing to the disease development through changing the topology, solubility and signal peptides of integral membrane proteins.[101] Further, aberrant splicing events have been identified in Parkinson's disease.[102] More specifically, in Spinal Muscle atrophy (SMA), the survival motor neuron (SMN) protein was revealed to be down regulated due to it exon skipping event of exon 7 from SMN m-RNA.[103] In cancer related genes, alternative splicing has been recruited to promote oncogenesis, tumor suppression[104] and metastasis.[105] Changes in alternative splicing patterns have been found in various types of cancer. [104-107]

Therapies for human diseases as well as cancer therapies through splicing modulation have been developed.[108-110] Therapeutic approaches in splicing associated diseases include mainly restoring splicing through Antisense Oligonucleotides (ASO) and targeting splicing pathways through small molecules.

To restore splicing, as ASO is used to target and specifically bind to 12-50 complementary nucleotides to sterically block splicesomal components interaction with splicing enhancers, and thus lead to exon skipping, or splicing silencers, and thus result in exon inclusion. The efficiency of the approach has been proved through using mouse model for SMA disease. Using an intracerebroventricular (ICV) injection of one ASO (ASO-10-27) lead to dose-dependent enhancement of exon 7 inclusion, SMN protein synthesis, improved muscle physiology, motor function and survival rates.[111] Interestingly, systemic delivery of (ASO-10-27) lead to increase of SMN protein level in both peripheral and central nervous system as well as increase survival from 10 days to 500 days[112]. Collectively, these good results in mouse models encouraged FDA to grant an approval on human clinical trials.

Through the small molecules approach, different splicesomal components or splicing factors can be targeted and lead to radical alterations in alternative splicing patterns. Many antitumor drugs have been found to possess splicing inhibition properties. Those drugs include herboxidiene, spliceostatin A, pladienolide and isoginkgetin.[113-

116].

All those drug showed antitumor potency over human cancer cell lines and xerograft mouse models.[115, 117-119] Herboxidiene and pladienolide B are structurally similar and both act through binding to SF3b components of the U2 snRNP thus inhibiting U2 snRNP function. Spliceostatin A also target a protein that is a component of U2 snRNP, but rather than promoting general splicing inhibition, it initiates alterations in splicing of cell cycle genes. [120]

Cdc-2 like kinases (Clk kinases) are involved in phosphorylation of SR proteins, a process found to be critical for SR proteins activity and nuclear localization. [121, 122]Targeting splicing through inhibition of such kinases can alter the activity of SR proteins splicing regulators. In previous studies, screening for molecules that alter the splicing of BCL-X, HIPK3, and RON transcripts hepatocarcinoma cell line have found amiloride as potent splicing modulator of those transcripts.[123] Cells treated with amiloride showed hypophosphorylated SR proteins and decreased level of SFSP3 (SRp20).[123] Consequently, small molecules altering the activity of certain splicing factors can be used to modulate the alternative splicing of cancer- related genes and thus exhibit therapeutic potentials.

### *1.8 Selective alteration of alternative splicing through targeting serine arginine rich splicing factors by Indole derivatives.*

Compounds of the Indole derivatives have proven to be potent splicing inhibitors. Derivatives of Indole chemical nucleus showed the ability to specifically inhibit exonic splicing enhancer ESE dependent splicing. This action is mediated by direct and selective interaction with the members of serine-arginine rich proteins family. [124]

It was early revealed that drugs that hinder kinase activity of topoisomerase I, and thus disrupt the phosphorylation status of SR proteins, showed a splicing inhibition potential and altered the splicing pattern of many genes.[125, 126] SR proteins phosphorylation posttranscriptional modification was shown to be essential for ESE-dependent splicing. [127] Indole derivatives was found to prevent phosphorylation of RS domain of SR proteins by topo I kinase and to lesser extent by Clk-sty kinases. However some Indole derivatives were able to inhibit both ESE events associated with SC35 or

SF2/ASF SR proteins, some other derivatives had specificity to one SR proteins.[124]

HIV-1 uses various 5' or 3' alternative splice sites to generate more than 40 different m-RNA and proteins for its replication. This process is mediated by interaction of specific HIV m-RNA with SR proteins [128]. Remarkably, compounds of Indole derivatives showed potent inhibition of HIV-RNA production in HIV chronically infected cells. Selective and direct interaction of Indole derivatives with SR proteins family members may explain the splicing inhibition and alteration of HIV-1 m-RNA appeared under the influence of Indole derivatives. [124]

Further confirmation of the therapeutic potentials of Indole derived drugs (IDCs) have been demonstrated through targeting ESE dependent splicing involved in cancer metastasis.[129] *Ron* (Recepteur d'origine nantais) proto-oncogene produces four different m-RNA transcripts. One specific transcript  $\Delta$  *Ron*, generated through exon 11 skipping event, has been associated with enhanced cell motility [130-132]. This splicing event has further linked to Epithelial- Mesenchymal transition process (EMT) and metastasis [133]. On molecular level, splicing of exon 11 is mediated through two elements, an exon silencer and exon enhancer, both present in constitutive neighboring exon 12.[134] Direct binding of SF2/ASF SR splicing factor, proto-oncogene up-regulated in many human tumors[135], to the enhancer sequence on exon 12 result eventually in formation of  $\Delta$  *Ron* transcript with its consequences on cell motility and EMT. [134]

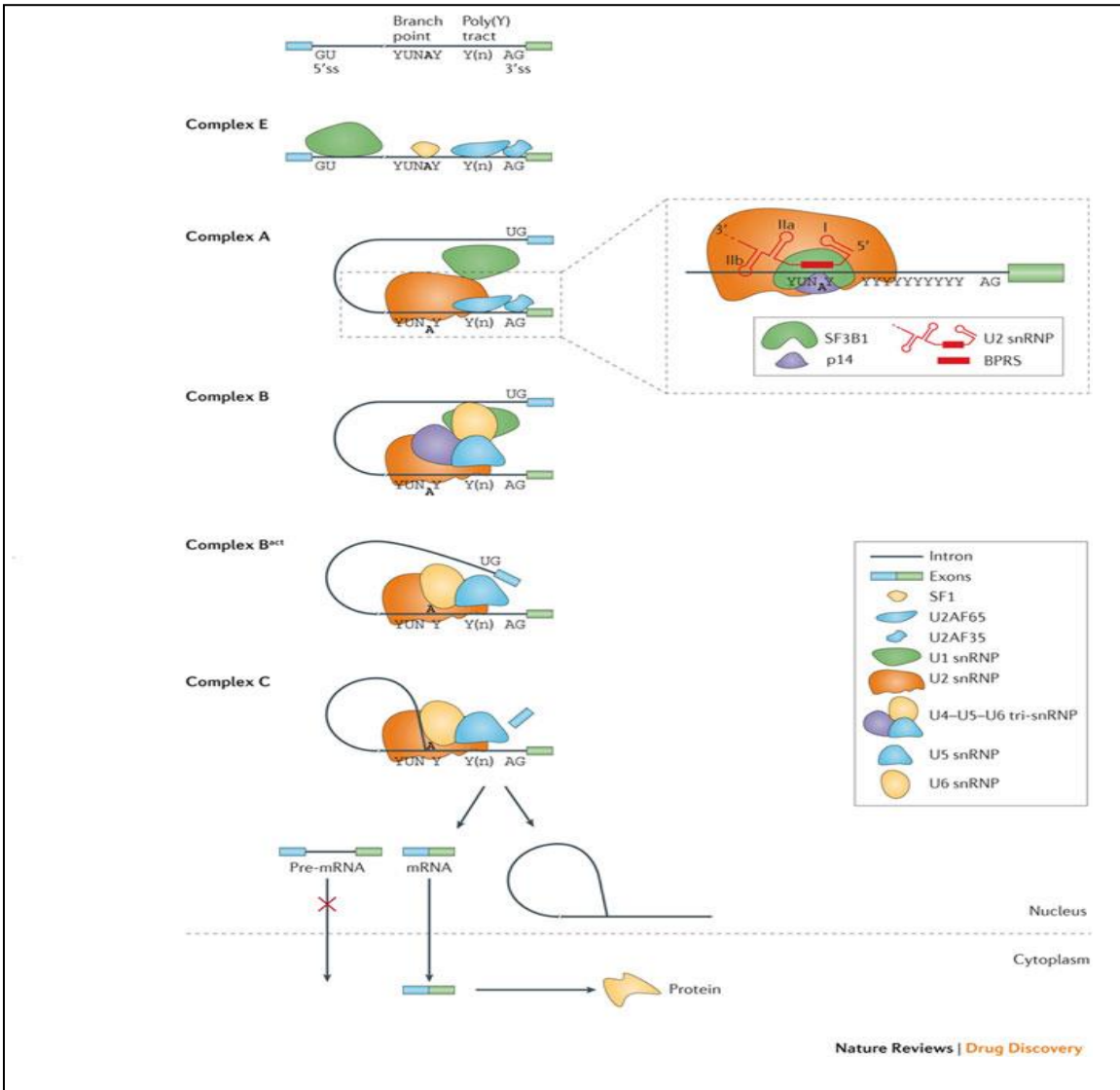
HeLa cells characterized by their preferential skipping of *Ron* exon 11 and expression of  $\Delta$  *Ron* transcript has been treated with 4 different IDCs Indole derivative, which showed previously potent inhibition of SF2/ASF ESE-dependent splicing. Three different IDCs, namely IDC48, IDC78 and IDC92 (nomenclature according to the original study) were found to specifically inhibit skipping of exon 11 and reduce the levels of  $\Delta$  *Ron* transcript. More importantly, dose effective curve showed that drug IDC92 was potent in *Ron* exon 11 inclusion even at 0.5  $\mu$ m while  $\Delta$  *Ron* became the less abundant transcript at 4  $\mu$ m, yet such concentration yield modest effect on *Caspase9* transcripts alternative splicing[129]. Collectively, all those findings support the therapeutic potency and specificity of targeting SR proteins mediated ESE-dependent alternative splicing by Indole derivatives.

### *1.9 Study objectives and design*

In this study, we are trying to modulate alternative splicing in plants through targeting serine-arginine rich proteins splicing factors by three different Indole derivatives, 6-Methylindole, 2, 5-Dimethylindole, 5-Bromoindole-2-Carboxylic Acid. Different derivatives of Indole nucleus have showed potency in inhibition of ESE dependent splicing in both in-vitro splicing systems and cell lines. To our knowledge, this is the first study to target ESE dependent splicing in plants.

This study aims mainly to reveal the scope of splicing alterations promoted by the three Indole derivatives under study. Splicing alterations shall be described in terms of global enhancement or decrease of AS under the effect of this derivatives. This study further investigate the functional categorization of genes experienced isoform shift between each treatment and control. Moreover, due to nature of coupling between gene expression and AS, the study aims too to elaborate the differential gene expression promoted by alteration in AS under Indole treatment. Finally, the study tries to give more insights into the potential molecular function changes or outcomes of certain differentially spliced transcripts to investigate the complex feedback and cross-regulatory roles between gene expression, post-transcriptional and post-translational modifications.

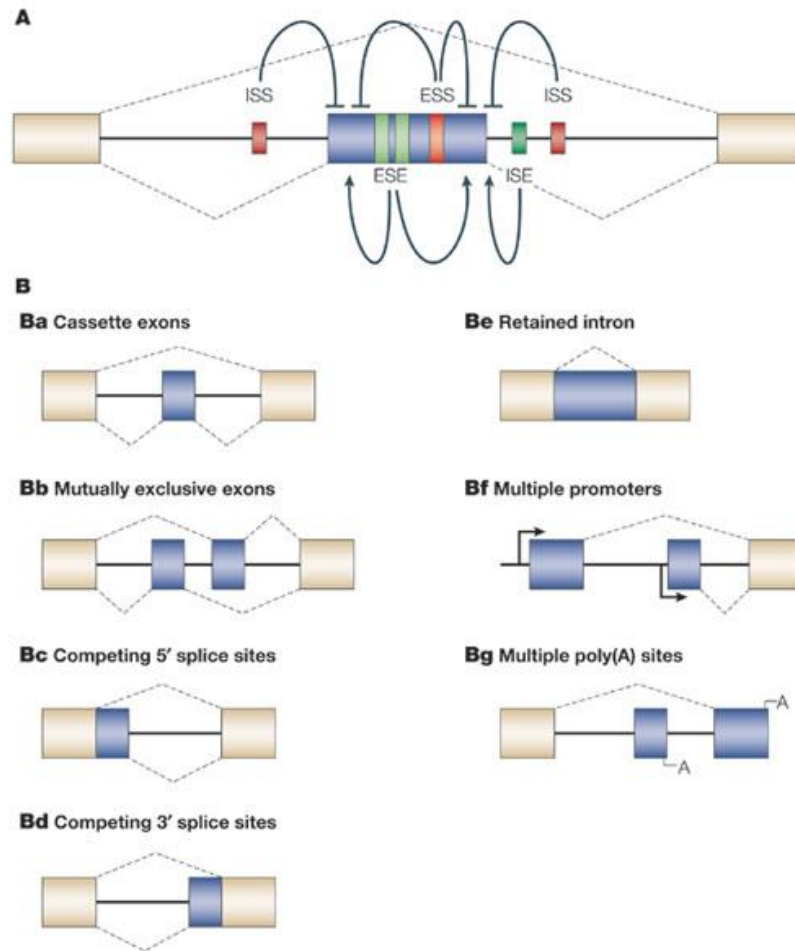
To achieve those objectives, paired-end RNA-seq of 32 samples (4 replicates\*2 time points\* 4 conditions [1 control + 3 treatments]) was used to give a snapshot of the transcriptome in different control, treatments and time points. Data generated was further analyzed for alternative splicing changes using JuncBASE, differential expression analysis using Tuxedo protocol, and functional insights into the differentially spliced transcripts using Integrative Genomic Viewer (IGV) visualization tool.



**Fig 1.1 Mechanism of Splicing and splicesome assembly on splice site.**

[136]

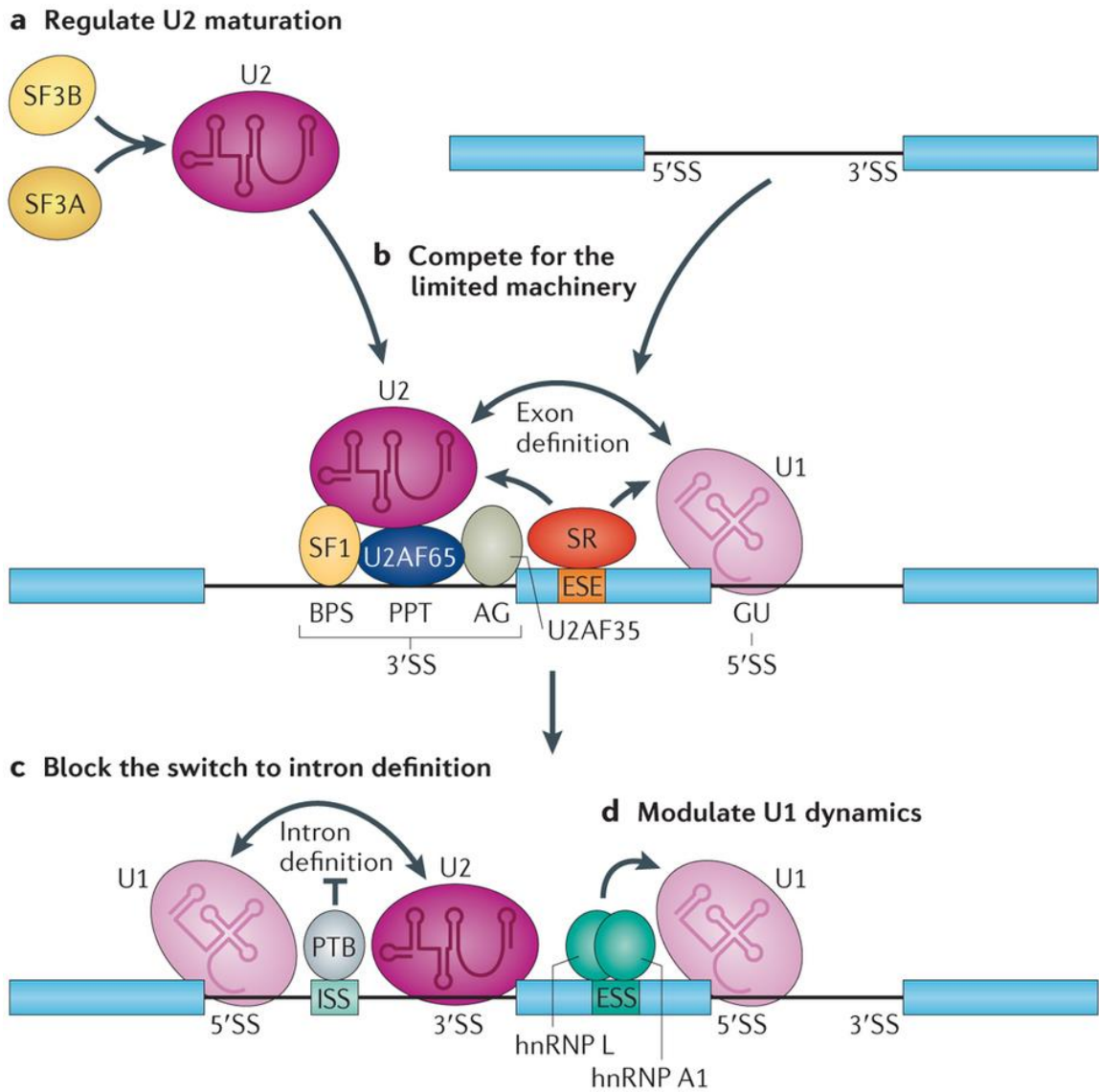




Nature Reviews | Molecular Cell Biology

**Fig 1.2** *Cis* acting factors and types of alternative splice events..

[54]



Nature Reviews | Genetics

Fig 1.3 Alternative splicing regulation through trans-acting splicing factors.

[31]

## CHAPTER 2: MATERIALS AND METHODS

### *2.1. Samples preparations and treatments*

Six days old *Arabidopsis Thaliana* Col-0 seedlings have been transferred to plates that contain 6-Methylindole (MW: 131.17), 2, 5-Dimethylindole (MW: 145.20) or 5-Bromoindole-2-Carboxylic Acid (MW: 240.05) at the same concentration of 10 $\mu$ m. Chemicals were ordered from Sigma-Aldrich, prepared in stock solutions of DMSO in 100mM concentration and stored at -20°C. Each drug treatment was extended to both 24 hours and 72 hours before RNA extraction. Control samples that received no treatment were prepared and harvested for RNA extraction also at both 24 and 72 hrs time points.

### *2.2. RNA extraction, library preparation and sequencing*

Whole seedlings have been used for subsequent RNA extraction and sequencing. For RNA extraction step, total RNA was extracted first using Triazol method. Strand-specific mRNA-seq library was prepared. Paired end sequencing was performed on Illumina HiSeq2000 sequencing platform (Illumina Inc., California, USA).

### *2.3. RNA-seq reads quality control and quality assessment*

Sequencing output included 64 fastq files for 32 samples sequenced in this study, so that each sample generated one file of reads for each end of any fragment sequenced. Fastqc(v 0.11.2)[137](available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) has been used to generate quality reports of all fastq files.

Fastqc reports showed that about 12 to 14 million reads per sample have been generated during sequencing, with a uniform read length 101. In the reports, per base sequence quality plots showed no base quality dropped below Phred quality score 33. Further, there were no overrepresented sequences, adapter content, nor reads flagged as poor sequences. Per Base N content was absent across all bases. Sequence quality scores plots showed average quality read around 37 with almost no read with overall quality below 33. However, a very few (around 9 bases) fluctuation in per base sequence content was found. Overall, as those results were consistent among the 32 study samples, the quality of the sequencing seemed promising and no trimming was needed to enhance the

quality of the data. So, Fastq files were proceeded to the reads mapping and alignment untrimmed.

#### *2.4 RNA-seq reads alignment and junction prediction*

Tophat software[138] has been used to align the generated reads of all the samples against the latest release of Arabidopsis genome sequence and annotation of gene models of TAIR10 (TAIR10.30 version), which was downloaded from ensemble plants File transfer portal at (<http://plants.ensembl.org/info/website/ftp/index.html>).

As a splice-aware aligner, Tophat first reads known junctions from the annotation file. From this annotation data, the software builds a transcriptome file that contains annotated exons and junctions coordinates. After that, Tophat uses Bowtie [139] aligner to map the RNA-seq reads to this transcriptome.

Interestingly, tophat has the ability to discover novel junctions, those junctions that their coordinates haven't been supported by annotation data currently available in TAIR10 (version TAIR1030). Default parameters have been used in tophat run.

Two important files have been produced from tophat run on each sample of the 32 samples in this study. First, a .bam file that contain the mapping information of the accepted reads hits. Second, a junction.bed file that contains the information of splice junctions found/predicted in each sample. In our basic tophat run, novel splice site discovery feature was enabled, so the output contained information of known as well as novel junctions. The output of this run, as being the most possible inclusive, has only been used in all downstream analysis.

Another side tophat run, where novel junction discovery feature is disable, has been conducted. The purpose of this side run was to gain quantitative information about difference between known and novel splice junctions per sample. As this run will generate only known junctions data, comparison with the first basic tophat run, that generate information about both known and novel combined, has helped to infer statistics about both types of junctions in each sample. However, this later side run outputs haven't been used in any of the downstream analysis.

Tophat version used was v.2.1.0. (<https://ccb.jhu.edu/software/tophat/index.html>). This version uses bowtie 2.2.6.0 as an aligner. Both programs were installed and used on Linux operating system on metagenomics-bio server (IP access 10.7.28.10) of the Biology

department, School of science and engineering, American University in Cairo. The terminal was accessed often through secure shell app of Google chrome internet browser.

### *2.5 Differential gene expression analysis*

Mapped reads haven been then used to assemble and quantify all transcripts in each of our samples, separately, through cufflinks software[140, 141] (v2.2.1), available to download at (<http://cole-trapnell-lab.github.io/cufflinks/install/>). Afterwards, annotation of all possible transcripts from each sample have been merged to generate one annotation file for all gene models possible in our samples pool. This has been achieved using cuffmerge program. Further, gene, isoform quantification and differential analysis between control and treatment in each time point has been performed through cuffdiff program. This differential expression analysis has been guided by the merged gene models annotation generated through cuffmerge in the previous step. Finally, visualization of the cuffdiff analysis output mainly utilized cummrbund bioconductor package (<http://bioconductor.org>) that runs on R statistical and computing programming environment. (<https://www.r-project.org>) (v3.2.3).

### *2.6 Alternative splice events annotation from confident splice junctions*

JuncBASE[142] (Junction Based Analysis of Splicing Events) protocol has been followed to accomplish identification and classification of splicing events in each sample. The protocol utilizes information from splice junction reads alignment of RNA-seq reads as well as annotated exon coordinates to analyze alternative splicing events. Furthermore, JuncBASE makes use of read counts assigned to exclusion and inclusion isoforms to develop “Percent-spliced in” value for each splicing event. JuncBASE software (v0.6) used in this study is available for download at (<http://compbio.berkeley.edu/proj/juncbase/Home.html>)

This protocol can further uses those values to identify splicing events that are differentially expressed across samples, which indicates a differential alternative splicing. JuncBASE protocol was initially developed in order to characterize both novel and annotated splicing events during *Drosophila* development as well as identify those events which altered upon knock down of splicing factors.

As this study aims mainly to study the consequences of targeting serine-arginine rich proteins splicing factors on alternative splicing, JuncBASE protocol was a very good candidate to perform such analysis.

The workflow of JuncBASE starts with building a database for both reference annotation exons and introns from reference gene models annotation file. The same gene model reference annotation file used for mapping reads with Tophat has been used for this step. Afterwards, in order to remove putative false positive splice junctions, a plot of entropy scores of each sample splice junctions has been generated using plot entropy feature of JuncBASE. An entropy score is a function of both the total reads assigned to a certain splice junction, the number of possible different offsets to which those reads map as well as the number of reads that mapped to every offset.[143] So, splice junctions where many reads mapped to each of the possible offsets positions across the junction will receive a higher entropy score than other junctions, where many reads mapped to only one or two offsets positions.

$p_i = \text{reads at offset } i / \text{total reads to junction window}$

$\text{Entropy} = - \sum_i (p_i * \log(p_i) / \log 2)$  [143]

In our study, and based on samples' splice junctions entropy plots generated, an entropy score of 2 has been used to filter putative false positive novel junctions. This value was chosen as it overlaps with normal distribution curve of entropy scores of known reference annotated junctions. Further, a minimum overhang of 10 alignment of reads on either side of intron-exon boundary was required to assign this reads to this intron-exon junction. Those two criteria were used to generate a set of confident splice junctions for each sample to be used for downstream alternative splicing events identification and quantification.

To identify and quantify alternative splice events, reference annotation of introns and exons coordinates, created in first step, has been used to annotate introns, exons, confirm first and last exons as well as assign gene names and decide whether each splicing event is novel or previously annotated. Finally, tables of raw and length normalized counts of both inclusion and exclusion isoforms of every splicing events have been developed to be used later for differential splicing analysis.

To generate counts of total and different events types in each sample, tables of length normalized reads counts for have been opened in Excel sheets (Microsoft office 2013, Microsoft Inc.). Then, plots of events counts have been generated after reading the tables in R programming environment through using ggplot2 CRAN package (<https://cran.r-project.org/web/packages/ggplot2/index.html>).

### *2.7 Differential alternative splicing analysis*

Tables of length normalized exclusion and inclusion counts have been utilized to calculate percent spliced in values for each event in each samples. Percent spliced in can be calculated as follows:

Percent spliced in =

$$\frac{\text{counts of reads mapped to inclusion isoform of the splicing event}}{\text{counts of reads mapped to both inclusion and exclusion isoforms}} * 100$$

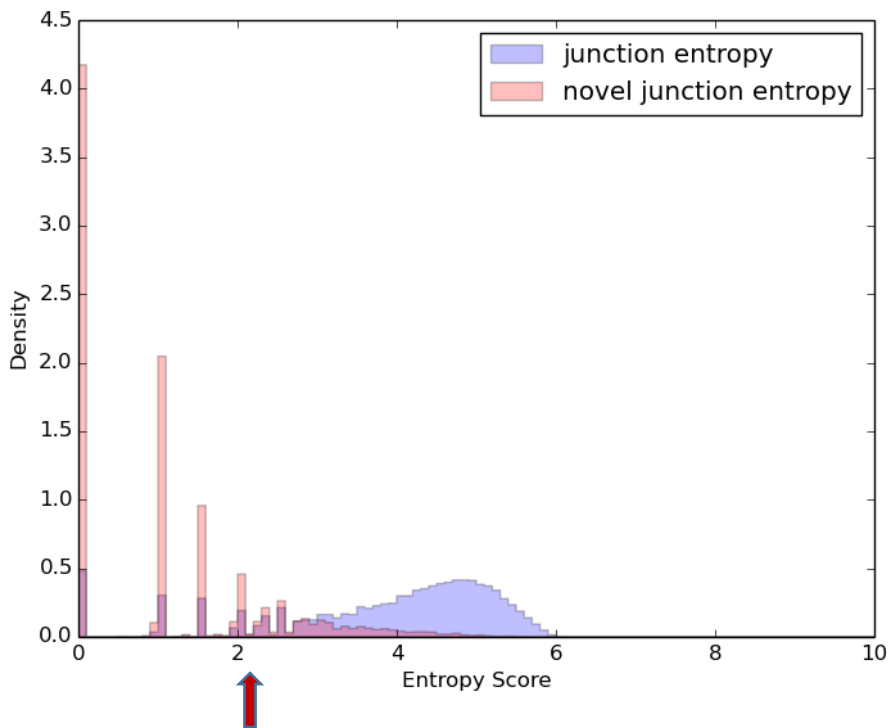
To perform differential alternative splicing analysis, paired t statistical test, have utilized the percent spliced-in values of every splicing event to calculate the p-value of the difference between each treatment and its corresponding control. In the context of our study, 6 paired t-test have been conducted between 4 replicates populations of control and three treatments at two time points. In each comparison, a table of the paired t-test results containing p-value of the difference in percent spliced in, the type of event, whether the event novel or annotated, gene name, chr number as well as genomic coordinates of inclusion and exclusion junctions and isoforms.

Further, the output tables of every paired t-test conducted has been filtered according to a cut-off p-value. Those events who showed a p-value lower than 0.01 have been chosen as statistically significant events.

Finally, and in order to determine the biological relevance of statistically significant events, lists of genes showed statistically significant differential alternative splicing have been created and submitted to DAVID (<https://david.ncifcrf.gov/>) for functional annotations as TAIR IDs (The Arabidopsis Information Resource ID). Results of functional ontologies and enrichment of each treatment group of genes have further plotted using ggplot2 in R programming and statistical computing environment.

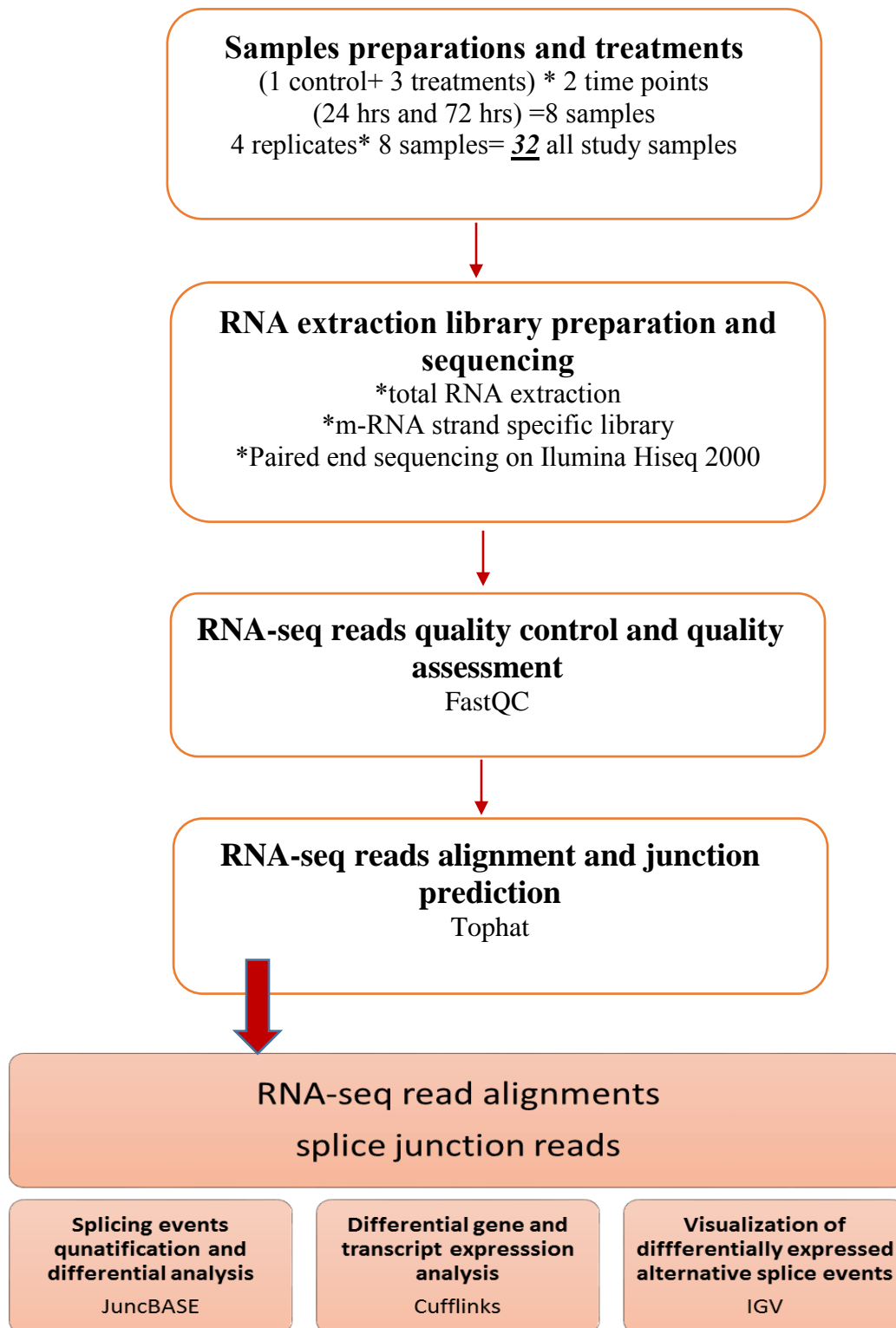
## 2.8 Visualization of differentially expressed alternative splice events

Visualization of the mapped reads to the genome in each sample and compare differential coverage to certain intronic or exonic sequences among samples can help to infer the functional consequences of such splice events. To achieve this step, integrative genomic viewer (IGV) [144] have been utilized to visualize differentially expressed alternative splice events. IGV starts with reads aligned in .sam format files. So, samtools have been used to convert .bam files for our 32 samples into .sam files. After that, IGV tools, feature of IGV, have been used to generate two types of files, sorted and indexed .sam files of each sample. All files have been loaded on visualization session of IGV to generate sashimi and coverage plots for differentially expressed alternative splice events.



**Fig2.1 Density distribution bar plot of known and annotated splice junctions in replicate 1 of control sample at 24 hours' time point.**





**Fig2.2 Flow chart of the study design**

## CHAPTER 3: RESULTS AND DISCUSSION

### 3.1 Global alterations of alternative splicing

From the RNA-seq data, 22478, 19166, 23577, 20897 alternative splice event have been annotated and discovered by JuncBASE ( with parameters specified at Materials and Methods) from control, 6-Methylindole, 2, 5-Dimethylindole, 5-Bromoindole-2-Carboxylic Acid treated samples at 24 hrs time point, respectively. At 72 hrs time point, 15340,18535,19319,26588 alternative splicing events have been described (See Fig 3.1). Majority of the events described at 24 hrs time point 4 samples are found to be novel and not previously annotated at *Arabidopsis* gene model annotation used. In contrast, majority of events described at 72 hrs time point were actually annotated. (See Fig3.2 and Fig3.3). Moreover, counting of each splice event type in each sample has been performed and plotted (See Fig 3.4). Eight types of alternative splicing events have been characterized in this analysis. Those events include; alternative donor or alternative 5' splice site, alternative acceptor or alternative 3' splice site, cassette exon, alternative first exon, alternative last exon and intron retention. In *Arabidopsis*, intron retention and alternative acceptor have been found to be the most prevalent splicing events, which is apparently different from humans, where cassette exon (also known as exon skipping) is the most prevalent splicing event. [22]

Our splicing events analysis has revealed global alterations in alternative splicing after treatment with Indole derivatives. In 4 out of 6 treated samples, alternative splicing analysis has showed enhancement of AS under the effect of Indole derivatives relative to control treatments. More interestingly, all three Indole derivative treated samples at 72 hrs time point showed enhancement of the alternative splicing. Pattern of splicing enhancement might seem to increase with the Indole derivative chemical structure complexity, as the degree of AS enhancement increased gradually under the effect of 6-Methylindole, 2, 5-Dimethylindole, 5-Bromoindole-2-Carboxylic Acid, respectively. However, 2, 5-Dimethylindole treated samples showed AS enhancement at 24 hrs time-point. Yet, both 6-Methylindole and 5-Bromoindole-2-Carboxylic Acid treated samples showed decrease in AS relative to the control sample at 24 hrs time point.

This may suggest that different Indole derivatives exhibit different actions towards splicing alteration. The fact that the majority of enhanced AS events at 72 hrs time point are belonging to previously annotated splice junctions or splice events supports the validity of the AS enhancement and discredit the probability that the enhanced AS observed may result from false positively discovered splice junction or splice event.

Overall, those observations of global alterations of AS, mainly towards enhancement, is consistent with the hypothesized action of the Indole derivatives compounds (IDDs), where the main mechanism of action is an inhibition of ESE dependent alternative splicing through targeting SR proteins. Such splicing inhibition shall promote failure of proper splicing or removal of nearby intronic sequences and as a consequence expand the degree of alternative splicing event. That's specifically valid and expected in an organism where intron retention represents the extreme majority of alternative splicing events. However, it remain unclear why the AS showed decrease in 6-Methylindole and 5-Bromoindole-2-Carboxylic Acid treated samples at 24 hrs time point.

More interesting and consistent with the later hypothesis, analyzing alteration in each type of splicing event separately revealed that always intron retention is the most altered splicing event. Intron retention, 3' alternative acceptor and 5' alternative donor splicing events alteration pattern is following the same overall splicing alteration pattern with no exceptions. Intron retention seems to be the most enhanced splicing event enriched specifically at 72 hrs time point. Samples treated with 5-Bromoindole-2-Carboxylic Acid showed almost more than double and half the number of splicing events appeared in the control samples. Moreover, alterations in exon skipping events, alternative first and last exon, coordinate exons as well as mutually exclusive exons showed slighter splicing alterations though those low representation of those splicing events in *Arabidopsis* transcriptome may discourage us from drawing conclusions on splicing pattern alteration they showed.

Collectively, observation of global and event specific splicing pattern alteration under the influence of Indole derivatives used in this study strongly recommend that after certain onset of time, 72 hrs, splicing inhibition promoted by those derivatives lead to significant enhancement of AS, mostly and specifically intron retention splicing events.

### *3.2 Differentially spliced genes associated with Indole derivatives treatment*

Enhancement of AS under Indole treatment lead to expression of alternative events that were not expressed in the control. Though, it is further interesting to investigate the isoform shift associated with each drug treatment and determine the functional scope of genes experienced such changes. Isoform shift mainly is a phenomena when a gene has two or more isoforms, yet only one is the most frequent relative to the others. In certain conditions, this distribution of isoforms is altered, so that the dominant isoform is shifted to be less frequent relative to one or two of the previously less frequent ones. Isoform shift may be recruited to produce proteins with altered or modified function, modify gene expression through NMD or whatever purpose to accommodate physiological changes associated with development, growth as well as response to environmental stresses.

JuncBASE produced percent spliced in values (for full description check Materials and Methods) and it was used to infer isoform shift but on splice event level. Splice events would be considered to contribute to isoform shift when a statistical significant ( $p\text{-value} > 0.01$ ) change is observed in percent spliced in values between control and treated samples. Only events that pass this cut-off  $p$ -value limit have been labeled as significant and utilized in downstream functional analysis. Plots of number of splice events who passed this cut-off and considered statistically significant differentially expressed alternative splice events have been generated. Further, plots of the number of genes experienced isoform shift under Indole treatment conditions at the two examined time points was plotted to infer how wide the scope of isoform shift promoted by derivatives under study. Moreover, functional categorization of the genes underwent isoform shift was determined and plotted to determine the scope of functions affected.

Our results (see Fig3.5) show that 293, 303 and 399 events in 6-Methylindole, 2, 5-Dimethylindole and 5-Bromoindole-2-Carboxylic Acid treated samples, respectively, showed a statistically significant differential splicing at 24 hrs time point. However, 187,146 and 366 statistically significant events were found, also respectively, at 72 hrs time point.

Moreover, plots of genes exhibited differential splicing (see Fig 3.6) or isoforms shift showed that 148, 142 and 199 genes at 24 hrs time point and 144, 83 and 214 genes at 72 hrs time point were significantly altered in 6-Methylindole, 2, 5-Dimethylindole and 5-Bromoindole-2-Carboxylic Acid samples, respectively.

Moreover, number of genes differentially spliced got plotted along with the types of splicing events to infer which splice event type experienced differential splicing more frequently under our study treatments. Interestingly, Alternative acceptor, rather than intron retention, was the most frequently differentially spliced event in most of treatments (see Fig 3.7). Alternative donor and intron retention have been found to be the second and third most frequently spliced event, respectively. Indeed, this may suggest splicing inhibition through Indole derivatives may recruit alternative 3' and 5' splice sites more than intron retention to introduce isoform shift in many genes.

Further, functional categorization of all genes got differentially spliced under our study treatments showed repeated enrichment of two functional categories across most of treated samples at the two time points, (See Fig 3.8-3.13). First function category enriched is RNA splicing genes (represented by terms; m-RNA processing, m-RNA splicing, RNA-binding). Most of those genes are found to be also phospho-proteins. This is consistent with the previous studies which revealed extensive phosphorylation of RNA splicing and processing proteins [91], which suggests post-translational level of regulation of RNA splicing genes through phosphorylation. Yet, as Indole derivatives inhibit phosphorylation of SR proteins and thus render them inactivated, prevalence of auto-regulation of SR proteins and cross regulation between them and other phospho-proteins splicing factors can give an a good explanation of the pattern of functional enrichment present. Inhibition of phospho- proteins serine arginine rich splicing factors lead to frequent alterations in their splicing patterns, as they tend to undergo auto-regulation of their alternative splicing.

Further, those alteration of SR proteins activity tend to cause alteration in other RNA processing and splicing factors due to cross regulation among RNA splicing factors. Glycine rich proteins, especially (GRP7 and GRP8), tend to be the most frequently splicing factor showed alteration of alternative splicing. More interestingly, enrichment of the term (kinases) in some samples may suggest a complex feedback or coupling between

alternative splicing post-transcriptional modification and phosphorylation posttranslational modification to regulate plant physiological changes and responses.

Another main functional category was found to be enriched in most of sample is chloroplast-related genes. This category was represented by the terms; chloroplast, thylakoid, plastid and photorespiration. Such enrichment may suggest retrograde regulation between chloroplast and nucleus that may be mediated through alternative splicing. It is further consistent with the recent idea suggesting the role of AS in plant's response to light and promotion of photo-respiration. [145]

Finally, enrichment of transcription, transcriptional regulation or nucleic acid binding related genes among differentially alternatively spliced genes in some treated samples may confirm another way of controlling expression level of downstream target genes through other than NMD which includes AS of transcription factors.

### *3.3 Differential gene expression analysis*

Gene expression differential analysis revealed that 9179,9325 and 12327 genes have been found to be differentially expressed between control and 6-Methylindole, 2, 5-Dimethylindole and 5-Bromoindole-2-Carboxylic Acid treated sample treated samples at 24 hrs time point, respectively. Also, 8870, 9140 and 9439 genes were further differentially expressed, in the same treated samples order, at 72 hrs time point. Differentially expressed genes are only those which showed differential expression p-value <0.01.

Due to that huge number of genes differentially expressed (which ranges from 25-35% of around 33 thousand genes in *Arabidopsis*), list of genes that showed the highest expression difference were generated by filtering genes ( $\geq 450$ ) absolute expression difference in FPKM unit. Clustered heatmaps of those genes were also plotted for each pair of comparison in this study. (See Fig 3.13-3.20).

Clusters of genes differentially expressed along different samples ( check Table 3.1 and 3.2) and time points showed potential consistency which may infer similarity in expression alteration in response to Indole derivatives under study. Most importantly, different members of photosystem, photo-respiration and photosynthesis have been found to be differentially expressed in almost all treatments.

Those members include; RBCS1A, RIBULOSE BISP HOSPHATE CARBOXYLASE SMALL CHAIN 1A, RBCS2B, RBCB3 all belong to RBSCO family known for its involvement in carbon fixation using light. Further, Light Harvesting complex of photosystem II family members; LHC1A, LHC1B, and others were also frequently differentially expressed. In most treatments, up-regulation of those genes have been observed, especially at 24 hrs time point treated samples. Yet, down-regulation of those photosynthesis related families have been observed under some treatments.

Both of our differential expression and splicing findings reveal that photosynthesis and photorespiration related genes are frequently altered on the level of alternative splicing and gene expression under the effect of Indole derivatives used in our study. Those findings strongly indicate regulation of photorespiration by AS. Overall, this is consistent with the recent findings that Phytochrome A and B both modulate gene transcription and alternative splicing in response to light and to promote photorespiration in *Arabidopsis*. [145]

### *3.4 Functional consequences of specific alternative splicing events promoted by different Indole derivative.*

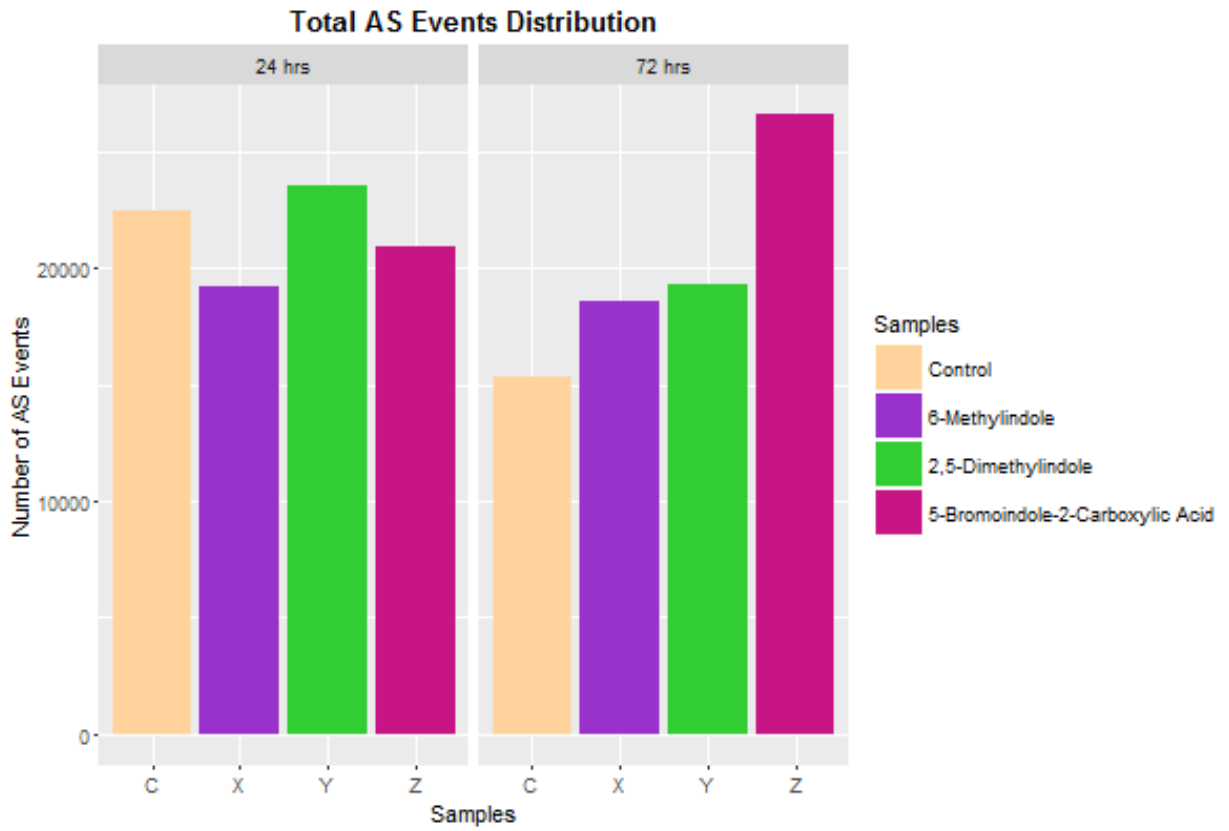
To get more insights into the molecular aspects of differential alternative splicing promoted by Indole derivatives under study, we utilized Integrative genomic viewer to plot three of the most statistically significant differentially splicing events in order to try to get more sense of the potential outcome of such splicing alteration.

First, plots of coverage data in both control and 2, 5-Dimethylindole of hydrolase-like protein (AT5G17670) against the annotated gene model (see Fig 3.21) revealed a differential usage of novel 5' splice site. First GT splice site has already been annotated in gene model and second splice site GT is a novel splice site that has been discovered in our data. Strikingly, this lower differential inclusion of 4 bases through alternative 5' splice site under 2, 5-Dimethylindole treatment lead to skipping of UAG( TAG on DNA level) pre-mature termination codon. Such pre-mature termination codon can have severe consequence on the abundance or functionality of protein, either through PTC-containing transcript degradation through NMD or through producing truncated protein. This 5' alternative splice event has been discovered through JuncBASE and differential 5' alternative donor was noted as statistically significant (p-value<0.01).

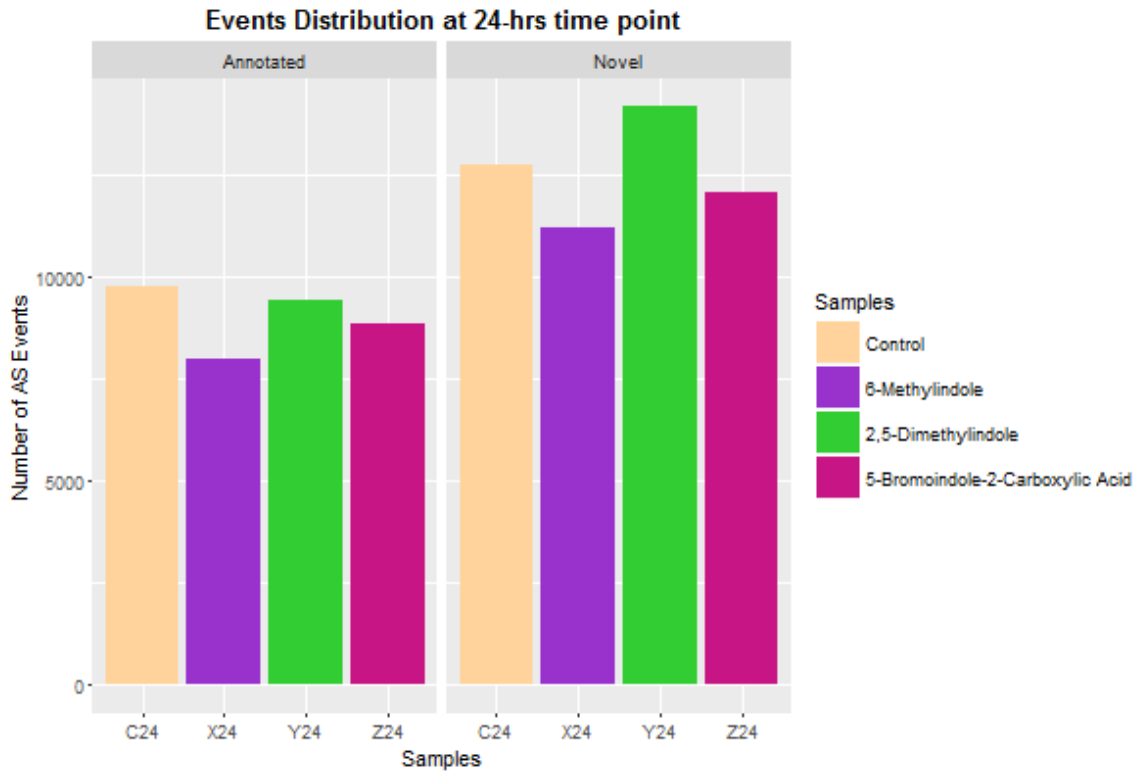
Another interesting example of alternative splicing alteration in our study is SR34A differential exon skipping. (See Fig 3.22) SR34A is a splicing factor and one of serine-arginine rich proteins splicing factor, which is hypothesized to be targeted by Indole derivatives and is known for auto-regulating itself. Differential exon skipping under the effect of 2, 5 Dimethyl Indole and 6-Methyl Indole relative to control was discovered by JuncBASE and noted to be statistically significant alternative splicing event ( $p$ -value $<0.01$ ). Although this exon have been annotated in previous studies, no information about the functional consequences of exon skipping is available. Yet, the fact that the exon lies in the 5' untranslated region of the gene model may indicate that skipping this exon may help the m-RNA generated to bypass regulatory mechanisms that target this area, like microRNA, in a similar fashion to metazoan. [146]

Finally, differential Intron retention event between control and 2,5-Dimethylindole at 72 hrs time point of CPK28 calcium dependent protein kinase was noted to be statistically significant ( $p$ -value $<0.01$ ). This novel intron retention has never been annotated in previous studies and its function remains to be elucidated. Yet, such AS mediated functional alteration of a protein kinase can further infer a strong coupling between AS and Phosphorylation, which was suggested by multiple findings in this study.

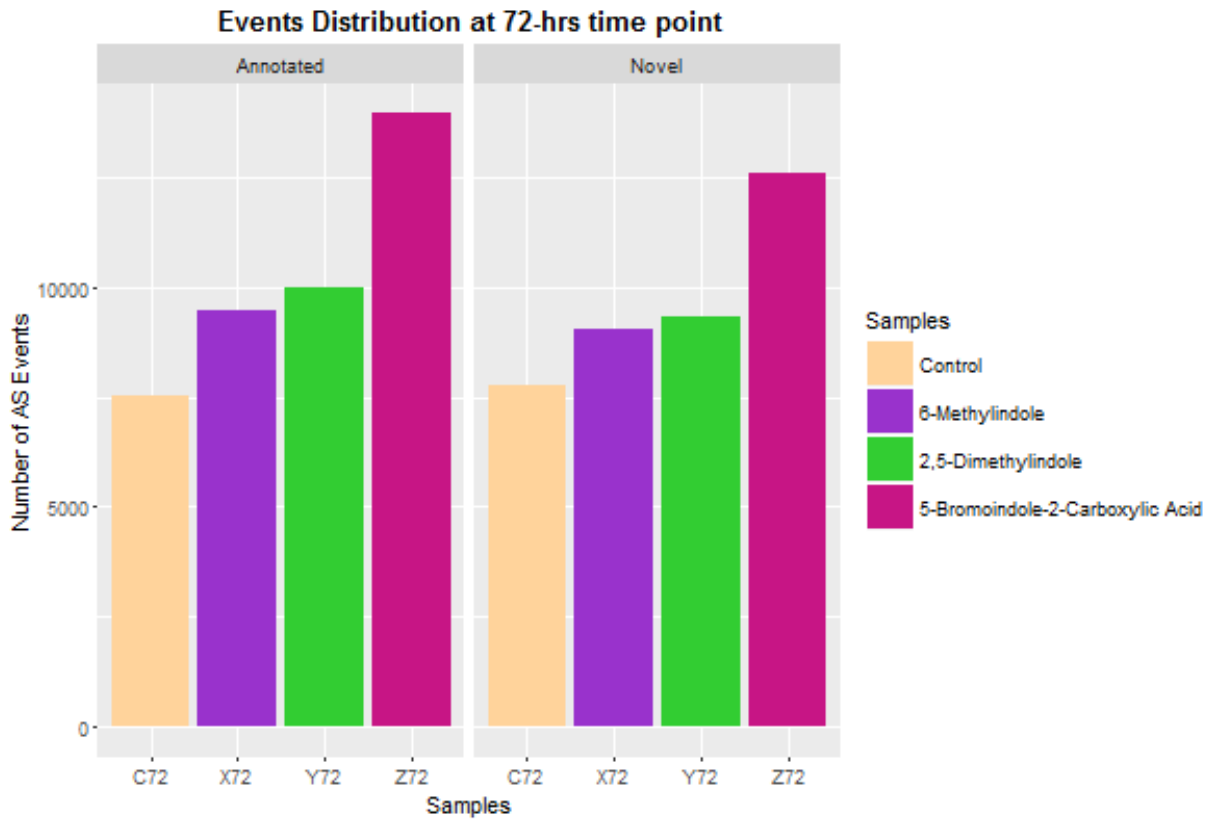




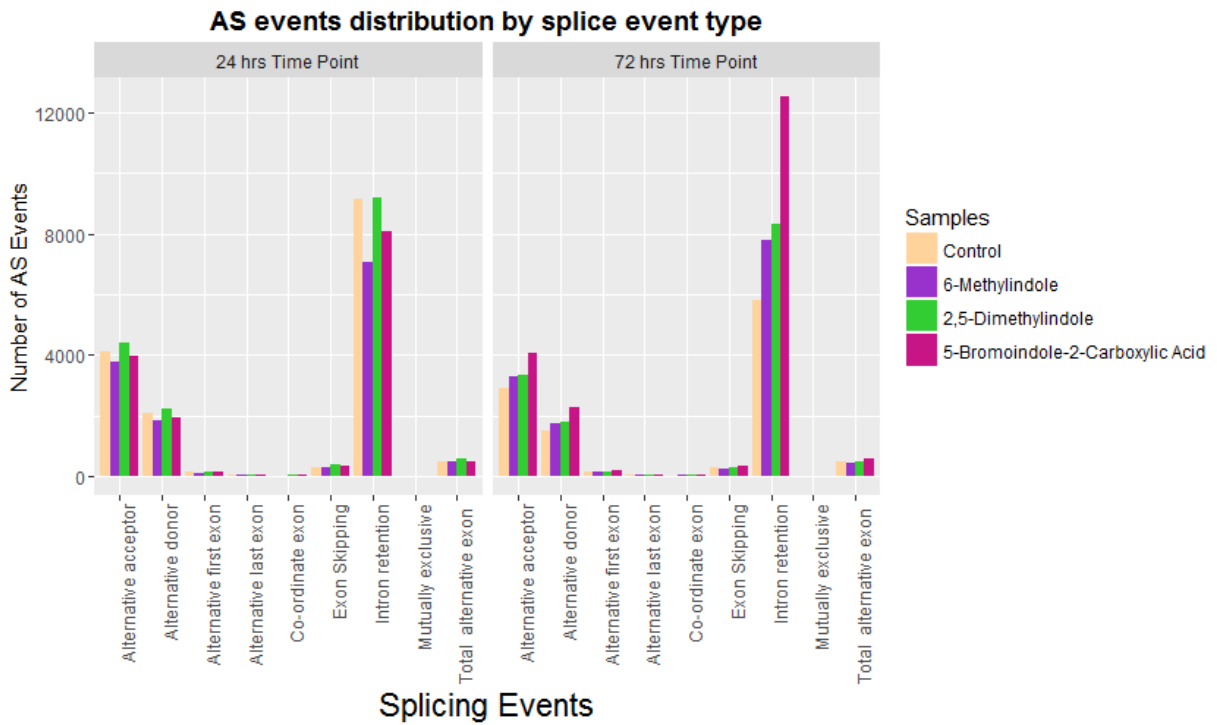
**Fig3.1 Total AS events distribution.** Counting of all splicing events discovered and supported by splice-junction reads from each control and treatment samples at both 24 hrs and 72 hrs time points.



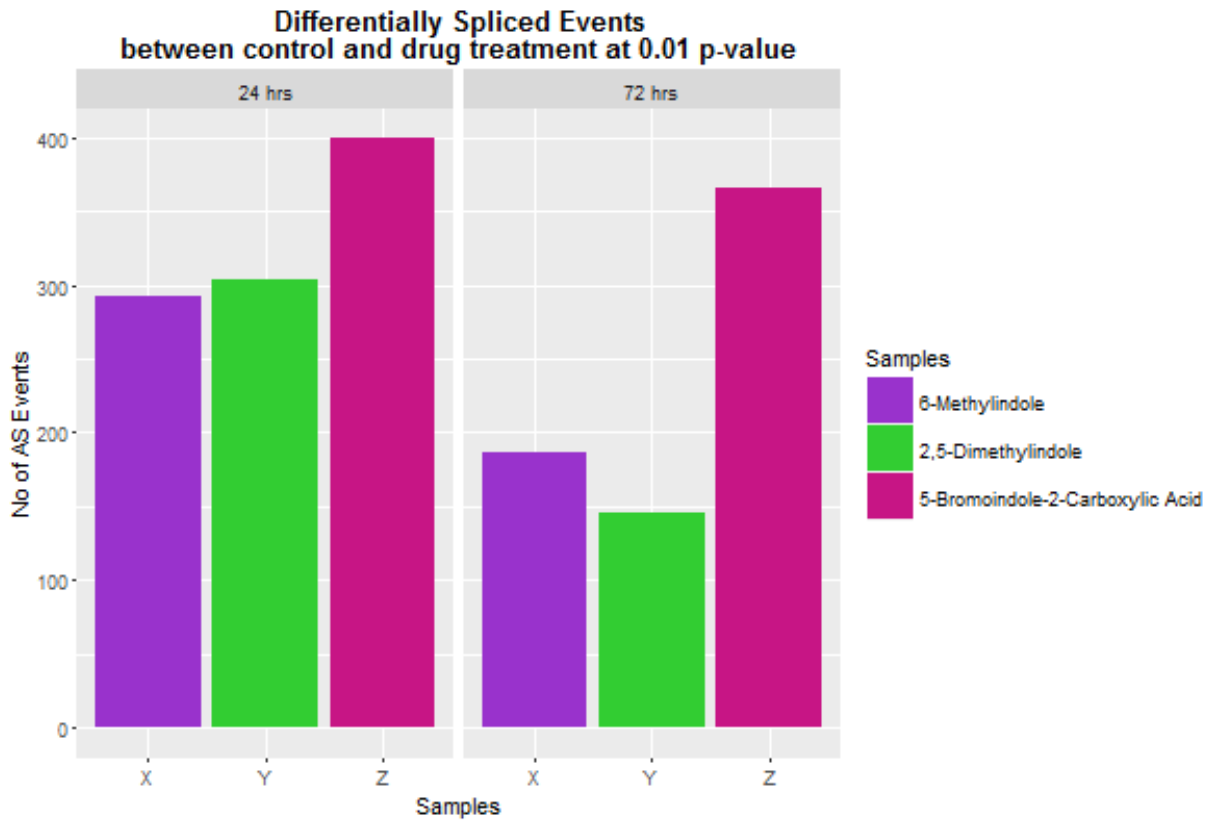
**Fig3.2 Annotated and Novel AS events distribution.** Comparison of annotated and novel events annotated and discovered in each sample at 24 hrs time point.



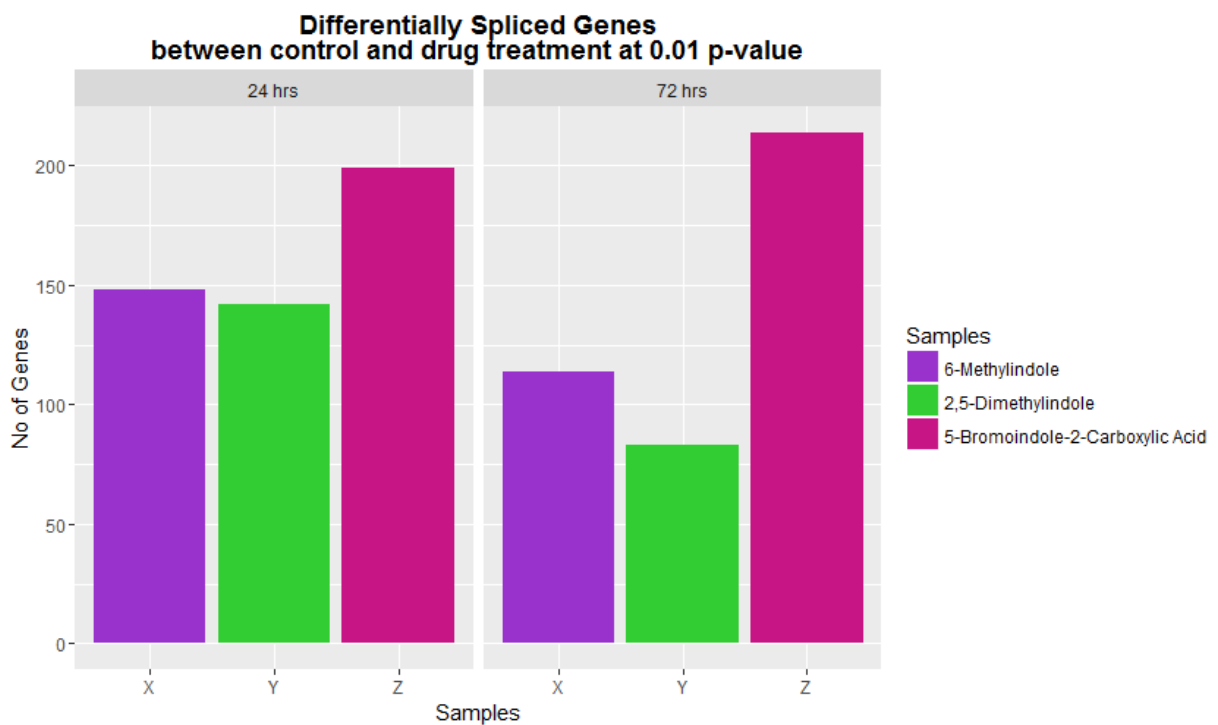
**Fig3.3 Annotated and Novel AS events distribution.** Comparison of annotated and novel events annotated and discovered in each sample at 72 hrs time point.



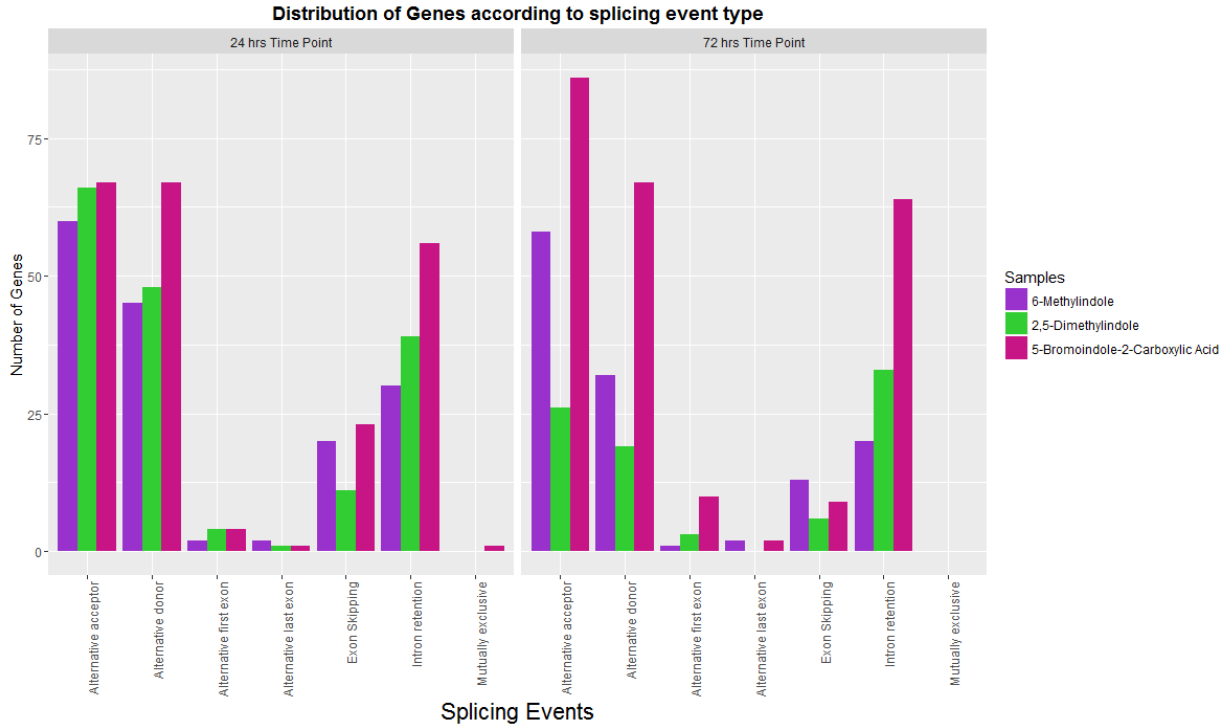
**Fig3.4 Different AS events distribution.** Plots of every splicing event on each sample at two time points. Splicing events plotted from the left to right in each graph are; alternative acceptor or 3' splice site, alternative donor or 5' splice site, alternative first exon, alternative last exon, co-ordinate exon, exon skipping, intron retention, mutually exclusive exons, and total alternative exons.



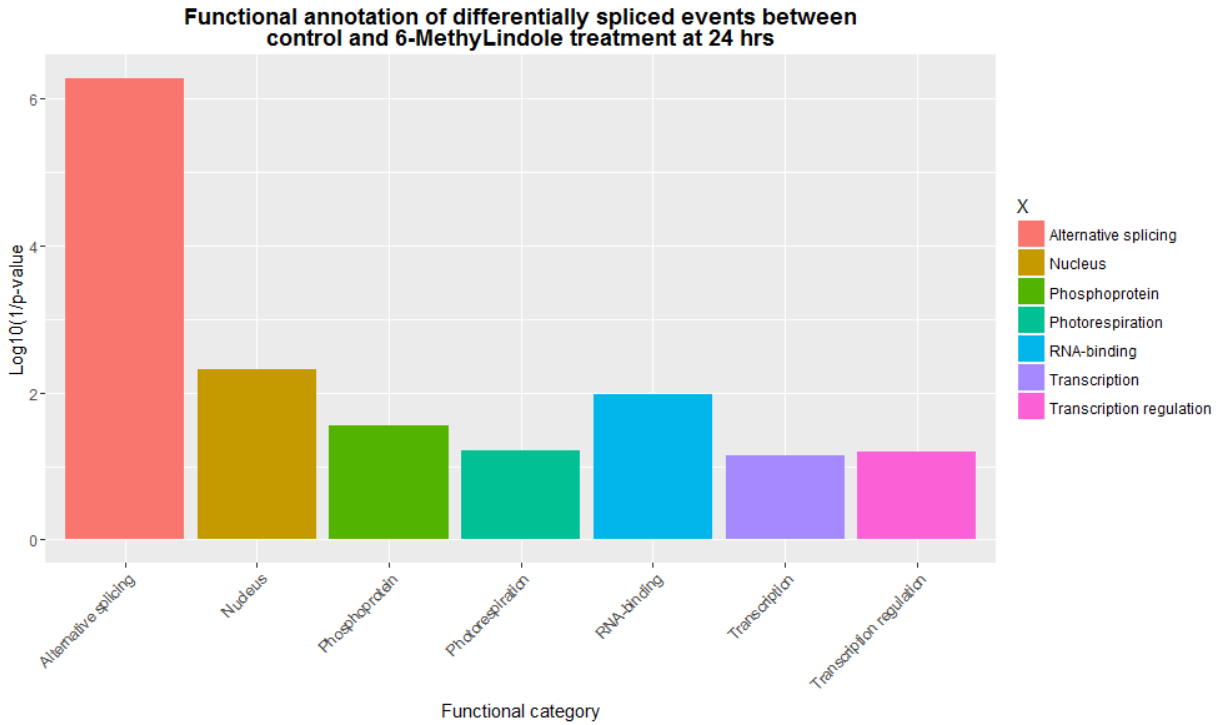
**Fig3.5 Distribution of statistically significant differentially expressed splicing events.** Plot shows the number of events differentially spliced at 0.01 p-value. Paired t-test statistical test used percent spliced in values of each splicing event between control and each treatment at the same time point to infer differential splicing.



**Fig3.6 Distribution of statistically significant differentially spliced genes.** Plot shows the number of genes differentially spliced at 0.01 p-value. Paired t-test statistical test used percent spliced in values of each splicing event between control and each treatment at the same time point to infer differential splicing.

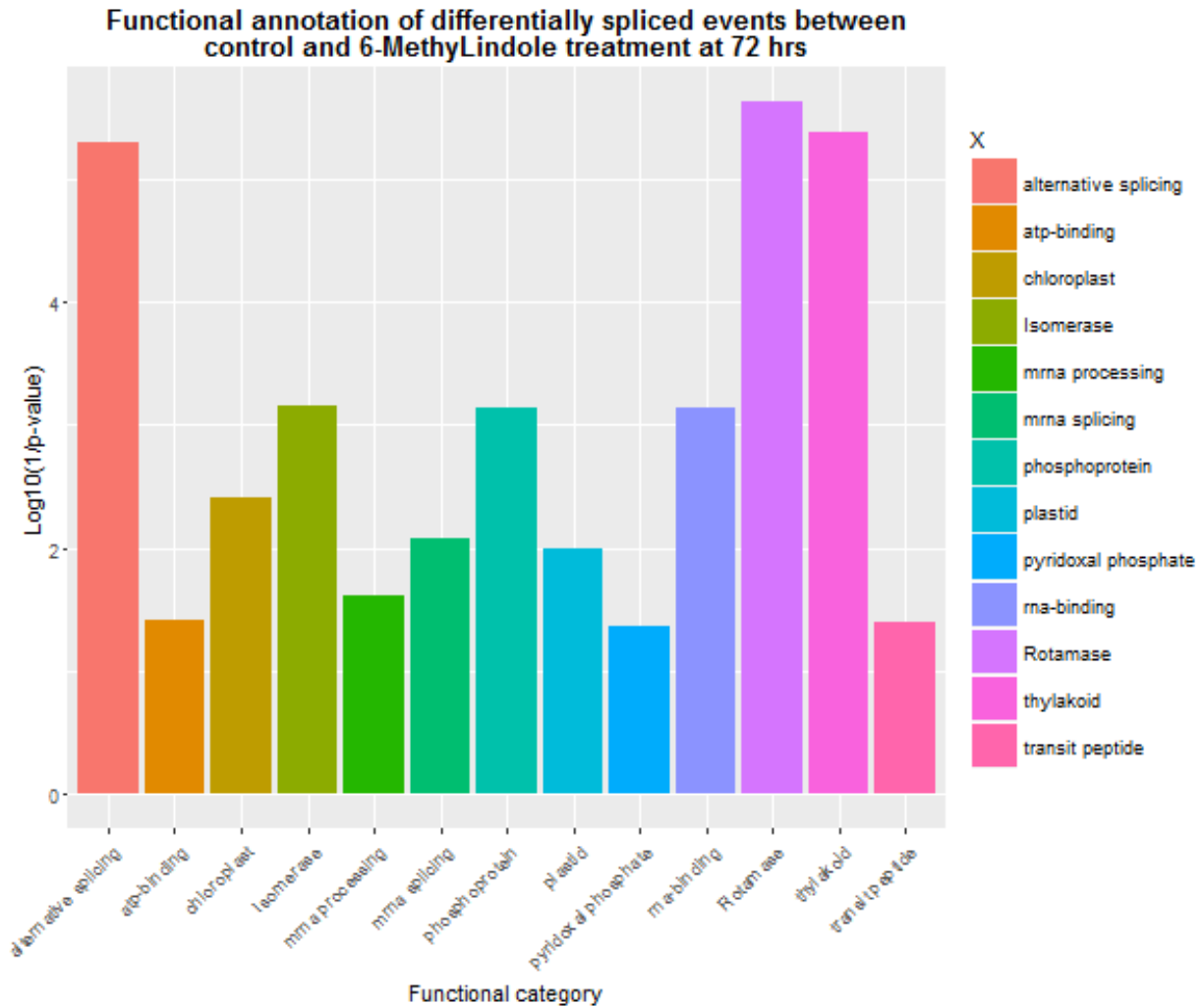


**Fig 3.7 Genes distribution according to event splice type affecting it.** Plots of genes differentially spliced by each splice event type in each sample at two time points. Splicing events plotted from the left to right in each graph are; alternative acceptor or 3' splice site, alternative donor or 5' splice site, alternative first exon, alternative last exon, exon skipping, intron retention and mutually exclusive exons.

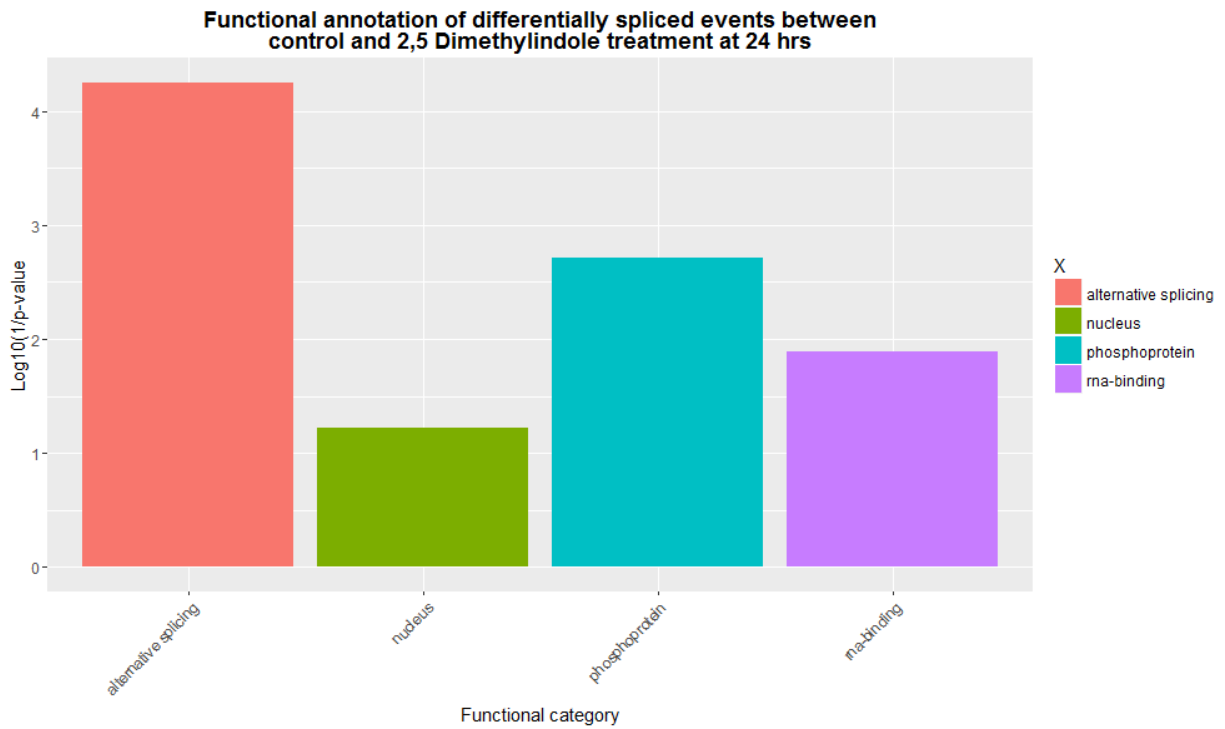


**Fig 3.8 Functional categorization of differentially spliced genes between control and 6-Methylindole treated samples at 24 hrs time point.** Plot contain functional categorization of differentially expressed genes between Control and 6-Methylindole treated samples at 24 hrs time point has been generated by David Functional annotation tool. All categories plotted exceed 1.3 log<sub>10</sub> (1/p-value) cut-off value which is equivalent to 0.05 p-value.

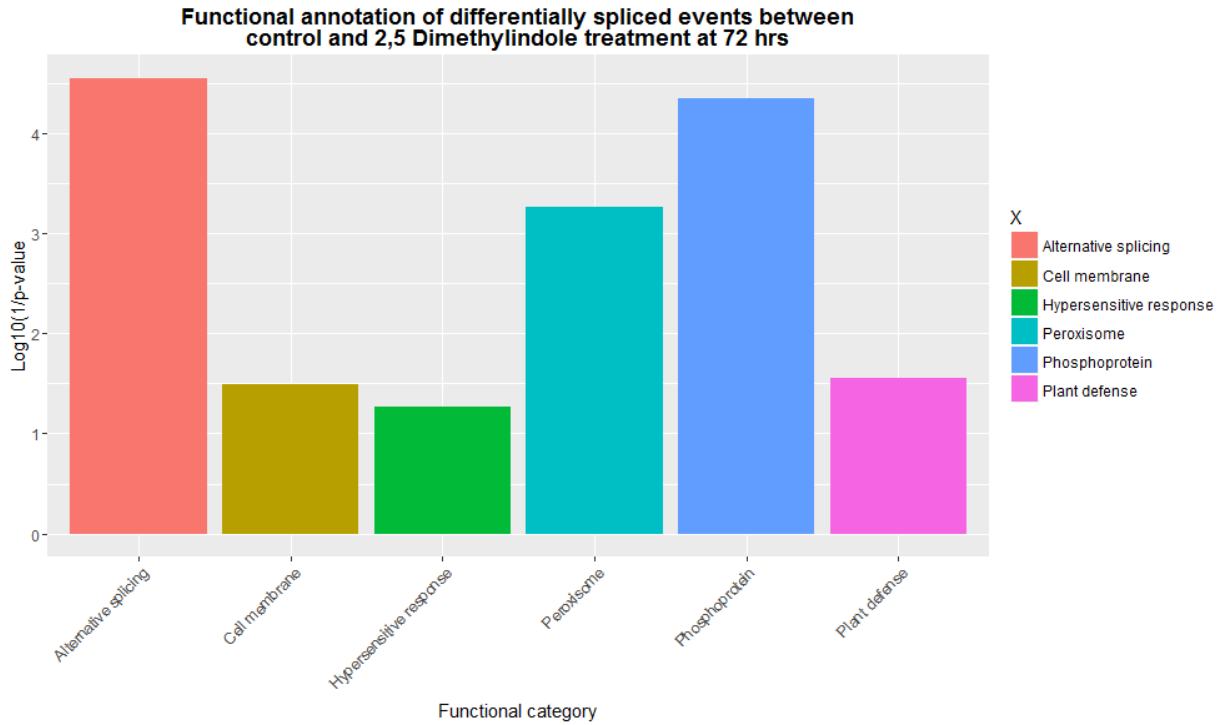




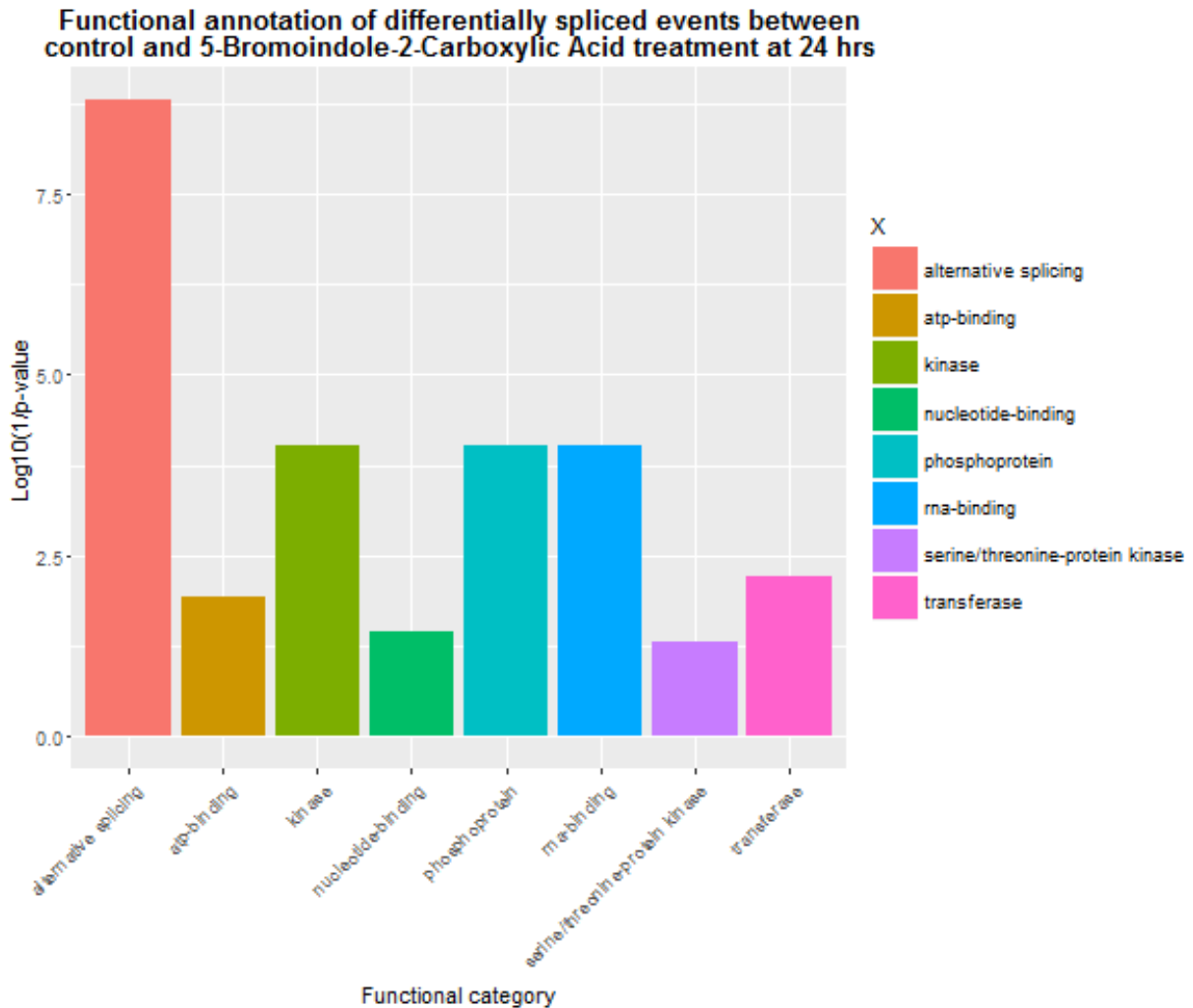
**Fig 3.9 Functional categorization of differentially spliced genes between control and 6-Methylindole treated samples at 72 hrs time point.** Plot contain functional categorization of differentially expressed genes between Control and 6-Methylindole treated samples at 72 hrs time point has been generated by David Functional annotation tool. All categories plotted exceed 1.3 log<sub>10</sub> (1/p-value) cut-off value which is equivalent to 0.05 p-value.



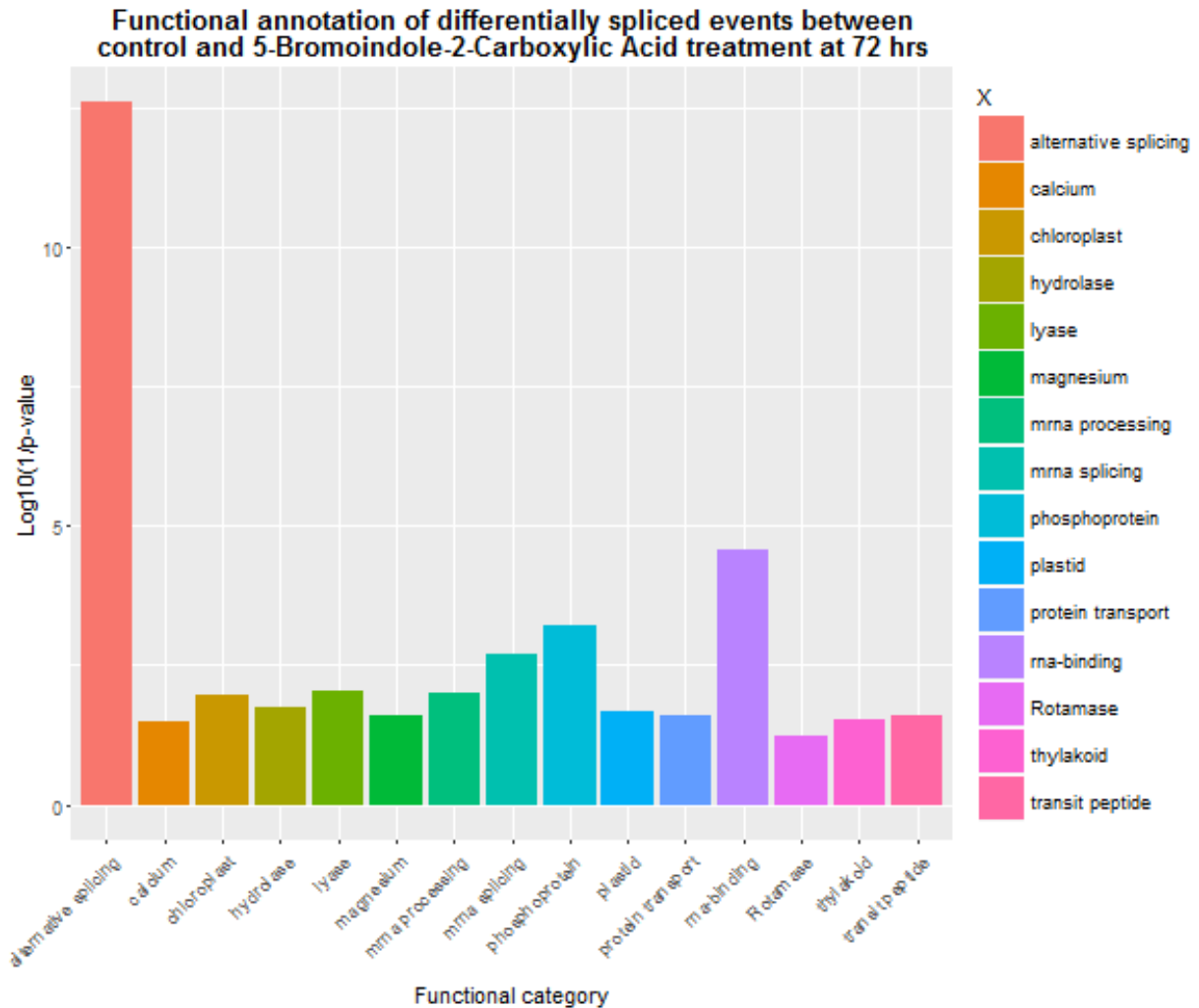
**Fig 3.10 Functional categorization of differentially spliced genes between control and 2, 5-Dimethylindole treated samples at 24 hrs time point.** Plot contain functional categorization of differentially expressed genes between Control and 2,5-Dimethylindole treated samples at 24 hrs time point has been generated by David Functional annotation tool. All categories plotted exceed 1.3 log<sub>10</sub>(1/p-value) cut-off value which is equivalent to 0.05 p-value.



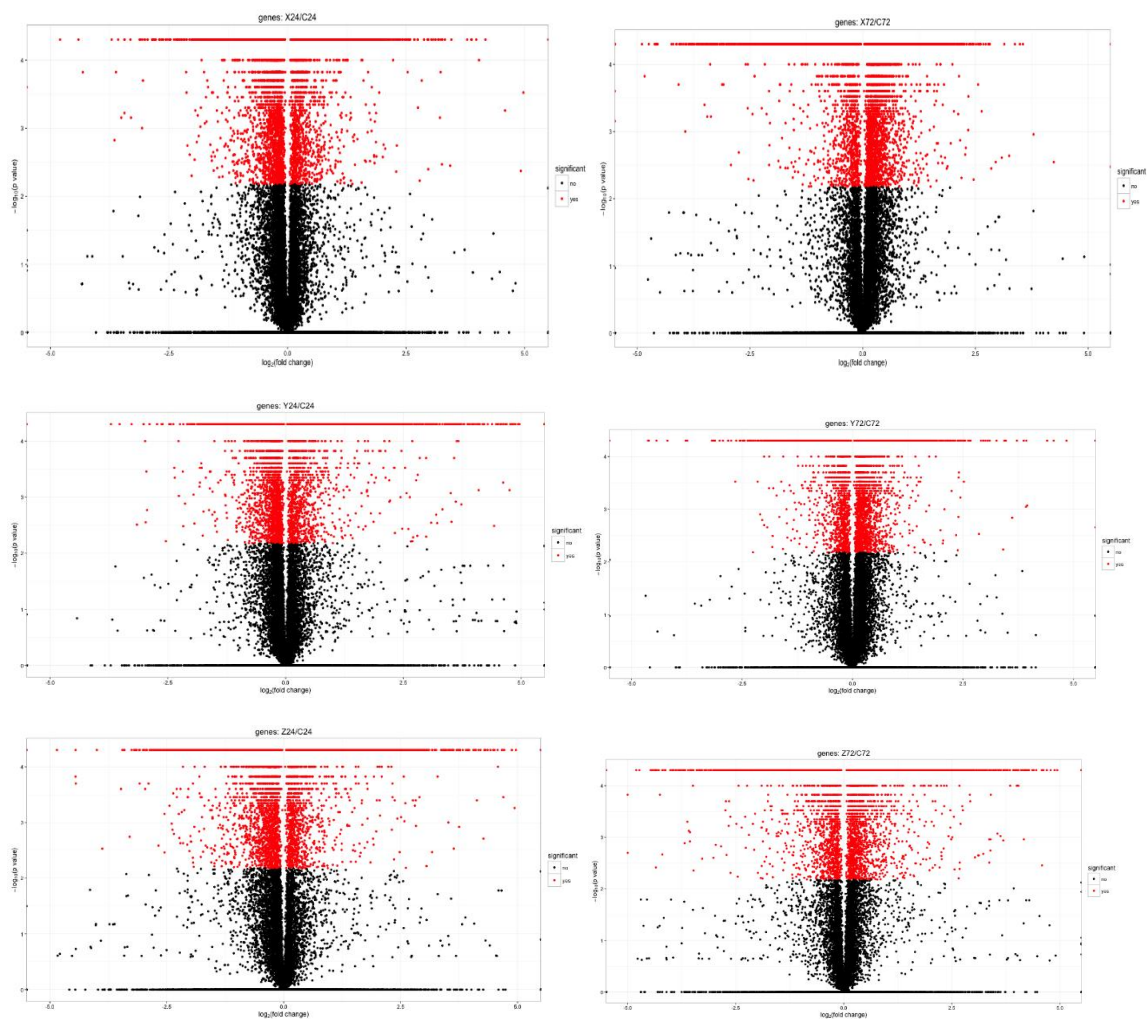
**Fig 3.11 Functional categorization of differentially spliced genes between control and 2, 5-Dimethylindole treated samples at 72 hrs time point.** Plot contain functional categorization of differentially expressed genes between Control and 2,5-Dimethylindole treated samples at 72 hrs time point has been generated by David Functional annotation tool. All categories plotted exceed 1.3 log<sub>10</sub> (1/p-value) cut-off value which is equivalent to 0.05 p-value.



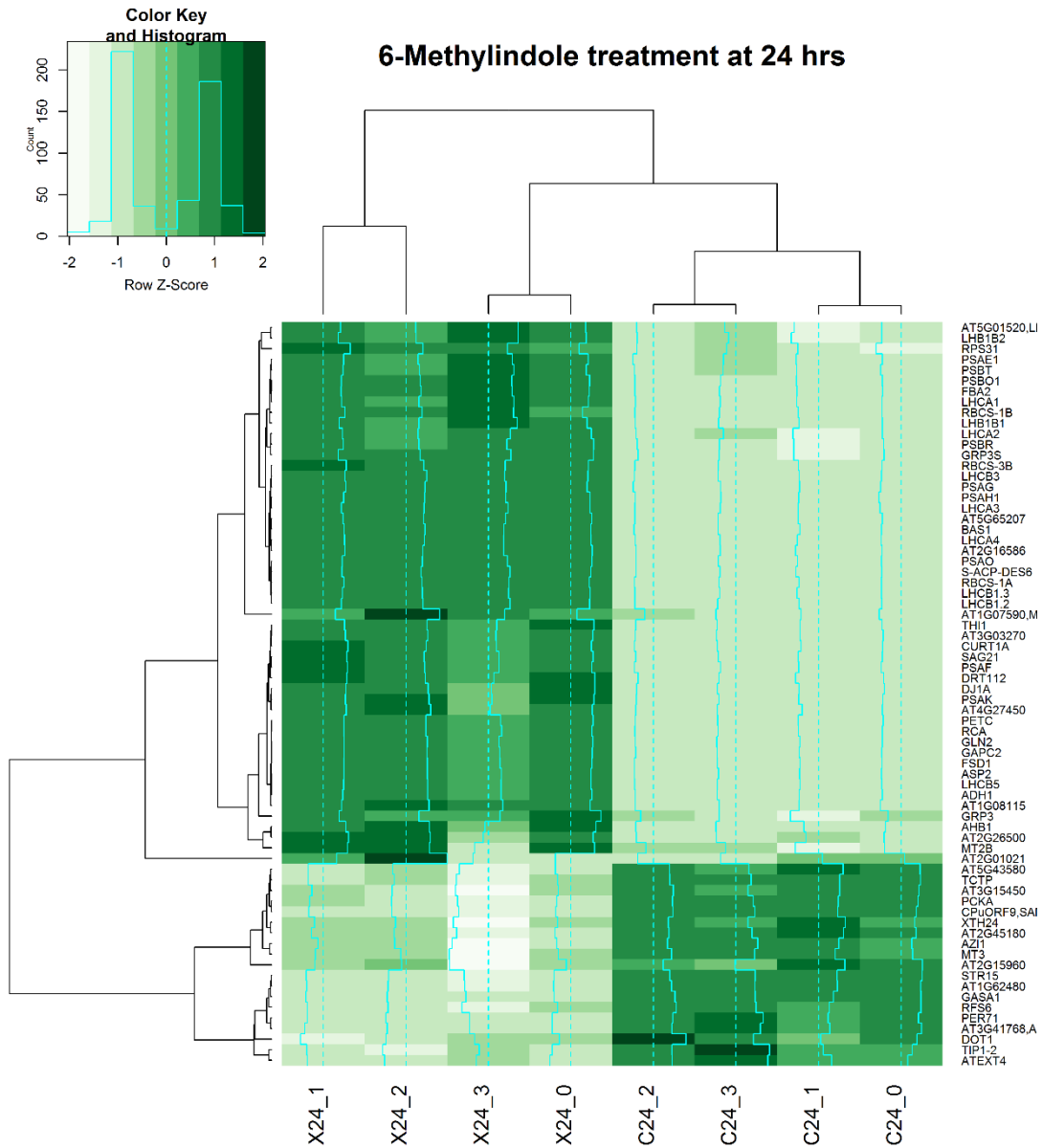
**Fig 3.12 Functional categorization of differentially spliced genes between control and 5-Bromoindole-2-Carboxylic Acid treated samples at 24 hrs time point.** Plot contain functional categorization of differentially expressed genes between Control and 5-Bromoindole-2-Carboxylic Acid treated samples at 24 hrs time point has been generated by David Functional annotation tool. All categories plotted exceed 1.3 log10 (1/p-value) cut-off value which is equivalent to 0.05 p-value.



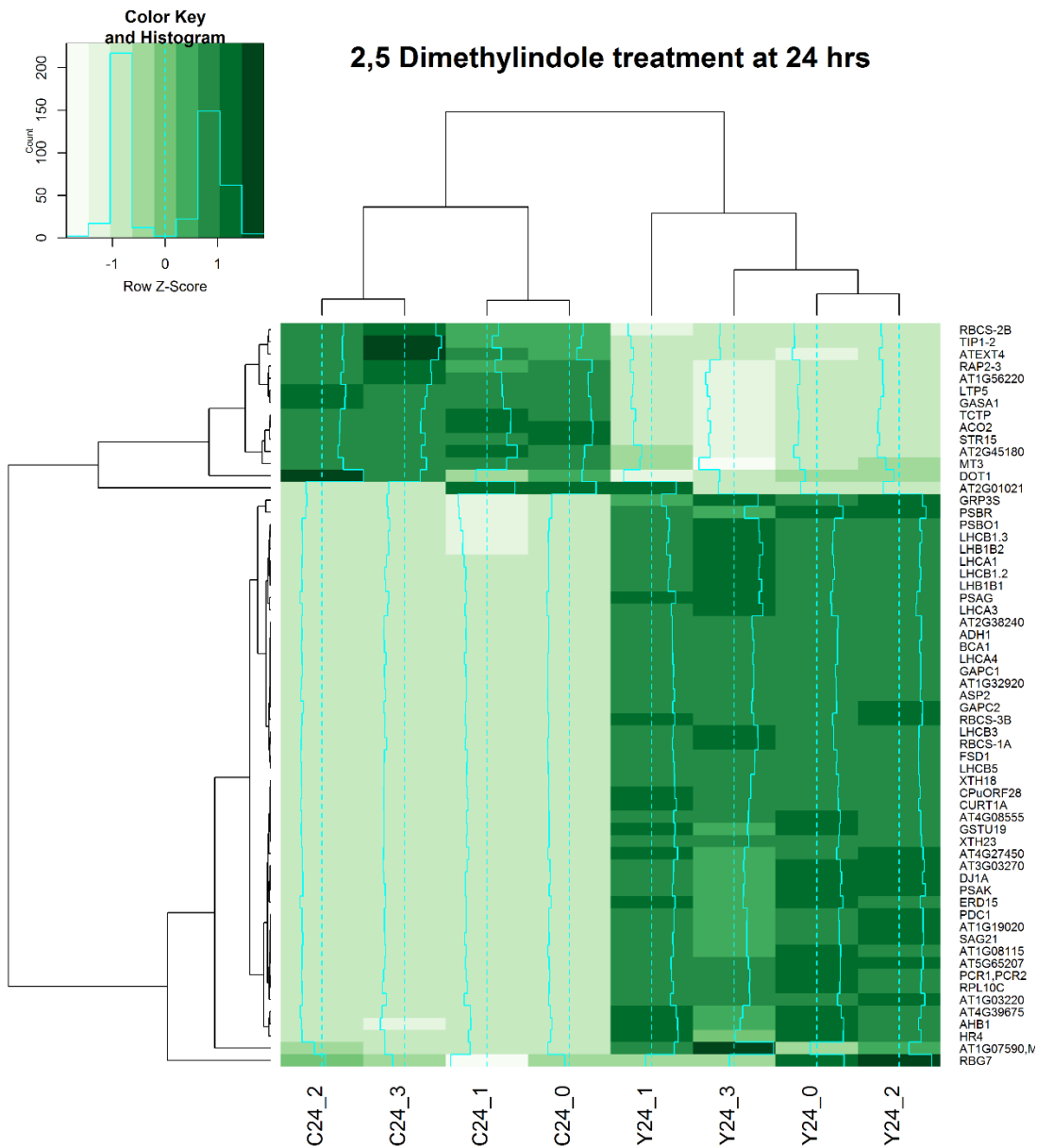
**Fig 3.13 Functional categorization of differentially spliced genes between control and 5-Bromoindole-2-Carboxylic Acid treated samples at 72 hrs time point.** Plot contain functional categorization of differentially expressed genes between Control and 5-Bromoindole-2-Carboxylic Acid treated samples at 72 hrs time point has been generated by David Functional annotation tool. All categories plotted exceed 1.3 log10 (1/p-value) cut-off value which is equivalent to 0.05 p-value.



**Fig 3.14** Volcano plots of expression alterations under (A) 6-Methylindole at 24 hrs time point (B) 6-Methylindole at 72 hrs time point (C) 2, 5-Dimethylindole at 24 hrs time point (D) 2, 5-Dimethylindole at 72 hrs time point (E) 5-Bromoindole-2-Carboxylic Acid at 24 hrs time point (F) 5-Bromoindole-2-Carboxylic Acid at 72 hrs time point.

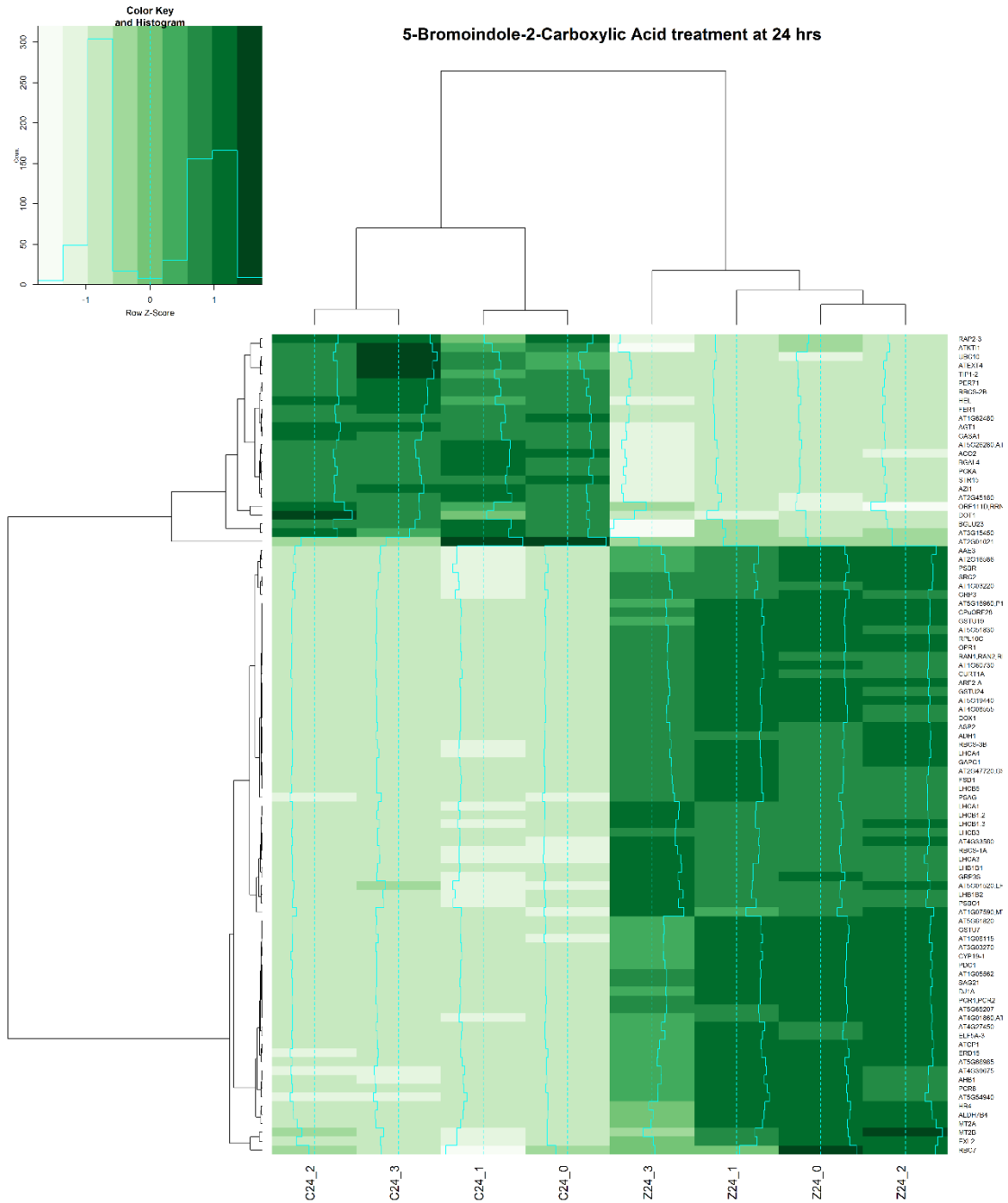


**Fig 3.15 Heatmap of top differentially expressed genes between control and 6-Methylindole treated samples at 24 hrs time point.** Genes plotted are ones which showed expression difference more than or equal to 450 between two samples. Clustering of genes through correlation have been done through hclust function in R statistical programming environment.

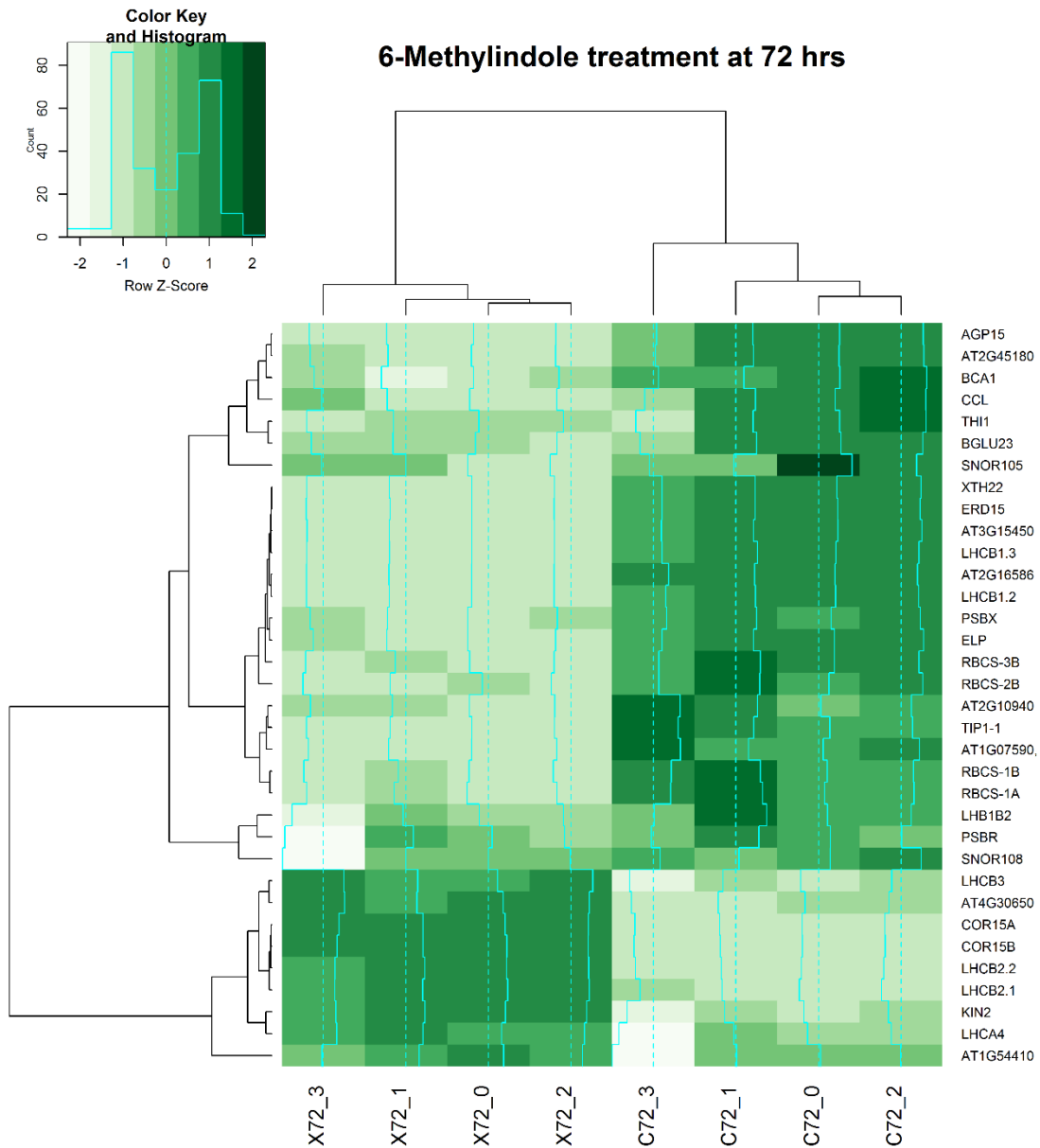


**Fig 3.16 Heatmap of top differentially expressed genes between control and 2, 5-Dimethylindole treated samples at 24 hrs time point.** Genes plotted are ones which showed expression difference more than or equal to 450 between two samples. Clustering of genes through correlation have been done through hclust function in R statistical programming environment.

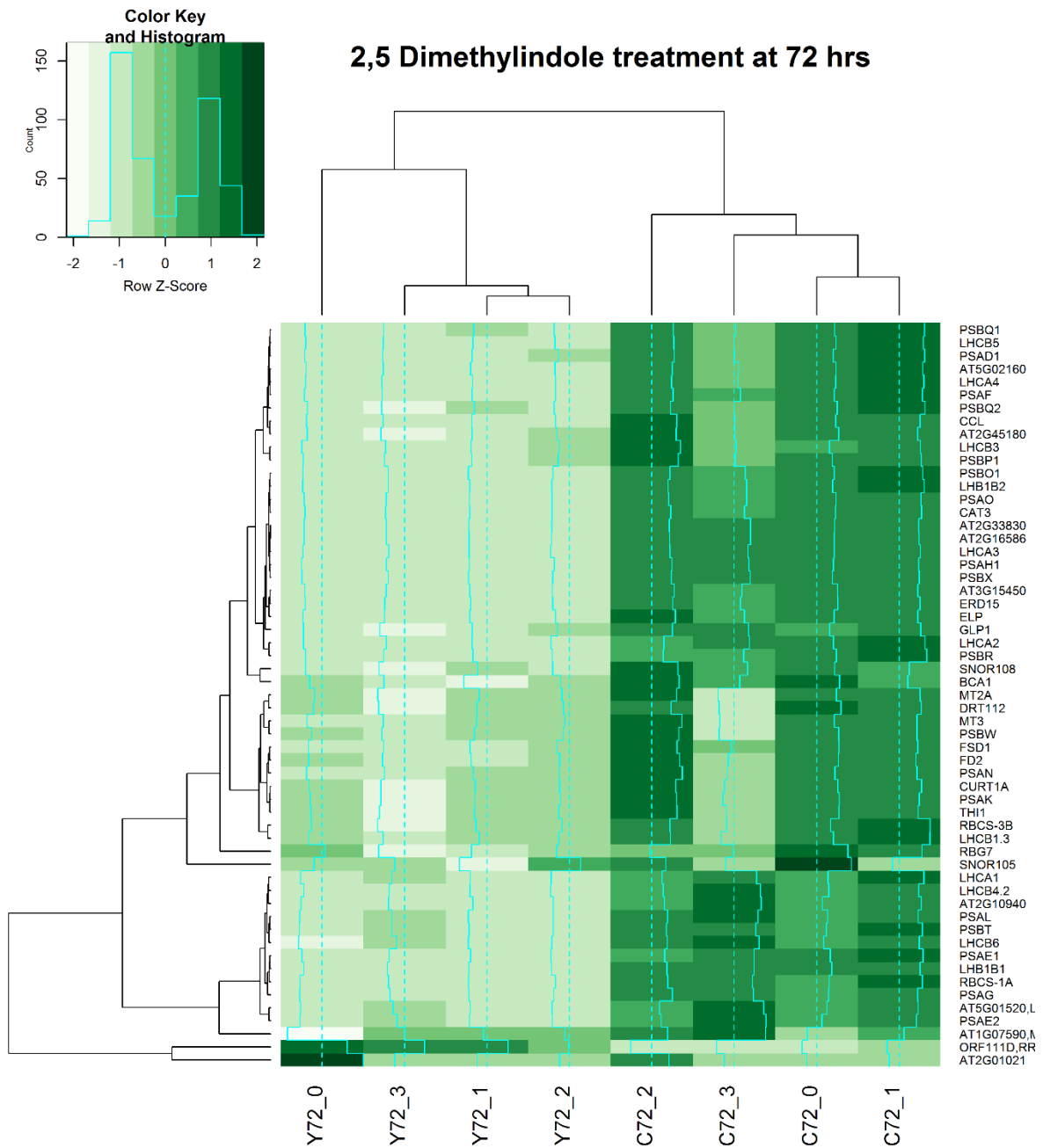




**Fig 3.17 Heatmap of top differentially expressed genes between control and 5-Bromoindole-2-Carboxylic Acid treated samples at 24 hrs time point.** Genes plotted are ones which showed expression difference more than or equal to 450 between two samples. Clustering of genes through correlation have been done through hclust function in R statistical programming environment.

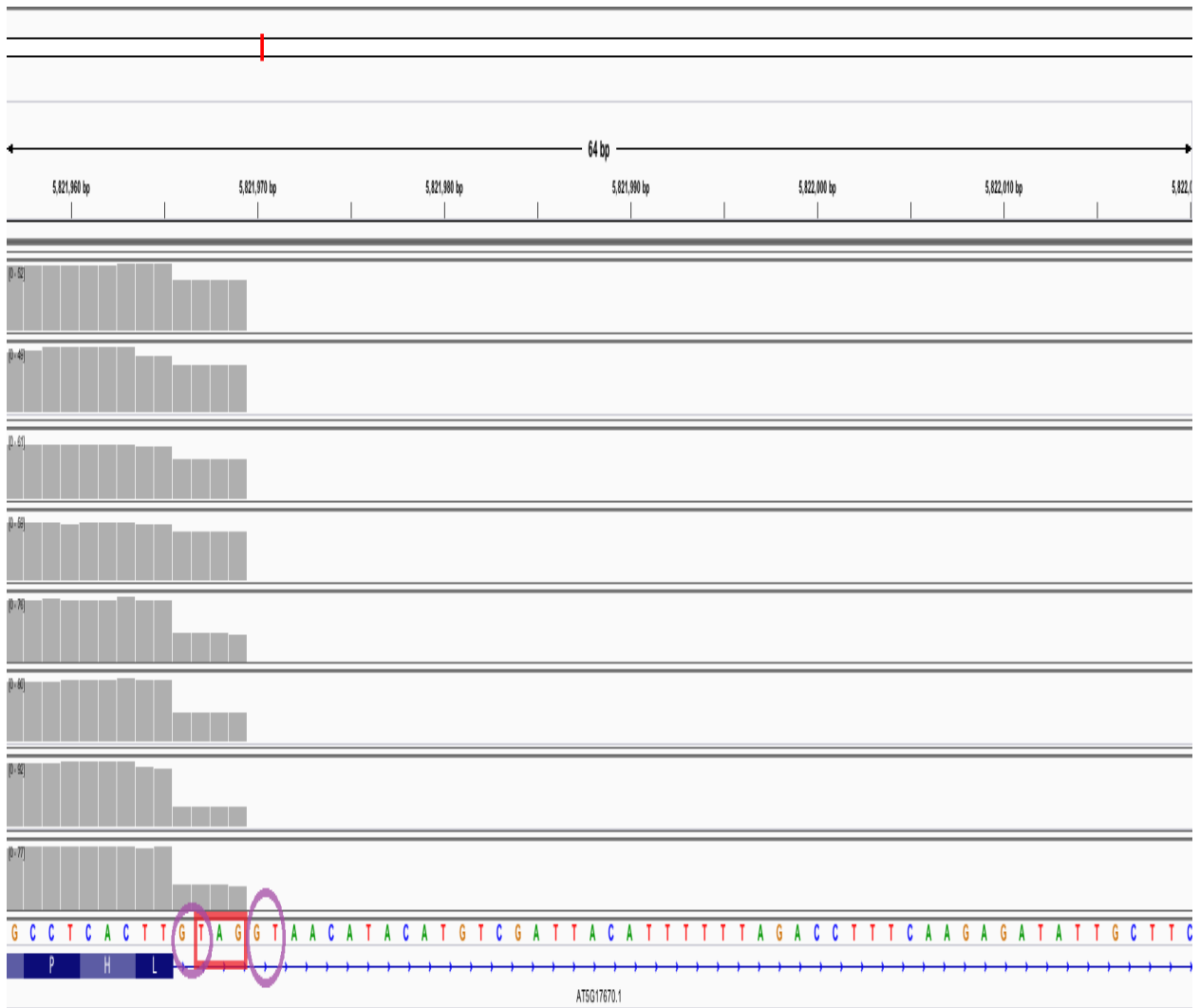


**Fig 3.18. Heatmap of top differentially expressed genes between control and 6-Methylindole treated samples at 72 hrs time point.** Genes plotted are ones which showed expression difference more than or equal to 450 between two samples. Clustering of genes through correlation have been done through hclust function in R statistical programming environment.

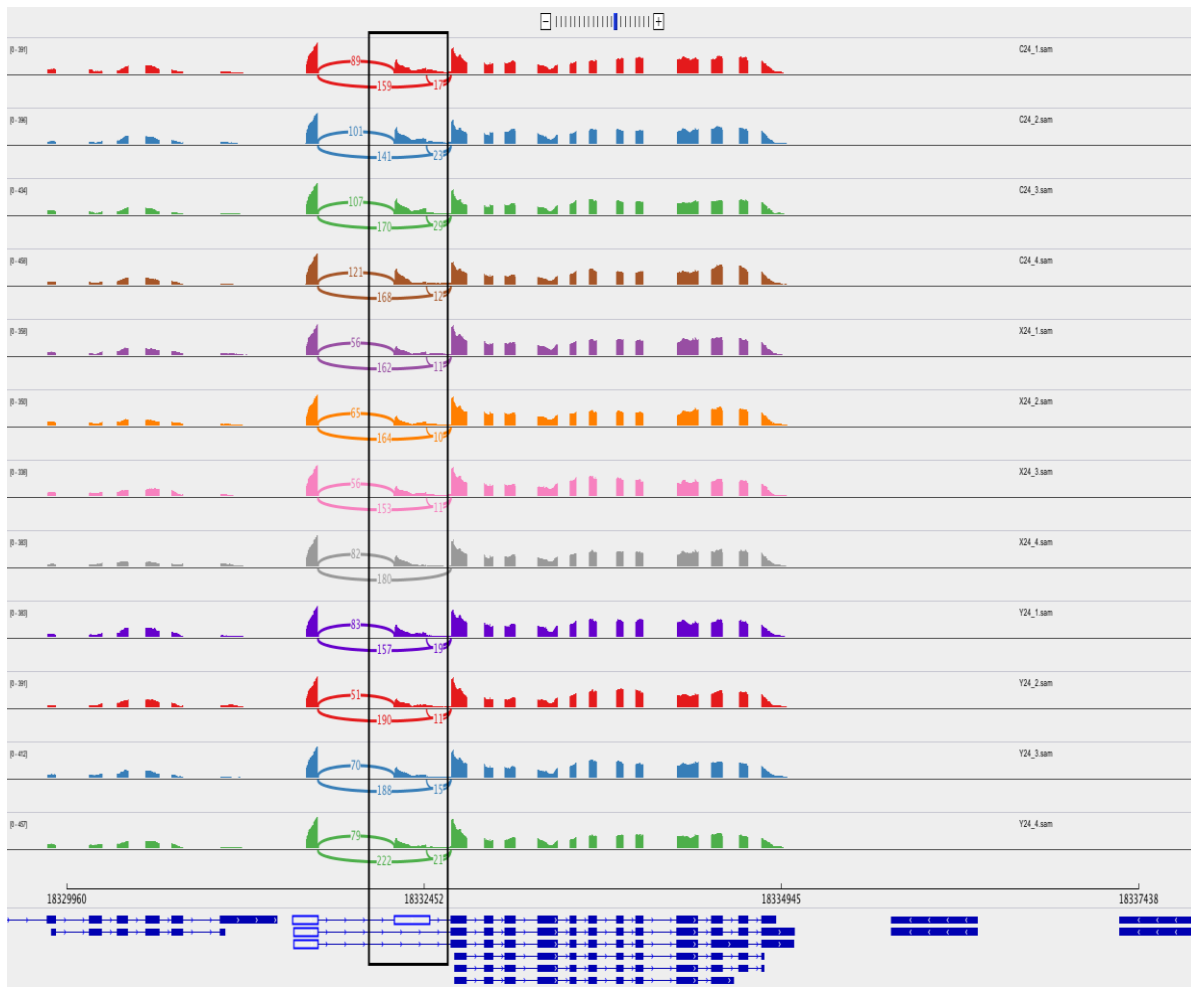


**Fig 3.19. Heatmap of top differentially expressed genes between control and 2, 5-Dimethylindole treated samples at 72 hrs time point.** Genes plotted are ones which showed expression difference more than or equal to 450 between two samples. Clustering of genes through correlation have been done through hclust function in R statistical programming environment.

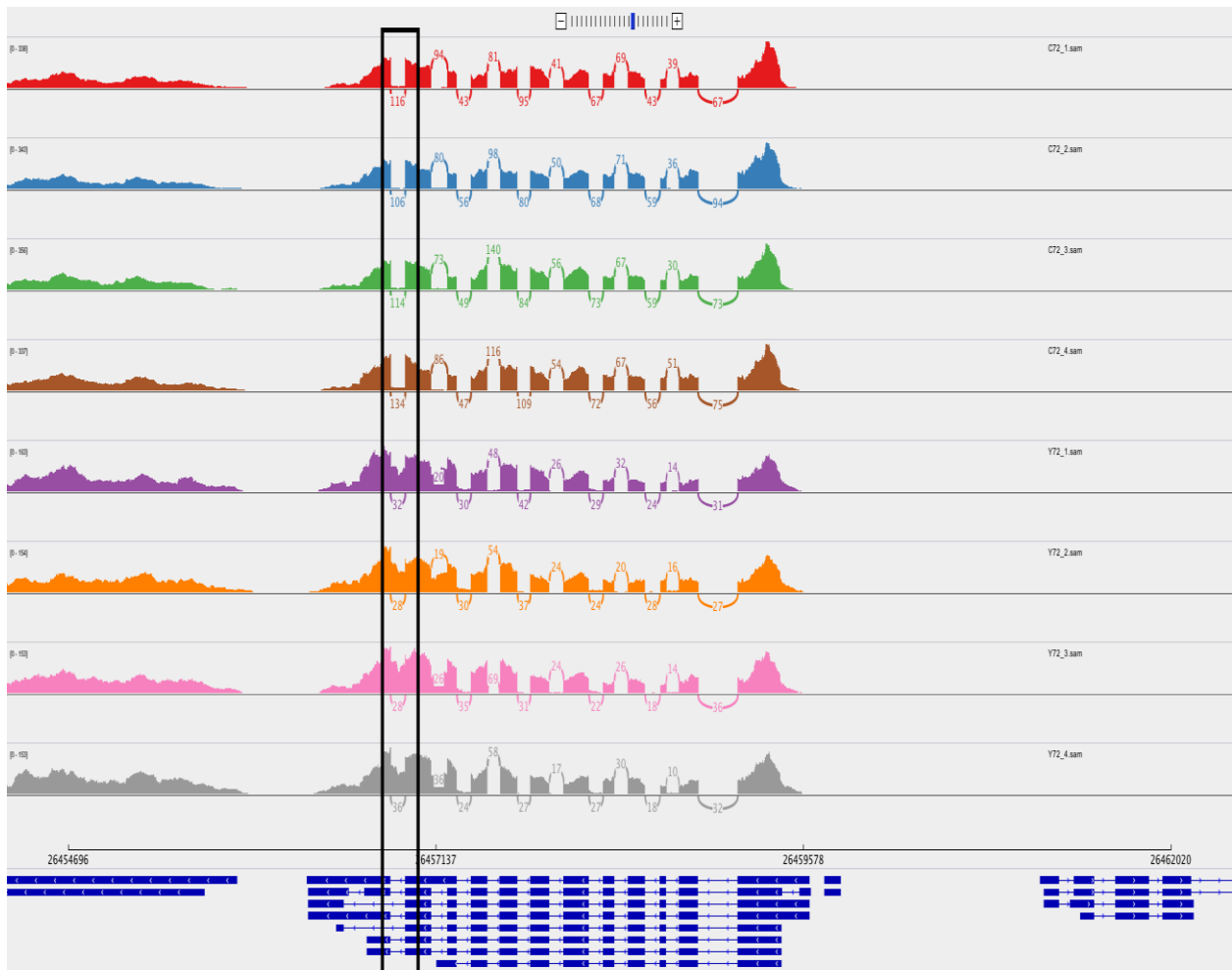




**Fig 3.21 Differential usage of alternative 5' splice site that skips a pre-mature termination codon between control and 2, 5-Dimethylindole of hydrolase-like protein (AT5G17670).** Comparison between differential inclusion of a pre-mature termination codon (highlighted in red square) between control sample (top four panels) 2, 5-Dimethylindole treatment (lower four panels). This differential inclusion is achieved through differential usage of two splice sites. First GT splice site (left purple circle) has already been annotated in gene model (plotted at bottom of the shot). Second splice site GT (right purple circle) is a novel splice site that has been discovered in our data. This 5' alternative splice event has been discovered through juncBASE and differential 5' alternative donor was noted as statistically significant (p-value<0.01).



**Fig 3.22 Differential exon skipping event in control vs. both 6-Methylindole and 2, 5-Dimethylindole treated samples at 24 hrs of SR34A serine arginine rich protein splicing factor.** Comparison between differential inclusion of exon 2 (highlighted in black square) between control sample (top four panels), 6-Methylindole (middle four panels) and 2, 5-Dimethylindole treatment (lower four panels). Less inclusion of exon 2 under the effect of both 6-Methylindole and 2, 5-Dimethylindole than control sample is observed. This exon skipping event has been annotated in genome reference gene model and differential exon skipping was noted as statistically significant (p-value<0.01).



**Fig 3.23 Differential Intron retention event between control and 2,5-Dimethylindole at 72 hrs time point of CPK28 calcium dependent protein kinase.** Comparison between differential inclusion of intron (highlighted in black square) between control sample (top four panels) 2, 5-Dimethylindole treatment (lower four panels). This novel intron retention event has been discovered through juncBASE and differential intron retention was noted as statistically significant (p-value<0.01).

**Table3.1. Differentially Expressed genes between control and treatments at 24 hrs time point**

<b>6-MethyIndole</b>	<b>2,5 Dimethylindole</b>	<b>5-Bromoindole 2-carboxylic acid</b>
LHCB1.3	AT1G03220	AT1G03220
PSAK	AT1G19020	AT1G05562
S-ACP-DES6	LHCB1.3	SRC2
LHCA3	PSAK	GSTU24
AT1G62480	AT1G32920	ARF2-A
ADH1	AT1G56220	LHCB1.3
PSBR	LHCA3	LHCA3
AT1G07590,MT1A	ACO2	ACO2
AT1G08115	RPL10C	AT1G62480
PSAO	ADH1	RPL10C
GAPC2	PSBR	ELF5A-3
DRT112	AT1G07590,MT1A	OPR1
LHCB1.2	AT1G08115	ADH1
PSAF	GAPC2	PSBR
PSAG	PCR1,PCR2	AT1G07590,MT1A
RBCS-1A	LHCB1.2	AT1G08115
GASA1	PSAG	PCR1,PCR2
ATEXT4	RBCS-1A	LHCB1.2
AT2G01021	CPuORF28	PCR8
GRP3S	GASA1	ALDH7B4
GRP3	ATEXT4	PSAG
AT2G15960	GSTU19	AT1G60730
AT2G16586	AT2G01021	RBCS-1A
AT2G26500	GRP3S	CPuORF28
LHB1B1	LHB1B1	ATKTI1
DOT1	DOT1	GASA1
RPS31	ERD15	ATEXT4
AT2G45180	AT2G45180	GSTU19
AHB1	AHB1	AT2G01021
LHB1B2	RBG7	GRP3S
RCA	LHB1B2	GRP3
CPuORF9,SAMDC1	AT2G38240	AT2G16586
BAS1	GAPC1	CYP19-1
DJ1A	DJ1A	LHB1B1
AT3G15450	RAP2-3	DOT1
AT3G41768,AT3G41979	HR4	ERD15



LHCA2	BCA1	AT2G45180
AT3G03270	AT3G03270	AT2G47720,GSTF8
MT3	MT3	AGT1
PSAH1	TCTP	AHB1
TCTP	TIP1-2	RBG7
PSBT	LHCA4	GSTU7
TIP1-2	LTP5	LHB1B2
LHCA4	LHCA1	GAPC1
LHCA1	CURT1A	DJ1A
CURT1A	AT4G08555	AT3G15450
PETC	LHCB5	RAP2-3
LHCB5	XTH23	HR4
STR15	STR15	DOX1
SAG21	AT4G39675	AT3G03270
AZI1	SAG21	HEL
FSD1	FSD1	BGLU23
AT4G27450	AT4G27450	MT2A
PSAE1	XTH18	TIP1-2
XTH24	PDC1	LHCA4
PCKA	ASP2	AAE3
FBA2	LHCB3	LHCA1
AT5G01520,LHCB4.1	PSBO1	CURT1A
ASP2	RBCS-3B	AT4G08555
RFS6	RBCS-2B	LHCB5
GLN2	AT5G65207	AT4G33560
LHCB3		STR15
THI1		AT4G39675
PSBO1		AT4G01860,AT4G01865,AT4G01870
MT2B		SAG21
RBCS-3B		AZI1
RBCS-1B		FSD1
AT5G43580		AT4G27450
PER71		PDC1
AT5G65207		PCKA
		AT5G01520,LHCB4.1
		AT5G19440
		ASP2
		RAN1,RAN2,RPT6B
		AT5G26280,AT5G26300
		ATCP1
		AT5G51830
		LHCB3
		BGAL4

		EXL2 PSBO1 AT5G66985 FER1 MT2B AT5G16960,P1 RBCS-3B RBCS-2B UBC10 AT5G54940 AT5G61820 PER71 AT5G65207 ORF111D,RRN5
--	--	---

**Table3.2. Differentially Expressed genes between control and treatments at 72 hrs time point**

<b>6-MethyIndole</b>	<b>2,5 Dimethylindole</b>	<b>5-Bromoindole 2- carboxylic acid</b>
LHCB1.3	PSBP1	PSB27-1
PSBR	CAT3	PSBP1
AT1G07590,MT1A	LHCB1.3	MT1C
ELP	PSAK	APX1
SNOR105	FD2	FD1
LHCB1.2	LHCA3	GSTU24
AT1G54410	PSBR	LFNR2
RBCS-1A	AT1G07590,MT1A	RPL34
AT2G16586	PSAO	LHCB1.3
TIP1-1	ELP	PSAK
ERD15	LHCB6	PRK
AT2G45180	DRT112	GAPB
LHCB2.2	SNOR105	PSAH2
LHCB2.1	PSAF	FD2
PSBX	PSAG	LHCA3
AT2G10940	RBCS-1A	PSBR
LHB1B2	GLP1	AT1G03140,PSAD2
COR15B	AT2G01021	AT1G07590,MT1A
COR15A	AT2G16586	PSAO
AT3G15450	PSAE2	ELP
CCL	LHB1B1	GAPA2
BCA1	ERD15	LHCB6
BGLU23	AT2G45180	DHAR1
LHCA4	PSBX	DRT112
AT4G30650	AT2G10940	EXT3
AGP15	RBG7	AT1G21500
KIN2	PSBW	SNOR105
SNOR108	AT2G33830	LHCB1.2
LHCB3	LHB1B2	PSAF
THI1	LHCB4.2	AGP21
RBCS-3B	AT3G15450	PSAG
RBCS-2B	CCL	RBCS-1A
RBCS-1B	LHCA2	PSBY
XTH22	BCA1	GLP1

MT2A	GSTU19
MT3	RPS17
PSAH1	AT2G01021
PSBT	PDF2.3
LHCA4	GRP3S
LHCA1	AT2G16586
CURT1A	PSAE2
LHCB5	AT2G26500
PSAL	GSTF9
PSBQ1	RPL28
PSAD1	LHB1B1
PSBQ2	TIP1-1
FSD1	RPS31
PSAE1	ERD15
AT5G01520,LHCB4.1	AT2G45180
AT5G02160	CURT1B
SNOR108	CP12-1
LHCB3	LHCB2.1
THI1	PSBX
PSBO1	AT2G10940
RBCS-3B	AGT1
PSAN	RBG7
ORF111D,RRN5	PSBW
	LHB1B2
	CP29B
	RCA
	COR15A
	GAPC1
	LHCB4.2
	BAS1
	AT3G15450
	TIP2-1
	GAPA1
	RPL12A
	LOX2
	TKL-1
	LHCA2
	CYP20-3
	CP12-2
	BCA1
	BGLU23
	MT2A
	PGK1,SDRB

		PSAH1 PSBT TIP1-2 LHCA4 PSBO2 CP29A RPL17 LHCA1 AT3G63160 CURT1A PETC ATPD LHCB5 PSAL AT4G16980 TUFA PSBQ1 ACP4 PSAD1 PSBQ2 GPX6 PER42 CP31A FSD1 PSAE1 AT4G32260 SSR16 SPE2 SHM1 FBA2 AT5G01520,LHCB4.1 AT5G02160 AGP15 BCA2 KIN2 NUDT24 GLN2 SNOR108 LHCB3 THI1 PSBO1 MT2B GER3
--	--	---

		VSP1,VSP2 RBCS-3B RBCS-2B RBCS-1B AT5G42530 XTH22 LTP4 PSAN
--	--	--

## CHAPTER 4: CONCLUSION AND FUTURE PERSPECTIVES

In this study, 3 different Indole derivatives, 6-Methylindole, 2, 5-Dimethylindole, 5-Bromoindole-2-Carboxylic Acid, have been used to target SR proteins in *Arabidopsis* 6 days old seedlings. Paired end RNA-seq have been used to provide a full snapshot of the plant's transcriptome after treatment at two time points 24 hrs and 72 hrs.

Interestingly, observation of global splicing pattern alteration under the influence of Indole derivatives used in this study indicate that in most treatments, 4 out of 6, and specifically after certain onset of time, 72 hrs, splicing inhibition promoted by those derivatives lead to significant enhancement of AS, mostly and specifically intron retention splicing events. This finding is strongly consistent with the hypothesis that ESE dependent splicing inhibition promoted by Indole derivative eventually result in splicing failure and more retained intron, especially in an organism where intron retention is extremely prevalent.

Identification of both genes differentially spliced and expressed revealed similarity in the affected systems and coupling of AS to gene expression. Most importantly, Photorespiration related genes were frequently and repeatedly altered, on AS or expression level, under different Indole derivatives treatments and time points. More interestingly RNA binding proteins and splicing factors were differentially spliced, a finding that is consistent with the prevalence of auto-regulatory nature of those genes.

Furthermore, insights into the molecular function of events showed differential splicing under treatment suggest various potential functional outcomes. In hydrolase-like protein (AT5G17670), novel alternative 5' splice site skipped a pre-mature termination codon. In SR34a SR splicing factor, exon skipping event at the 5' untranslated region may render the spliced transcript resistant to regulatory pathways target this region similar to what happens in metazoans. In CPK32 protein kinase, novel intron retention event, which function is unknown, may suggest an AS mediated alteration of kinase activity. This may further indicate strong coupling between post-transcriptional splicing and post-translation phosphorylation.

Future directions of this research line shall include two main research lines. First, systematic structure activity relationship of Indole derivatives determination shall be conducted. More Indole derivatives, beyond those used in this study may be experimented on in-vitro splicing system and on different SR proteins shall give more insights into the specificity exhibited by different derivatives. Those studies are critically important for the development of Indole derivatives as splicing modulators in both therapeutic and research context.

Moreover, more splicing modulators that target different components of splicesomal machinery, rather than SR proteins, may be used to reveal more into the consequences of splicing modulation on Arabidopsis (or other plants)'s transcriptome. Alteration of AS through AOS against specific splicing event may be further used. This modulation of AS by targeting different splicing players shall continue to clarify more into the nature of landscape AS regulation mediated of gene expression, post-transcriptional and post-translational modifications.



## References

1. Moore, M.J., *From birth to death: the complex lives of eukaryotic mRNAs*. Science, 2005. **309**(5740): p. 1514-1518.
2. Cheng, Y. and X. Chen, *Posttranscriptional control of plant development*. Current opinion in plant biology, 2004. **7**(1): p. 20-25.
3. Simpson, G.G., *The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of Arabidopsis flowering time*. Current opinion in plant biology, 2004. **7**(5): p. 570-574.
4. Staiger, D., et al., *Emerging role for RNA-based regulation in plant immunity*. New Phytologist, 2013. **197**(2): p. 394-404.
5. Carvalho, R.F., C.V. Feijão, and P. Duque, *On the physiological significance of alternative splicing events in higher plants*. Protoplasma, 2013. **250**(3): p. 639-650.
6. Staiger, D. and R. Green, *RNA-based regulation in the plant circadian clock*. Trends in plant science, 2011. **16**(10): p. 517-523.
7. Lorković, Z.J., *Role of plant RNA-binding proteins in development, stress response and genome organization*. Trends in plant science, 2009. **14**(4): p. 229-236.
8. Rataj, K. and G.G. Simpson, *Message ends: RNA 3' processing and flowering time control*. J Exp Bot, 2014. **65**(2): p. 353-63.
9. Sharp, P.A., *Split genes and RNA splicing (Nobel lecture)*. Angewandte Chemie International Edition in English, 1994. **33**(12): p. 1229-1240.
10. Gilbert, W., *Why genes in pieces?* Nature, 1978. **271**(5645): p. 501.
11. Kornblihtt, A.R., et al., *Alternative splicing: a pivotal step between eukaryotic transcription and translation*. Nat Rev Mol Cell Biol, 2013. **14**(3): p. 153-165.
12. Black, D.L., *Mechanisms of alternative pre-messenger RNA splicing*. Annual review of biochemistry, 2003. **72**(1): p. 291-336.
13. Smith, C.W., J.G. Patton, and B. Nadal-Ginard, *Alternative splicing in the control of gene expression*. Annual review of genetics, 1989. **23**(1): p. 527-577.
14. Braunschweig, U., et al., *Dynamic integration of splicing within gene regulatory pathways*. Cell, 2013. **152**.
15. Stamm, S., et al., *Function of alternative splicing*. Gene, 2005. **344**: p. 1-20.
16. Bhuvanagiri, M., et al., *NMD: RNA biology meets human genetic medicine*. Biochemical Journal, 2010. **430**(3): p. 365-377.
17. Mühlemann, O. and J. Lykke-Andersen, *How and where are nonsense mRNAs degraded in mammalian cells?* RNA biology, 2010. **7**(1): p. 28-32.
18. Reddy, A.S., *Alternative splicing of pre-messenger RNAs in plants in the genomic era*. Annu. Rev. Plant Biol., 2007. **58**: p. 267-294.
19. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes*. Nature, 2008. **456**(7221): p. 470-476.
20. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing*. Nature Genet., 2008. **40**: p. 1413-1415.

21. Zhang, G., et al., *Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome*. *Genome research*, 2010. **20**(5): p. 646-654.
22. Filichkin, S.A., et al., *Genome-wide mapping of alternative splicing in Arabidopsis thaliana*. *Genome Res*, 2010. **20**.
23. Marquez, Y., et al., *Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis*. *Genome Res.*, 2012. **22**: p. 1184-1279.
24. Rühl, C., et al., *Polypyrimidine tract binding protein homologs from Arabidopsis are key regulators of alternative splicing with implications in fundamental developmental processes*. *The Plant Cell*, 2012. **24**(11): p. 4360-4375.
25. Zhang, X.C. and W. Gassmann, *Alternative splicing and mRNA levels of the disease resistance gene RPS4 are induced during defense responses*. *Plant Physiol.*, 2007. **145**: p. 1577-1587.
26. Kazan, K., *Alternative splicing and proteome diversity in plants: the tip of the iceberg has just emerged*. *Trends in plant science*, 2003. **8**(10): p. 468-471.
27. Staiger, D. and J.W. Brown, *Alternative splicing at the intersection of biological timing, development, and stress responses*. *Plant Cell*, 2013. **25**.
28. Dinesh-Kumar, S. and B.J. Baker, *Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance*. *Proceedings of the National Academy of Sciences*, 2000. **97**(4): p. 1908-1913.
29. Filichkin, S.A., et al., *Genome-wide mapping of alternative splicing in Arabidopsis thaliana*. *Genome research*, 2010. **20**(1): p. 45-58.
30. Ali, G. and A. Reddy, *Regulation of alternative splicing of pre-mRNAs by stresses*, in *Nuclear pre-mRNA Processing in Plants*. 2008, Springer. p. 257-275.
31. Fu, X.-D. and M. Ares Jr, *Context-dependent control of alternative splicing by RNA-binding proteins*. *Nat Rev Genet*, 2014. **15**(10): p. 689-701.
32. Matera, A.G. and Z. Wang, *A day in the life of the spliceosome*. *Nat Rev Mol Cell Biol*, 2014. **15**(2): p. 108-121.
33. Will, C.L. and R. Lührmann, *Spliceosome structure and function*. *Cold Spring Harbor perspectives in biology*, 2011. **3**(7): p. a003707.
34. Wahl, M.C., C.L. Will, and R. Lührmann, *The spliceosome: design principles of a dynamic RNP machine*. *Cell*, 2009. **136**: p. 701-718.
35. Chen, W. and M.J. Moore, *Spliceosomes*. *Current Biology*, 2015. **25**(5): p. R181-R183.
36. Tharun, S., *Roles of eukaryotic Lsm proteins in the regulation of mRNA function*. *International review of cell and molecular biology*, 2008. **272**: p. 149-189.
37. Cordin, O. and J.D. Beggs, *RNA helicases in splicing*. *RNA Biology*, 2013. **10**(1): p. 83-95.
38. Brown, J., G. Feix, and D. Frendewey, *Accurate in vitro splicing of two pre-mRNA plant introns in a HeLa cell nuclear extract*. *The EMBO journal*, 1986. **5**(11): p. 2749.
39. Barta, A., et al., *The expression of a nopaline synthase—human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue*. *Plant Molecular Biology*, 1986. **6**(5): p. 347-357.

40. Wiebauer, K., J.-J. Herrero, and W. Filipowicz, *Nuclear pre-mRNA processing in plants: distinct modes of 3'-splice-site selection in plants and animals*. Molecular and cellular biology, 1988. **8**(5): p. 2042-2051.
41. Iwata, H. and O. Gotoh, *Comparative analysis of information contents relevant to recognition of introns in many species*. BMC genomics, 2011. **12**(1): p. 45.
42. Brown, J., et al., *Splicing signals and factors in plant intron removal*. Biochemical Society Transactions, 2002. **30**(2): p. 146-149.
43. Wang, B.-B. and V. Brendel, *The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing*. Genome Biol, 2004. **5**(12): p. R102.
44. Koncz, C., et al., *The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator*. Frontiers in plant science, 2012. **3**: p. 9.
45. Nilsen, T.W. and B.R. Graveley, *Expansion of the eukaryotic proteome by alternative splicing*. Nature, 2010. **463**(7280): p. 457-463.
46. Nicholson, P. and O. Mühlemann, *Cutting the nonsense: the degradation of PTC-containing mRNAs*. Biochemical Society Transactions, 2010. **38**(6): p. 1615-1620.
47. Isken, O. and L.E. Maquat, *The multiple lives of NMD factors: balancing roles in gene and genome regulation*. Nature reviews Genetics, 2008. **9**(9): p. 699-712.
48. Maquat, L.E., *Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics*. Nature reviews Molecular cell biology, 2004. **5**(2): p. 89-99.
49. Hentze, M.W. and A.E. Kulozik, *A perfect message: RNA surveillance and nonsense-mediated decay*. Cell, 1999. **96**(3): p. 307-310.
50. Arciga-Reyes, L., et al., *UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in Arabidopsis*. The plant journal, 2006. **47**(3): p. 480-489.
51. McGlincy, N.J. and C.W. Smith, *Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?* Trends in biochemical sciences, 2008. **33**(8): p. 385-393.
52. Chen, M. and J.L. Manley, *Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches*. Nat Rev Mol Cell Biol, 2009. **10**.
53. Barash, Y., *Deciphering the splicing code*. Nature, 2010. **465**: p. 53-59.
54. Matlin, A.J., F. Clark, and C.W.J. Smith, *Understanding alternative splicing: towards a cellular code*. Nat Rev Mol Cell Biol, 2005. **6**(5): p. 386-398.
55. Barta, A., M. Kalyna, and A.S. Reddy, *Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR proteins) in plants*. Plant Cell, 2010. **22**.
56. Lopato, S., et al., *Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors*. Proceedings of the National Academy of Sciences, 1996. **93**(7): p. 3074-3079.
57. Reddy, A.S., et al., *Localization and dynamics of plant splicing regulators*. Plant Physiology, 2011: p. pp. 111.186700.
58. Kalyna, M. and A. Barta, *A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions?* Biochemical Society Transactions, 2004. **32**(4): p. 561-564.

59. Kalyna, M., S. Lopato, and A. Barta, *Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development*. Molecular biology of the cell, 2003. **14**(9): p. 3565-3577.
60. Lopato, S., et al., *atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes*. Genes & development, 1999. **13**(8): p. 987-1001.
61. Lazar, G. and H.M. Goodman, *The Arabidopsis splicing factor SR1 is regulated by alternative splicing*. Plant molecular biology, 2000. **42**(4): p. 571-581.
62. Lopato, S., E. Waigmann, and A. Barta, *Characterization of a novel arginine/serine-rich splicing factor in Arabidopsis*. The Plant Cell, 1996. **8**(12): p. 2255-2264.
63. Palusa, S.G. and A.S. Reddy, *Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay*. New Phytologist, 2010. **185**(1): p. 83-89.
64. Golovkin, M. and A.S. Reddy, *The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins*. The Plant Cell, 1998. **10**(10): p. 1637-1647.
65. Iida, K. and M. Go, *Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants*. Molecular biology and evolution, 2006. **23**(5): p. 1085-1094.
66. Petrillo, E., et al., *A chloroplast retrograde signal regulates nuclear alternative splicing*. Science, 2014. **344**(6182): p. 427-430.
67. Martinez-Contreras, R., *hnRNP proteins and splicing control*. Adv. Exp. Med. Biol., 2007. **623**: p. 123-147.
68. Boutz, P.L., et al., *A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons*. Genes & development, 2007. **21**(13): p. 1636-1652.
69. Izquierdo, J.M., et al., *Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition*. Molecular cell, 2005. **19**(4): p. 475-484.
70. Stauffer, E., et al., *Polypyrimidine tract-binding protein homologues from Arabidopsis underlie regulatory circuits based on alternative splicing and downstream control*. The Plant Journal, 2010. **64**(2): p. 243-255.
71. Simpson, C.G., et al., *Arabidopsis PTB1 and PTB2 proteins negatively regulate splicing of a mini-exon splicing reporter and affect alternative splicing of endogenous genes differentially*. New Phytologist, 2014. **203**(2): p. 424-436.
72. Staiger, D., et al., *The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA*. The Plant Journal, 2003. **33**(2): p. 361-371.
73. Ciuzan, O., et al., *The evolutionarily conserved multifunctional glycine-rich RNA-binding proteins play key roles in development and stress adaptation*. Physiologia plantarum, 2015. **153**(1): p. 1-11.
74. Mangeon, A., R.M. Junqueira, and G. Sachetto-Martins, *Functional diversity of the plant glycine-rich proteins superfamily*. Plant signaling & behavior, 2010. **5**(2): p. 99-104.

75. Carpenter, C.D., J.A. Kreps, and A.E. Simon, *Genes encoding glycine-rich Arabidopsis thaliana proteins with RNA-binding motifs are influenced by cold treatment and an endogenous circadian rhythm*. Plant Physiology, 1994. **104**(3): p. 1015-1025.
76. Schöning, J.C., et al., *Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in Arabidopsis*. Nucleic acids research, 2008. **36**(22): p. 6977-6987.
77. Iida, K., et al., *Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences*. Nucleic Acids Res, 2004. **32**.
78. Christensen, A.H., R.A. Sharrock, and P.H. Quail, *Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation*. Plant molecular biology, 1992. **18**(4): p. 675-689.
79. Larkin, P.D. and W.D. Park, *Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism*. Plant molecular biology, 1999. **40**(4): p. 719-727.
80. Ding, F., et al., *Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis*. BMC Genomics, 2014. **15**(1): p. 1-14.
81. von Koskull-Döring, P., K.-D. Scharf, and L. Nover, *The diversity of plant heat stress transcription factors*. Trends in plant science, 2007. **12**(10): p. 452-457.
82. Sugio, A., et al., *The cytosolic protein response as a subcomponent of the wider heat shock response in Arabidopsis*. The Plant Cell, 2009. **21**(2): p. 642-654.
83. Liu, J., et al., *An autoregulatory loop controlling Arabidopsis HsfA2 expression: role of heat shock-induced alternative splicing*. Plant physiology, 2013. **162**(1): p. 512-521.
84. Palusa, S.G., G.S. Ali, and A.S.N. Reddy, *Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses*. Plant Journal, 2007. **49**.
85. Tanabe, N., et al., *Differential expression of alternatively spliced mRNAs of Arabidopsis SR protein homologs, atSR30 and atSR45a, in response to environmental stress*. Plant Cell Physiol, 2007. **48**(7): p. 1036-49.
86. Gullledge, A.A., et al., *Mining Arabidopsis thaliana RNA-seq data with Integrated Genome Browser reveals stress-induced alternative splicing of the putative splicing regulator SR45a*. American Journal of Botany, 2012. **99**(2): p. 219-231.
87. Mastrangelo, A.M., et al., *Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity*. Plant Sci, 2012. **185–186**.
88. Leviatan, N., et al., *Genome-wide survey of cold stress regulated alternative splicing in Arabidopsis thaliana with tiling microarray*. PLoS One, 2013. **8**(6): p. e66511.
89. Duque, P., *A role for SR proteins in plant stress responses*. Plant signaling & behavior, 2011. **6**(1): p. 49-54.

90. Ali, G. and A. Reddy, *Spatiotemporal organization of pre-mRNA splicing proteins in plants*, in *Nuclear pre-mRNA Processing in Plants*. 2008, Springer. p. 103-118.
91. van Bentem, S.d.l.F., et al., *Phosphoproteomics reveals extensive in vivo phosphorylation of Arabidopsis proteins involved in RNA metabolism*. *Nucleic Acids Research*, 2006. **34**(11): p. 3267-3278.
92. Marquez, Y., et al., *Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis*. *Genome Res*, 2012. **22**.
93. Miller, M.J., et al., *Quantitative Proteomics Reveals Factors Regulating RNA Biology as Dynamic Targets of Stress-induced SUMOylation in Arabidopsis*. *Molecular & Cellular Proteomics*, 2013. **12**(2): p. 449-463.
94. Meyer, M., et al., *Deciphering 3' ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing*. *Molecular cell*, 2011. **43**(6): p. 1033-1039.
95. Eperon, L.P., et al., *Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase?* *Cell*, 1988. **54**: p. 393-401.
96. DRAPER, D.E., *A guide to ions and RNA structure*. *RNA*, 2004. **10**(3): p. 335-343.
97. Mullen, M.A., S.M. Assmann, and P.C. Bevilacqua, *Toward a digital gene response: RNA G-quadruplexes with fewer quartets fold with higher cooperativity*. *Journal of the American Chemical Society*, 2011. **134**(2): p. 812-815.
98. Mills, J.D. and M. Janitz, *Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases*. *Neurobiology of aging*, 2012. **33**(5): p. 1012. e11-1012. e24.
99. Yap, K. and E.V. Makeyev, *Regulation of gene expression in mammalian nervous system through alternative pre-mRNA splicing coupled with RNA quality control mechanisms*. *Molecular and Cellular Neuroscience*, 2013. **56**: p. 420-428.
100. Poulos, M.G., et al., *Developments in RNA splicing and disease*. *Cold Spring Harbor perspectives in biology*, 2011. **3**(1): p. a000778.
101. Mittendorf, K.F., et al., *Tailoring of membrane proteins by alternative splicing of pre-mRNA*. *Biochemistry*, 2012. **51**(28): p. 5541-5556.
102. Fu, R.-H., et al., *Aberrant alternative splicing events in Parkinson's disease*. *Cell transplantation*, 2013. **22**(4): p. 653-661.
103. Pearn, J., *Classification of spinal muscular atrophies*. *The Lancet*, 1980. **315**(8174): p. 919-922.
104. Kaida, D., T. Schneider-Poetsch, and M. Yoshida, *Splicing in oncogenesis and tumor suppression*. *Cancer science*, 2012. **103**(9): p. 1611-1616.
105. Hagen, R.M. and M.R. Ladomery, *Role of splice variants in the metastatic progression of prostate cancer*. *Biochemical Society Transactions*, 2012. **40**(4): p. 870-874.
106. Sampath, J. and L.M. Pelus, *Alternative splice variants of survivin as potential targets in cancer*. *Current drug discovery technologies*, 2007. **4**(3): p. 174-191.
107. Miura, K., W. Fujibuchi, and I. Sasaki, *Alternative pre-mRNA splicing in digestive tract malignancy*. *Cancer science*, 2011. **102**(2): p. 309-316.

108. Singh, R.K. and T.A. Cooper, *Pre-mRNA splicing in disease and therapeutics*. Trends in molecular medicine, 2012. **18**(8): p. 472-482.
109. Spitali, P. and A. Aartsma-Rus, *Splice modulating therapies for human disease*. Cell, 2012. **148**(6): p. 1085-1088.
110. Douglas, A.G. and M.J. Wood, *RNA splicing: disease and therapy*. Briefings in functional genomics, 2011. **10**(3): p. 151-164.
111. Passini, M.A., et al., *Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy*. Science translational medicine, 2011. **3**(72): p. 72ra18-72ra18.
112. Hua, Y., et al., *Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model*. Nature, 2011. **478**(7367): p. 123-126.
113. O'Brien, K., et al., *The biflavonoid isoginkgetin is a general inhibitor of Pre-mRNA splicing*. Journal of Biological Chemistry, 2008. **283**(48): p. 33147-33154.
114. Kaida, D., et al., *Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA*. Nature chemical biology, 2007. **3**(9): p. 576-583.
115. Hasegawa, M., et al., *Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product*. ACS chemical biology, 2011. **6**(3): p. 229-233.
116. Kotake, Y., et al., *Splicing factor SF3b as a target of the antitumor natural product pladienolide*. Nature chemical biology, 2007. **3**(9): p. 570-575.
117. Mizui, Y., et al., *Pladienolides, new substances from culture of Streptomyces platensis Mer-11107 III. In vitro and In vivo antitumor activities*. The Journal of antibiotics, 2004. **57**(3): p. 188-196.
118. Nakajima, H., et al., *New antitumor substances, FR901463, FR901464 and FR901465. II. Activities against experimental tumors in mice and mechanism of action*. The Journal of antibiotics, 1996. **49**(12): p. 1204-1211.
119. Yoon, S.-O., et al., *Isoginkgetin inhibits tumor cell invasion by regulating phosphatidylinositol 3-kinase/Akt-dependent matrix metalloproteinase-9 expression*. Molecular cancer therapeutics, 2006. **5**(11): p. 2666-2675.
120. Corriero, A., B. Miñana, and J. Valcárcel, *Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A*. Genes & development, 2011. **25**(5): p. 445-459.
121. Ngo, J.C.K., et al., *Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2*. Molecular cell, 2005. **20**(1): p. 77-89.
122. Prasad, J. and J.L. Manley, *Regulation and substrate specificity of the SR protein kinase Clk/Sty*. Molecular and cellular biology, 2003. **23**(12): p. 4139-4149.
123. Chang, J.-G., et al., *Small molecule amiloride modulates oncogenic RNA alternative splicing to devitalize human cancer cells*. PloS one, 2011. **6**(6): p. e18643.
124. Soret, J., et al., *Selective modification of alternative splicing by indole derivatives that target serine-arginine-rich protein splicing factors*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(24): p. 8764-8769.

125. Labourier, E., et al., *Poisoning of Topoisomerase I by an Antitumor Indolocarbazole Drug: Stabilization of Topoisomerase I-DNA Covalent Complexes and Specific Inhibition of the Protein Kinase Activity*. Cancer Research, 1999. **59**(1): p. 52-55.
126. Pilch, B., et al., *Specific Inhibition of Serine- and Arginine-rich Splicing Factors Phosphorylation, Spliceosome Assembly, and Splicing by the Antitumor Drug NB-506*. Cancer Research, 2001. **61**(18): p. 6876-6884.
127. Soret, J., et al., *Altered Serine/Arginine-Rich Protein Phosphorylation and Exonic Enhancer-Dependent Splicing in Mammalian Cells Lacking Topoisomerase I*. Cancer Research, 2003. **63**(23): p. 8203-8211.
128. Ropers, D., et al., *Differential effects of the SR proteins 9G8, SC35, ASF/SF2, and SRp40 on the utilization of the A1 to A5 splicing sites of HIV-1 RNA*. Journal of Biological Chemistry, 2004. **279**(29): p. 29963-29973.
129. Ghigna, C., et al., *Pro-metastatic splicing of Ron proto-oncogene mRNA can be reversed: therapeutic potential of bifunctional oligonucleotides and indole derivatives*. RNA biology, 2010. **7**(4): p. 495-503.
130. Ghigna, C., et al., *Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene*. Mol Cell, 2005. **20**(6): p. 881-90.
131. Zhou, Y.Q., et al., *Altered expression of the RON receptor tyrosine kinase in primary human colorectal adenocarcinomas: generation of different splicing RON variants and their oncogenic potential*. Oncogene, 2003. **22**(2): p. 186-97.
132. Lu, Y., H.P. Yao, and M.H. Wang, *Multiple variants of the RON receptor tyrosine kinase: biochemical properties, tumorigenic activities, and potential drug targets*. Cancer Lett, 2007. **257**(2): p. 157-64.
133. Camp, E.R., et al., *RON, a tyrosine kinase receptor involved in tumor progression and metastasis*. Ann Surg Oncol, 2005. **12**(4): p. 273-81.
134. Ghigna, C., *Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene*. Mol. Cell, 2005. **20**: p. 881-890.
135. Karni, R., et al., *The gene encoding the splicing factor SF2/ASF is a proto-oncogene*. Nat Struct Mol Biol, 2007. **14**(3): p. 185-93.
136. Bonnal, S., L. Vigevani, and J. Valcarcel, *The spliceosome as a target of novel antitumour drugs*. Nat Rev Drug Discov, 2012. **11**(11): p. 847-859.
137. Andrews, S., *FastQC: A quality control tool for high throughput sequence data*. Reference Source, 2010.
138. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq*. Bioinformatics, 2009. **25**(9): p. 1105-1111.
139. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nat Meth, 2012. **9**(4): p. 357-359.
140. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation*. Nat Biotech, 2010. **28**(5): p. 511-515.
141. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks*. Nature protocols, 2012. **7**(3): p. 562-578.
142. Brooks, A.N., et al., *Conservation of an RNA regulatory map between Drosophila and mammals*. Genome research, 2011. **21**(2): p. 193-202.



143. Graveley, B.R., et al., *The developmental transcriptome of Drosophila melanogaster*. Nature, 2011. **471**(7339): p. 473-479.
144. Robinson, J.T., et al., *Integrative genomics viewer*. Nat Biotech, 2011. **29**(1): p. 24-26.
145. Shikata, H., et al., *Phytochrome controls alternative splicing to mediate light responses in Arabidopsis*. Proc Natl Acad Sci U S A, 2014. **111**(52): p. 18781-6.
146. Bartel, D.P. and C.-Z. Chen, *Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs*. Nat Rev Genet, 2004. **5**(5): p. 396-400.