Understanding the Epigenetics of alternative splicing in *Arabidopsis thaliana*

by

Ibtissam Jabre

Canterbury Christ Church University

Thesis submitted

for the Degree of Doctor of Philosophy

September 2019

Table of Contents

Word count (Without references): 50500 words.

Acknowledgements

<u>Declaration</u>

I declare that I'm the author of this thesis and the experimental work, data analysis, and the results presented here are conducted by myself. Work that belongs to other researchers has been clearly identified with references. I hereby declare that this thesis is my own and has not been submitted to this university or others. Given the nature of the work conducted in this study, no ethical approval is needed. Henceforth, this work doesn't include any ethical approval forms.

Ibtissam Jabre

Abstract

Being sessile and photosynthetic necessitates that plants must have a certain degree of predictability for ambient conditions during the daily cycles to maximise efficiency and survival. However, subtle or sudden changes in weather conditions alongside different growth and developmental phases necessitate that plants must continuously monitor different environmental cues and synchronise them with their physiology and metabolism in a time, growth phase and development-stage dependent manner. Plants use complex gene regulatory mechanisms to overcome environmental challenges. Alternative splicing (AS) of pre-mRNAs, a process that generates two or more transcripts from multi-exon genes, adds another layer of complexity to gene regulatory mechanisms to modulate transcriptome diversity in a tissue- and condition-dependent manner. In mammals, mounting evidence indicates that chromatin structure can regulate co-transcriptional AS. Recent evidence supports co-transcriptional regulation of AS in plants, but how dynamic changes in the chromatin influence the AS process upon cold stress remains poorly understood and is the subject of this study. In order to answer this question, four approaches were followed in parallel; (1) *Arabidopsis thaliana* (Arabidopsis) plants with identical DNA sequence but differential DNA methylation and nucleosome occupancy (Ctrl and AzadC plants with wild type DNA methylation and hypomethylation, respectively) were developed to perform (2) RNA sequencing (RNA-seq) and (3) Micrococcal Nuclease sequencing (MNase-seq) for Ctrl and AzadC grown at 22°C and at 4°C as cold treatment for 24 hours, and (4) whole genome bisulphite sequencing (WGBS) for AzadC grown at 22°C and at 4°C as cold treatment for 24 hours. This strategy allowed us to understand how epigenetic variations between AzadC treatment derived lines and Ctrl plants affect AS under cold conditions, without the confounding effects of sequence variation.

Interestingly, RNA-seq, MNase-seq, and WGBS show a strong reprogramming of AS patterns upon cold stress associated with changes in epigenetic features (i.e. DNA methylation and nucleosome occupancy). To my knowledge, this is the first study in Arabidopsis that demonstrates that changes in transcriptional and AS patterns coincide with genome-wide changes in nucleosome occupancy and DNA methylation patterns upon temperature shift.

List of abbreviations

AS	Alternative splicing
MES	N-Morpholino ethanesulfonic acid
µg	Micrograms
µl	Microliter
µM	Micromolars
5mC	5-methylcytosine
A	Adenosine
A3'SS	Alternative 3' splice site
A5'SS	Alternative 5' splice site
ABA	Abscisic acid
AF/AL	Alternative first/Alternative last
Alt3'	alternative acceptor
Alt3'	Acceptor splice
Alt5'	alternative donor
Alt5'	Alternative donor
AtRTDv2	Arabidopsisthaliana reference transcriptome dataset version 2
BAM	Binary Alignment Map
Bcl	Binary base call
bp	Base pairs
CCA1	*CIRCADIAN CLOCK ASSOCIATED 1*
CG	Cytosine Guanine
ChIP	Chromatin immunoprecipitation
circRNAs	Circular RNAs
Col-0	Columbia-0
comma delimited	Csv
CPM	Count per million
CS	Constitutive splicing
CTCF	CCCTC-binding factor
CTD	Carboxyl-terminal domain
CTOB	Converted Original bottom strand
CTOT	Converted Original top strand
Cytosine	Cytosine
DAS	Differentially alternatively spliced
DE	Differentially expressed
DMRs	Differentially methylated regions
DOG1	*DELAY OF GERMINATION 1*
DPNs	Differentially positioning nucleosome
DREB2	*DEHYDRATION-RESPONSIVE ELEMENT BINDING 2*
dsDNA	Double-stranded DNA
eIFs	Eukaryotic initiation factors
EIs	Exitrons
ES	Exon skipping

ESEs	Exonic splicing enhancers
ESSs	Exonic splicing silencers
FDR	False discovery rate
FLC	*FLOWERING LOCUS C*
FLM	*FLOWERING LOCUS M*
G	Guanine
g	Grams
GO	Gene Ontology
gRNA	Guide RNA
GRO-seq	Global run-on sequencing
GRP	Glycine-Rich RNA Binding Proteins
hnRNPs	Heterogeneous nuclear ribonucleoproteins
HR-RT PCR	High resolution reverse transcription polymerase chain reaction
HSFs	Heat sock transcription factors
IDPs/IDRs	Intrinsically disordered proteins or regions
IGV	integrated genome visulaser
iNPS	improved nucleosome positioning
IR	Intron retention
ISEs	Intronic splicing enhancers
Iso-Seq	Isoform sequencing
ISSs	Intronic splicing silencers
L	Litre
LC-MS/MS	Liquid chromatography–mass spectrometry
LHY	*LATE ELONGATED HYPOCOTYL*
LSM4	*Sm-like4*
METTL3	*N6-adenosine-methyltransferase-like 3*
MICU	*MITOCHONDRIAL CALCIUM UPTAKE*
miRNAs	microRNAs
Ml	Millilitres
ml	Millilitre
MNase-seq	Micrococcal Nuclease sequencing
mRNA	Messenger RNA
MS	Mass-Spectrometry
MS	Murashige and Skoog
mTORC1	Rapamycin Complex 1
MXE	Mutually exclusive exons
NASC	Arabidopsis Stock Centre
ng	Nanograms
NGS	Next genereation sequencing
nM	Nanomolars
NMD	Non-sense mediated decay
OB	Original bottom
ORF	Open reading frame
Os*Met1-2*	CG methyltransferase mutant
OT	Original top
PAS	Polyadenylation site

PCA	Principal components analysis
pNET-Seq	Plant native elongating transcript sequencing
Poly(A)	Polyadenylation
PPR9	*PSEUDO RESPONSE REGULATOR 9*
pre-mRNAs	Precursor mRNAs
PSI	Percent spliced in
PTB	Polypyrimidine Tract Binding Proteins
PTC	Premature termination codon
PTMs	Post-translational modifications
QC	Quality control
q-PCR	Quantitative polymerase chain reaction
Ribo-Seq	Ribosome profiling/foot-printing along with next-generation sequencing
RNAi	RNA Interference
RNAPII	RNA polymerases II
RNA-seq	RNA sequencing
RNP	Ribonucleoprotein particle
RPGs	Ribosomal protein genes
rpm	Rotation per minute
SAM	Sequence Alignment Map
SB	Sodium bisulfite
SF-RBPs	Splicing factors RNA-binding proteins
SFs	Splicing factors
SJs	Splice junctions
snRNPs	Ribonucleoprotein particles
SPB	Succinimidyl-[4-(psoralen-8-yloxy)]-butyrate
SR	Serine/arginine-rich
SS	splice sites
ssODNs	oligodeoxynucleotide
SUA	*SUPPRESSOR OF ABI3-5*
SURF2	*SURGEIT LOCUS PROTEIN 2*
TES	Transcription end site
TFs	Transcription factors
TPM	Transcripts per million
tRNA	Transfer RNA
TSS	Transcription start site
U	Uracil
UPF	Up-frameshift
UTR	Untranslated region
WGBS	Whole genome bisulphite sequencing
WT	Wild-type
ZIFL1	*ZINC-INDUCED FACILITATOR-LIKE 1*
ΔPSI	Splicing change
ΔPS	Δ percent spliced
μE m−2 s−1	microeinsteins per second per square meter
μm	Micrometres

L_2FC Log$_2$fold change

List of figures

List of tables

Chapter 1. Introduction

1.1 Overview of gene expression

In eukaryotes, gene expression begins with transcription whereby, the genetic information stored in the DNA sequences (genes) is transmitted to an intermediate molecule called messenger RNA (mRNA) (Chambon, 1978). The mRNA molecules are then transported into the cytoplasm and serve as a template for the production of proteins, through a process called translation, to make the cellular machineries (Chambon, 1978). Gene expression is dynamic and varies during different developmental stages and in response to different cellular and environmental conditions. To fine-tune cellular physiology and metabolism under normal as well as stress conditions, gene expression is controlled by a complex of regulatory networks at different levels to ensure the production of the correct amount and type of proteins, which in turn undergo post-translational processing to increase protein diversity and control a variety of cellular functions (Lelli *et al.,* 2012).

At the transcription level, the DNA-dependent RNA polymerases II (RNAPII) uses the DNA stored in the nucleus as a template to transcribe, in a number of distinct phases, precursor mRNAs (pre-mRNAs) from protein-coding genes (Svejstrup, 2004; Proudfoot, 2011). At the pre-initiation stage of transcription, RNAPII and other transcription factors (TFs) assemble over specific sequences in the promoter region located approximately 25 nucleotides upstream of the transcription initiation site of each gene (i.e. TATA box), as well as other regulatory sequences located upstream of the initiation binding site that regulate the frequency of gene expression (Hahn, 2004). Beside TATA box, which is a conserved element sequence among eukaryotes, genes contain other TFs binding sites that are targeted by their cognate proteins upon specific stressful or development conditions; hence ensuring gene transcriptional activation or repression in a condition- and tissue- specific manner (Svejstrup, 2004). The level of chromatin condensation is very dynamic throughout the cell cycle and can restrict or allow transcription initiation through modulating the binding affinity of TFs. (Li *et al.,* 2007). At the simplest compaction level, ~165 base pair of double-stranded DNA (dsDNA) wrapped around a core of eight proteins, two of each of the histone proteins H2A, H2B, H3 and H4 to form the basic packing unit of the DNA called a nucleosome core (García, González and Antequera, 2017). Two neighbouring nucleosome cores are then joined by a fragment of 50 bp of linker dsDNA sealed by linker H1 or H5 histone proteins. The nucleosome core with ~165 bp of DNA together with the linker histone is called the chromatosome. The chromatosome and the

additional linker DNA constitute a higher compaction level of DNA called nucleosomes. For a higher compaction level, each nucleosome folds up to form a 30-nanometer chromatin fiber, resulting in loops averaging 300 nanometres (nm) in length. The 300 nm fibers are compressed and folded to produce a 250 nm-wide fiber, which is tightly coiled into the chromatid of a chromosome (Cutter and Hayes, 2015). Upon RNAPII and TFs binding, the two DNA strands disassociate to form an open complex with the RNAPII, which subsequently allow the initiation stage of transcription to start (Li, Carey and Workman, 2007). At this stage, RNAPII is released from the promoter regions towards the gene body to start the synthesis of the first complementary nucleotides (adenine (A), uracil (U), guanine (G) and cytosine (C)), which are connected together through a phosphodiester bond to form a nascent RNA that remains bound to the DNA throughout the elongation stage (Li, Carey and Workman, 2007). The RNA chain grows as more nucleotides are added to its 3' end during the elongation stage. Finally, transcription ends when RNAPII reaches terminator sequences and stops the addition of nucleotides to the RNA chain. This latter is then released as nascent pre-mRNA from the DNA template alongside RNAPII (Proudfoot, 2011; Mischo and Proudfoot, 2013; Lemay and Bachand, 2015).

Co- and post-transcriptional regulation of gene expression ensure the correct processing of nascent pre-mRNA to produce mature mRNA ready to be exported from the nucleus to the cytoplasm for translation. The first control at the co-transcriptional level is 5'capping during which, a guanine nucleotide carrying a methyl group (m 7 G) is added to the first nucleotide at the 5'end (5' cap) of the pre-mRNA chain. The 5'cap structure protects the nascent pre-mRNA from exonucleolytic degradation and plays important roles in RNA stability, nuclear export, splicing, and translation efficiency (Cowling and Cole, 2010). Splicing is a second mechanism of co-transcriptional regulation of gene expression, whereby the nascent pre-mRNA can undergo two types of splicing known as constitutive and alternative splicing (CS and AS respectively). CS consists of intron removal and exon joining to produce one mRNA from multi-exon genes. Unlike CS, AS uses differential AS sites to engender multiple transcript variants from multi-exon genes (see section 1.3 and 1.4 for details). The process of splicing is catalysed by the spliceosome, which is composed of five small nuclear ribonucleoprotein particles (snRNPs) designated as U1, U2, U4, U5, and U6 and additional spliceosome-associated non-snRNP proteins (Wahl, Will and Lührmann, 2009; Day *et al.*, 2012). The final step of gene expression control at the co- and post-transcriptional level involves a polyadenylation step during which, RNA binding proteins cleave enzymatically at the UA-rich

cleavage signal located at the 3' end of nascent pre-mRNA. Afterwards a poly(A) tail of a few hundred nucleotides long (~ 250 adenine residues) is added at the 3' end by poly(A) polymerase (Darnell, 2013).

In eukaryotes, once pre-mRNA processing is completed, mature mRNA is transported to the cytoplasm by large ribonucleoprotein particle (RNP) complexes, the ribosomes, for translation. Translation initiation begins with the binding of eukaryotic initiation factors (eIFs), eIF-1, eIF-1A, and eIF-3 to the 40S ribosomal subunit and the association of eIF-2 (in a complex with GTP) with the initiator methionyl transfer RNA (tRNA) (Sonenberg and Hinnebusch, 2009; Jackson, Hellen and Pestova, 2010). Afterwards, the initiation factors recognise the 5' cap (eIF-4E) at the 3' end (eIF-4G and eIF-4E) of the mRNA; thus accounting for simultaneous translation and polyadenylation. Then, the initiation factors eIF-4E and eIF-4G associate with eIF-4A and eIF-4B to bind the mRNA to the 40S ribosomal subunit. Once assembled, the triplex, 40S ribosomal subunit, eIFs, and the methionyl tRNA start mRNA scanning to identify the AUG initiation codon A (Sonenberg and Hinnebusch, 2009; Jackson, Hellen and Pestova, 2010). Once identified, the eIF-2 is released from the translation initiation complex, followed by the binding of the large ribosomal subunit 60S to form the 80S initiation complex of eukaryotic cells (Kozak, 1989, 1992; Jackson, Hellen and Pestova, 2010). Once the initiation complex is assembled, the elongation step begins, whereby the initiating *N*-formylmethionyl tRNA occupies the first ribosomal binding site designated P (peptidyl) as a complement to the AUG start codon. Then, the ribosome moves along and read the mRNA in frame of codons that represent nucleotide triplets, and the aminoacyl tRNA carries an amino acid together with an anticodon adaptor that has the complementary sequence of a specific codon to the second ribosomal binding site designated A (aminoacyl). Then, a peptide bond is formed, followed by the translocation of the first two amino acids to the P site and the uncharged tRNA to the third tRNA-binding site termed E (exit). A new aminoacyl tRNA can then be added in the site A for addition of the next amino acid in the growing peptide chain (Dever and Green, 2012). The elongation of polypeptide chain terminates when UAA, UAG, or UGA stop codons cannot be identified by aminoacyl tRNA in the ribosomal binding site A. Alternatively, stop codons are recognized at the site A by a single release factor eRF-1, which in cooperation with eRF-2 stimulate the disassembly of the ribosomal subunits and the release of polypeptide chain (Dever and Green, 2012).

Translational regulation adds another layer of control to gene expression regulation of eukaryotic cells. Translation regulation mechanisms can either affect specific mRNAs translation efficiency or the overall translational activity (Gebauer, Preiss and Hentze, 2012). In the first case, regulator proteins bind their complementary sequences in the mRNA to block translation through interfering with cap recognition and binding of the 40S ribosomal subunit or to stabilize the mRNA through its protection from nuclease degradation (Hinnebusch and Lorsch, 2012; Roux and Topisirovic, 2012). Another mechanism of translational regulation is the localization of mRNAs to different cellular regions of eggs or embryos by different regulator proteins; thus, allowing condition- and development stage- specific translation (Chao, Yoon and Singer, 2012; Lasko, 2012). Nevertheless, modulating the overall translational activity involve the control of eIF-2, responsible for initiation complex formation. For eIF-2 to be in an active state and to bind to the initiator methionyl tRNA, eIF-2B catalyses the exchange of bound GDP for GTP to form eIF-2/GTP complex ready for translation initiation. In particular and to block translation initiation, regulatory protein kinases can phosphorylate eIF-2 to inhibit the exchange of bound GDP for GTP (Hinnebusch and Lorsch, 2012; Pavitt and Ron, 2012). A third mechanism of translational regulation is to control the polyadenylation of mRNAs, whereby untranslated mRNAs with short poly(A) tail are stored in the nucleus in the early development stages and are subsequently recruited for translation at the appropriate stage of development by the lengthening of their poly(A) tails (Cui *et al.*, 2013; Lim *et al.*, 2016). At the post-translational level, a variety of chemical changes termed post-translational protein modification (PTMs) are catalysed by enzymes that attach covalent chemical moieties to specific amino acid residues (Lothrop, Torres and Fuchs, 2013; Strumillo and Beltrao, 2015). In eukaryotes, the common PTMs are acetylation, phosphorylation, glycosylation and ubiquitylation (Kaikkonen, Lam and Glass, 2011). These modifications regulate various aspects of cellular functionalities such as protein structure, folding, subcellular localisation, and protein-substrate/protein interactions, and their functional state (Kaikkonen, Lam and Glass, 2011; Nachtergaele and He, 2017). PTMs are reversible by the action of deconjugating enzymes hence, allowing the control of protein function (Kaikkonen, Lam and Glass, 2011; Nachtergaele and He, 2017).

In brief, eukaryotes regulate gene expression throughout a complex regulatory mechanisms to control cell-to-cell interactions, orchestrate multiple stages of development, and adapt to environmental changes. This is mainly achieved by: (1) transcriptional control of gene expression that control the amount of mRNA transcribed from a gene in a condition- and time-

specific manner, (2) co- and post- transcriptional regulation that dictate precise temporal and spatial mRNA translatability and translation efficiency, and (3) PTMs that affect multiple facades of protein functions and viability. In plants, aforementioned genetic mechanisms enable adaptation to stressful conditions; however, locking every stress experience in the form of a genetic code may not be an effective strategy considering sessile nature of plants, the diversity of stressful conditions, and the day/night cycle fluctuations. Henceforth, plants must have a certain degree of predictability for ambient conditions during the daily cycles to maximise their efficiency and survival. However, subtle or sudden changes in weather conditions alongside different growth and developmental phases necessitate that plants must continuously monitor different environmental cues and synchronise them with their physiology and metabolism in a time, growth phase and development-stage dependent manner. AS of pre-mRNAs has emerged as an important regulatory co-transcriptional mechanism that influence plant gene expression patterns under normal and stressful conditions through fine-tuning their transcriptome and protein diversity (Syed *et al.*, 2012; Reddy *et al.*, 2013; Chaudhary *et al.*, 2019; Jabre *et al.*, 2019). In plants, clear body of evidence indicates that AS patterns vary upon different physiological conditions and in response to various environmental stresses to ensure their survival in a changing environment (Allan B James, Syed, Bordage, *et al.*, 2012; Allan B James, Syed, Brown, *et al.*, 2012; Calixto *et al.*, 2018; Filichkin *et al.*, 2018). In mammals, recent evidence indicates that epigenetic changes such as DNA methylation, chromatin modifications, regulate RNAPII processivity, co-transcriptional AS, and the stability as well as the translation efficiency of splice isoforms (Luco *et al.*, 2010, 2011; Gonzalez *et al.*, 2015; Cheng *et al.*, 2017). In plants, the role of epigenetic modifications in regulating transcription rate and mRNA abundance under stress is beginning to emerge (Core, Waterfall and Lis, 2008; Zhu *et al.*, 2018; Jabre *et al.*, 2019). However, how plants modulate AS responses through epigenetic modifications to adapt environmental challenges is still elusive.

Previously, high number of studies and reviews have described the process of spliceosome assembly and the splicing cycle as well as the mechanisms of AS (Wahl, Will and Lührmann, 2009; Will and Lührmann, 2011), yet a big gap persists in the field of epigenetic regulation of AS in plants. For this reason, I am providing brief account of pre-mRNA splicing and presenting detailed information how the chromatin structure and the crosstalk at the co/post-transcriptional level regulate the fate of alternatively spliced transcripts in plants upon stress responses. Similarly, to which extent alternatively spliced transcripts in plants contribute to proteome diversity under normal as well as stressful conditions is poorly understood and is the

focus of this chapter. I will also discuss in detail how environmental cues dictate transcripts destined for translation, nuclear sequestration, or degradation via such crosstalk; thus, affecting protein stability and function. To my knowledge, no previous experimental work or literature review have linked the effect of environmental changes on chromatin re-modelling to regulate co-transcriptional AS outcomes and subsequently protein diversity. Hence, this chapter addresses the lack of knowledge in the field of plants epigenetic regulation of co-transcriptional regulation of AS upon environmental stresses.

1.2 Transcription and splicing dynamics in plants

Transcription is a fundamental process to orchestrate gene expression patterns in response to different developmental and environmental cues. Surprisingly, limited information is available on the mechanism of transcription in plants (Hetzel *et al.*, 2016). Human promotors are GC-rich (Core, Waterfall and Lis, 2008; Hetzel *et al.*, 2016), whereas plant promoters are AT-rich and tend to inhibit nucleosome formation, promoting DNA flexibility and transcription factor recruitment (Zuo and Li, 2011). Comparison of RNA-seq and global run-on sequencing (GRO-seq) data sets in Arabidopsis revealed a high correlation between nascent and steady-state transcripts (Hetzel *et al.*, 2016). Further, stable transcripts were associated with biological functions like translation, photosynthesis and metabolic functions. On the other hand, unstable transcripts had a higher representation of stimulus response genes, signal transduction, and hormones (Hetzel *et al.*, 2016). These results highlight that conserved genes associated with housekeeping functions are more stable compared with highly regulated transcripts. In view of these findings, it would be reasonable to speculate that AS transcripts, as a result of their dynamic nature, would be more suited for regulatory roles. Previous GRO-seq data showed that plant promoters lack promoter-proximal pausing of RNAPII and divergent transcription, which are prevalent in humans as well as yeast and drosophila (Nechaev *et al.*, 2010; Hetzel *et al.*, 2016). However, very recent GRO-seq and plant native elongating transcript sequencing (pNET-seq) experiments from Arabidopsis indicate that RNAPII with an unphosphorylated carboxyl-terminal domain (CTD) indeed accumulates downstream of transcription start sites (TSS) (Zhu *et al.*, 2018). However, promoter-proximal pausing in Arabidopsis is much more loose (broad peak) compared with mammals where pausing occurs in a narrow window of 25-50 nucleotides (Zhu *et al.*, 2018). These findings indicate that efficient RNAPII recruitment, as well as release from promoter-proximal pausing is necessary for efficient transcriptional response in Arabidopsis. Interestingly, plant promoters also show Ser2P CTD RNAPII

accumulation adjacent to the 3' polyadenylation site (PAS), suggesting the presence of a surveillance mechanism before transcription termination (Zhu *et al.*, 2018). In vitro work in yeast proposed that RNAPII pausing after PAS may increase surveillance time and aid in mRNA degradation (Anamika *et al.*, 2012). In addition, Ser5P CTD RNAPII elongates more slowly in exons compared with introns to provide more time for the spliceosome to appropriately select splice sites in Arabidopsis (Zhu *et al.*, 2018). These data show that RNAPII CTD phosphorylation is a dynamic process and maybe even more important for sessile organisms like plants to maintain appropriate transcriptional and splicing dynamics under varied conditions. Since AS is largely co-transcriptional, distinctive features of plant transcription (transcription initiation and TSS/PAS proximal RNAPII pausing) may have a bearing on the transcriptional, splicing, and processing dynamics before a transcript is released from the transcription and splicing machinery (Irimia *et al.*, 2014; Hetzel *et al.*, 2016; Zhu *et al.*, 2018).

1.3 Pre-mRNA splicing

Pre-mRNA splicing is catalysed by the spliceosome, a large ribonucleoprotein complex that recognises various *cis*-sequences in pre-mRNAs, including 5' and 3' splice sites, branch points, polypyrimidine tracts, and other splicing regulatory elements (suppressors and enhancers) The core spliceosome is composed of five uridine-rich snRNPs (U1, U2, U4, U5 and U6) and additional spliceosome-associated proteins Other non-snRNP SFs, predominantly serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs), target splicing enhancers and suppressors located in exons and introns, and regulate splice site selection by the spliceosome (Wahl, Will and Lührmann, 2009; Will and Lührmann, 2011; Lee and Rio, 2015).

AS occurs when the spliceosome differentially selects splice sites. Common types of AS include exon skipping (ES), mutually exclusive exons (MXE), intron retention (IR), and selection of alternative donor (Alt5') and acceptor splice (Alt3') sites (Kim, Magen and Ast, 2007). Recently characterised exitrons (EIs) complement the repertoire of AS events (Marquez *et al.*, 2015; Staiger and Simpson, 2015). EIs are alternatively spliced internal regions of reference protein-coding exons. Majority of EIs have lengths divisible by three and they broadly impact protein function by affecting protein domains, disordered regions, and the availability of sites for various PTMs (Marquez *et al.*, 2015).

Different splice isoforms display various fates in plants that may include (1) nuclear sequestration and further splicing to generate full-length mRNAs (Yang, Wightman and Meyerowitz, 2017; Hartmann, Wießner and Wachter, 2018), (2) translation into functional or truncated proteins (Penfield, Josse and Halliday, 2010; Mastrangelo *et al.*, 2012; Liu *et al.*, 2013), and (3) degradation via nonsense-mediated mRNA decay (NMD) (Hori and Watanabe, 2005, 2007; Arciga-Reyes *et al.*, 2006; Schwartz *et al.*, 2006; Kerényi *et al.*, 2008; Nyikó *et al.*, 2009; Palusa and Reddy, 2010).

1.4 Alternative splicing mediates plant responses to abiotic/biotic stresses

Plants resort to AS to fine-tune their physiology and metabolism to maintain a balance between carbon fixation and resource allocation under normal as well as biotic and/or abiotic stress conditions such as pathogen infection, temperature, salt, drought, wounding, and light (Calixto *et al.*, 2018; Filichkin *et al.*, 2018; Seaton *et al.*, 2018). RNA-Seq and single molecule isoform sequencing (Iso-Seq) data from poplar leaf, root, and stem xylem tissues under drought, salt and temperature fluctuations revealed that stress-induced changes in AS profiles modulate plants transcriptome to abiotic stresses and highlighted IR as the predominant type of AS, where intron-containing isoform ratios change across all treatments and tissue types (Filichkin *et al.*, 2018). Recent studies show that salt stress and high temperature alter the splicing patterns of more than 6000 genes in Arabidopsis and 1000 genes in grapes (Feng *et al.*, 2015; Jiang *et al.*, 2017). Arabidopsis heat-shock TFs (HSFs) are the most well-known genes, which are extensively modulated by AS in response to extreme heat (62). Interestingly, the expression of HSFs is regulated by the DEHYDRATION-RESPONSIVE ELEMENT BINDING 2 (DREB2)TFs, which undergoes stress-induced AS (Liu *et al.*, 2013; Z. Liu *et al.*, 2017). Another example of heat-induced AS is Arabidopsis ZINC-INDUCED FACILITATOR-LIKE1 (ZIFL1) pre-mRNA that produces a full-length splice variant (ZIFL1.1), which is translated into a protein isoform that localizes to the vacuolar membrane of root cells to regulate auxin transport (Remy *et al.*, 2013). The second splice isoform, ZIFL1.3 transcript, encodes a truncated protein that lacks two *C*-terminal domains and is targeted to the plasma membrane of leaf stromal guard cells to mediate drought tolerance (Remy *et al.*, 2013). Cold stress is also known to induce AS in plants. Recently, time-series RNA-Seq data from Arabidopsis plants exposed to cold stress identified 8,949 genes which were differentially expressed (DE) and differentially alternatively spliced (DAS) of which 1,647 genes were regulated only at the AS

level (Calixto *et al.*, 2018). In rice, cold-induced IR and ES of OsCYP19-4 result in eight AS isoforms, which are required for stress tolerance (Lee *et al.*, 2016). In wheat, global profiling of AS landscape after drought, and heat treatments, and their combination show that 200, 3576 and 4056 genes exhibit significant DAS to drought stress, heat stress and their combination, respectively (Liu *et al.*, 2018). Additionally, investigating the influence of external environmental signals on circadian clock genes' rhythmic oscillations showed that temperature transitions and pathogen infection dictate not only the ratios of nonsense transcript isoforms but the timing of their expression as well (Filichkin *et al.*, 2015). In Arabidopsis, IR1 in the 5'UTR of Arabidopsis *LATE ELONGATED HYPOCOTYL (LHY)* plays a critical role in plants' adaptation to temperature fluctuations (Brown *et al.,* 2018). Interestingly, polypyrimidine tract binding protein (PTB), U2AF65A, and SUPPRESSOR OF ABI3-5 (SUA) display temperature–dependent isoform switching of premature termination codon (PTC) containing transcripts to regulate the levels of their fully spliced protein-coding isoforms through AS (Brown *et al.,* 2018). Changes in AS patterns of SFs were congruent with IR1 detected in *LHY* (Brown *et al.,* 2018). These findings indicate that stress-induced AS of SFs coordinate the splicing patterns of downstream target stress-responsive genes.

Similarly, AS is also key in biotic stress responses, for example, data from soybean show that PsAvr3c, a *Phytophthora sojae* pathogen effector, may manipulate host spliceosomal machinery to shift splicing profiles and overcome the host immune system (Huang *et al.*, 2017). This is achieved by the localization of PsAvr3C to the nucleus to stabilize and inhibit the proteosomal degradation of soybean genes rich in serine, lysine, and arginine (GmSKRP1/2), which is shown to be a negative regulator of plant immunity (Huang *et al.*, 2017). Interestingly, co-immunoprecipitation assay of GmSKRP1/2 show its interaction with other SFs that subsequently affect the splicing patterns of over 400 genes (Huang *et al.*, 2017). Additionally, *Pseudomonas Syringae* Pv. leaf infection in Arabidopsis induces the retention of the long intron of *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)* and causes a moderate increase of the PTC *LHY* cassette exon isoform (Filichkin *et al.*, 2015). In rice, upon pathogen infection, the WRKY family of TFs OsWRKY62 and OsWRKY76 undergo CS and AS to promote plant defence response (Liu *et al.*, 2016). Interestingly, the majority of genes encoding splicing regulators in plants are subject to extensive AS and change the profile of their splicing patterns in response to various environmental stresses (Palusa *et al.,* 2010; Zhang *et al.*, 2013). For example, in Arabidopsis, profiling of 19 SR gene splicing events under different developmental stages and in response to cold, heat and hormone treatment show that 80% of SR genes tested encode at

least 95 transcripts to increase the transcriptome complexity by six-fold, which resulted in increasing the splicing patterns of 49% of all intron-containing genes (Palusa, Ali and Reddy, 2007). Additional data also show the differential recruitment of SR splice variants for translation under normal conditions and in response to stresses (Palusa and Reddy, 2015). Taken together, biotic and abiotic stress responses in plants are coordinated through a network involving AS of SFs and their downstream target genes. How plants perceive environmental stresses and regulate their AS profiles in a condition-dependant manner is patchy; however, epigenetic mechanisms may mediate the crosstalk between different stresses and corresponding transcriptional and AS responses.

1.5 Regulation of Alternative splicing

Regulation of AS and the fate of alternatively spliced transcripts is mainly driven by the concentration of SFs and their proportions (largely due to competition between SR proteins as positive regulators and hnRNPs as negative regulators for binding to *cis*-regulatory elements in particular cell types/conditions. Additionally, the structure of pre-mRNAs also regulates splicing significantly (Shen, Julie L.C. Kan and Green, 2004; Y. Ding *et al.*, 2014). In both mammals and plants, chromatin, which carries differential DNA methylation and multiple histone modifications, may mediate RNAPII processivity to influence splicing outcomes (Schwartz, Meshorer and Ast, 2009; Tilgner *et al.*, 2009; Luco *et al.*, 2010; Lyko *et al.*, 2010; Malapeira, Khaitova and Mas, 2012; Gelfman *et al.*, 2013; Malapeira and Mas, 2013; Wan *et al.*, 2013; Ullah *et al.*, 2018). Hence, splicing regulation is mediated through a complex cellular network referred to as the "splicing code" that fine-tunes gene expression in response to different conditions (Barash *et al.*, 2010; Reddy *et al.*, 2012).

1.5.1. Cis-elements and trans-acting factors

Spliceosome assembly during intron removal and exon joining is regulated by *cis*-regulatory elements of the pre-mRNA including; splice sites, branch point, polypyrimidine tract, sequence elements enhancer and suppressor of splicing (Lorković *et al.*, 2000; Lee and Rio, 2015). The sequence elements include exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs), depending on their location (in introns or exons), promote or inhibit splicing in exon or intron, based on their nomenclature (Lee and Rio, 2015). The richness of introns with UA and GC, respectively also help the spliceosome to define intron/exon borders and subsequently identifying a conserved

GU dinucleotide at the donor site (at the beginning of the intron) and an AG dinucleotide at the acceptor site (at the end of the intron) (Schwartz, Meshorer and Ast, 2009; Tilgner *et al.*, 2009).

Cis-regulatory elements are recognised by *trans*-acting factors to regulate splicing. The most common SFs known in plants that control both CS and AS are SR proteins, hnRNPs proteins (Matlin, Clark and Smith, 2005). In plants and metazoa, SR proteins are highly conserved and are characterized by a RS domain rich in serine and arginine involved in protein-protein interaction at their C-terminal, and one or two RNA recognition motifs that bind the mRNA at their N-terminus (Kalyna and Barta, 2004; Richardson *et al.*, 2011). Interestingly, plants possesses almost a double number of SR genes present in mammals pointing towards a wide range of mechanisms to control AS at different development stages and in response to environmental changes in plants (Kalyna and Barta, 2004; Richardson *et al.*, 2011). Moreover, SR genes undergo extensive AS in response to environmental stresses to expand the variation, abundance, and activity of SFs (Palusa, Ali and Reddy, 2007; Tanabe *et al.*, 2007; Duque, 2011).

The second group of the RNA binding proteins are hnRNPs proteins, which are characterized by high molecular weight and act in homopolymer complexes (Han, Tang and Smith, 2010). hnRNPs proteins bind as well to dsDNA and are involved in nucleic acid metabolism and multiple biological processes (Wachter, Rühl and Stauffer, 2012). Depending on their binding position and cellular concentration, hnRNPs proteins mediate CS and AS using different mechanisms such as interfering with spliceosomal components interaction, coupling splicing with other steps in gene expression, and interfering with mRNA structure (Matlin, Clark and Smith, 2005). hnRNP are divided into two groups which are the Polypyrimidine Tract Binding Proteins (PTB), and Glycine-Rich RNA Binding Proteins (GRP). In Arabidopsis, three PTB protein homologs were identified, designated as PTB1 (At3g01150), PTB2 (At5g53180), and PTB3 (At1g43190) (Rühl *et al.*, 2012). It is noteworthy that pre-mRNA splicing of all three PTB homologs from Arabidopsis generates two types of splice variants of which, one encodes the full-length protein, whereas the alternative variant contains a PTC and is subject to degradation via NMD (Rühl *et al.*, 2012). In Arabidopsis, 21 glycine-rich RNA binding proteins were identified as homologous for human hnRNP A1and hnRNP A2/B1 however, AtGRP7 and AtGRP8 are the most investigated proteins in Arabidopsis and are known to cross auto-regulate their own pre-mRNA via the mean of AS NMD (Heintzen *et al.*, 1997; Staiger *et al.*, 2003; Schmal, Reimann and Staiger, 2013) .

1.5.2 Co-transcriptional regulation of alternative splicing

An extensive body of evidence suggests that splicing is predominantly coupled to transcription in metazoans, and is dependent on chromatin structure, which is modulated by DNA methylation, histone PTMs and chromatin adapter complexes (Cramer *et al.*, 1999; Listerman, Sapra and Neugebauer, 2006; Swinburne *et al.*, 2006; Carrillo Oesterreich, Preibisch and Neugebauer, 2010a; Khodor *et al.*, 2011). CTD of RNAPII serves as a landing pad for the recruitment of proteins involved in capping, splicing, polyadenylation and export of transcripts (Alexander *et al.*, 2010; Luco *et al.*, 2010; Fusby *et al.*, 2015). Various studies have shown that RNAPII CTD phosphorylation facilitates the recruitment of SFs including SR proteins to influence both CS and AS (Hirose and Manley, 2000; Gasch *et al.*, 2006; Lenasi and Barboric, 2010; Hajheidari, Koncz and Eick, 2013). Recruitment and kinetic models have been proposed to explain the mechanism by which transcriptional machinery controls AS (Brody *et al.*, 2011; Luco *et al.*, 2011; Jimeno-González *et al.*, 2015; Dvinge, 2018). The recruitment model states that the transcription machinery interacts directly or indirectly with SFs and thereby affects splicing outcomes. The kinetic model proposes that decreasing the speed of RNAPII allows additional time for an upstream exon with weak splice sites to recruit the splicing machinery before a downstream exon with stronger splice sites emerges during pre-mRNA synthesis (Roberts *et al.*, 1998; Brody *et al.*, 2011).

Similar to mammals (Nojima *et al.*, 2015), very recent NET-seq data from Arabidopsis also showed that phosphorylation of RNAPII at Ser 5P mediates interactions with the spliceosome (Zhu *et al.*, 2018). In addition, RNAPII elongation speed in Arabidopsis was also found to be slower in exons than introns, facilitating exon and splice site recognition. Accumulation of RNAPII Ser 5P at 5′ splice sites, in concert with the splicing machinery, facilitates 5′ splice site recognition and cleavage during elongation (Zhu *et al.*, 2018). Interestingly, plants can employ a signaling molecule from chloroplasts to regulate AS in the nucleus under different light conditions (Petrillo, Godoy Herz, Fuchs, *et al.*, 2014). The nature of this chloroplast-derived retrograde signal is not clear, although a nuclear regulatory mechanism that affects AS of a subset of Arabidopsis genes has been revealed (Petrillo, Godoy Herz, Fuchs, *et al.*, 2014; Godoy Herz *et al.*, 2019). Interestingly, RNAPII elongation speed is faster under light conditions than in darkness. In addition, greater RNAPII processivity is associated with a more open chromatin structure, which favors RNAPII elongation (Petrillo, Godoy Herz, Fuchs, *et al.*, 2014; Godoy Herz *et al.*, 2019). These results provide strong evidence that plants can

control nuclear events such as AS by coupling environmental and physiological cues to RNAPII elongation speed, and thereby elicit an appropriate plant responses (Godoy Herz *et al.*, 2014, 2019; Petrillo, Godoy Herz, Barta, *et al.*, 2014; Petrillo, Godoy Herz, Fuchs, *et al.*, 2014). Similarly, the spliceosome disassembly factor NTR1 is essential for appropriate expression and splicing of the *DELAY OF GERMINATION 1* (*DOG1*) gene. AtNTR1-deficient plants display a higher RNAPII elongation rate, preference for downstream 5′ and 3′ splice sites, and increased exon skipping (Dolata *et al.*, 2015). Interestingly, AtNTR1 also co-localizes with RNAPII to achieve splicing of target genes (Dolata *et al.*, 2015). Recent data from plants have also identified a strong relationship between chromatin changes, transcriptional control and AS regulation. For example, quantitative variation in the transcription of the *FLOWERING LOCUS C* (*FLC*) gene in Arabidopsis was associated with H3K36me3 and H3K4me2 histone marks, suggesting that different chromatin states influence initiation and elongation rates that affect splicing of *FLC* (Wu *et al.*, 2016). Chromatin-bound RNA was more abundant inside exon 1 of *FLC* than at the exon1–intron1 junction, suggesting that splicing at intron 1 is mostly co-transcriptional (Wu *et al.*, 2016). Additionally, *FLC* intron 1 retention is associated with a high level of H3K27me3, which is coincident with low cytosine-guanine (CG) methylation and H3K36me3/H3K4me1 marks, demonstrating a link between chromatin features and splicing outcomes in the *FLC* gene (Mahrez *et al.*, 2016). Recently, Ullah *et al.* (Ullah *et al.*, 2018) investigated the relationship between open chromatin and intron retention in Arabidopsis and rice. They showed that the chromatin structure is more open in retained introns. Based on this correlation, it was suggested that the open chromatin architecture in retained introns enhances the RNAPII elongation rate, which leads to skipping of splice sites by the spliceosome (Ullah *et al.*, 2018). Together these studies strongly suggest that splicing is also co-transcriptional in plants, and that the chromatin environment has a strong effect on RNAPII processivity to modulate the transcriptional and splicing dynamics of plant genes.

1.5.2.1 DNA methylation and regulation of alternative splicing

Plants exhibit extensive variation in DNA methylation and gene expression under different developmental and stress conditions (Dubin *et al.*, 2015; Chwialkowska *et al.*, 2016; Hossain *et al.*, 2017; Kawakatsu *et al.*, 2017; Lu *et al.*, 2017). In eukaryotes, DNA methylation occurs in symmetric CG and CHG (H = A, T or C) and asymmetric CHH contexts (Ehrlich *et al.*, 1982). However, DNA methylation is largely dependent on the CpG context in plants. In the

Arabidopsis genome, 24% of CG sites are methylated, compared with only 6.7% of CHG and 1.7% of CHH sites (Cokus *et al.*, 2008; Lister *et al.*, 2008). Interestingly, nucleosomal DNA is highly methylated, and exons rather than the introns are marked at the DNA level by high occupancy of nucleosomes. These are preferentially positioned at intron-exon and exon-intron boundaries in both mammals and Arabidopsis (Mavrich *et al.*, 2008; Schwartz, Meshorer and Ast, 2009; Chodavarapu *et al.*, 2010; M.-J. Liu *et al.*, 2015). Additionally, nucleosome occupancy is also lower in alternatively spliced exons compared with constitutively spliced exons (Nahkuri, Taft and Mattick, 2009; Schwartz, Meshorer and Ast, 2009; Tilgner *et al.*, 2009; Chen, Luo and Zhang, 2010; Gelfman *et al.*, 2013). Since DNA is packaged into nucleosomes, RNAPII elongation rate is inherently subject to frequent pausing at constitutively spliced exons with high GC levels (Churchman and Weissman, 2011; Shukla and Oberdoerffer, 2012), and regions of high nucleosome density slow down RNAPII to facilitate the recruitment of SFs to weaker upstream splice sites (Tilgner *et al.*, 2009; Chen, Luo and Zhang, 2010; Shukla *et al.*, 2011; Fong *et al.*, 2014).

An example of this is found in the honey bee, in which DNA methylation is almost exclusively found in exons with a strong correlation between methylation patterns on alternative exons and splicing patterns of these exons in workers and queens (Lyko *et al.*, 2010). Intriguingly, a reduction in methylation of the *dnmt3* gene encoding a methyltransferase via RNAi results in widespread changes in AS in honey bee fat tissues (Li-Byarlay, Li and Stroud, 2013). Additionally, a DNA-binding protein, CCCTC-binding factor (CTCF), promotes inclusion of weak upstream exons in the *CD45* gene by causing local RNAPII pausing in mammals. Methylation of exon 5 abolished CTCF binding and resulted in the complete loss of exon 5 from *CD45* transcripts (Shukla *et al.*, 2011). Interestingly, a direct link was very recently unveiled between DNA methylation and AS in humans by perturbing DNA methylation patterns of alternatively spliced exons. In this study, the authors used CRISPR-dCas9 proteins (for details, see the 'Engineering splicing variation' in chapter 5) and methylating/demethylating enzyme fusions (Shayevitch *et al.*, 2018). This work clearly demonstrates that changes in the methylation pattern of alternatively spliced exons mediates their inclusion, but has no effect on introns or constitutively spliced exons (Shayevitch *et al.*, 2018).

Recent work in plants demonstrated abundant DNA methylation and splicing variation under different growth and stress conditions, and during different developmental stages. For example,

quantification of AS in wild-type (WT) and *OsMet1-2* (CG methyltransferase mutant) rice lines revealed widespread differences in splicing variation (Wang *et al.*, 2016). Consistent with the metazoan data (Wang *et al.*, 2016), CG methylation was found to be higher in WT exons compared with adjacent introns, and was not solely dependent on the CG composition of exons and introns (Wang *et al.*, 2016). Further evidence from cotton showed similar CG methylation levels in constitutive and alternative exons, but variable patterns during different fibre development stages (M. Wang *et al.*, 2018). By contrast, CG methylation was higher in alternative introns than constitutive introns. Furthermore, differential CG methylation has a strong influence on nucleosome formation since constitutive exons displayed higher nucleosome occupancy than alternative exons. However, alternative exons exhibited higher nucleosome density than constitutive introns (M. Wang *et al.*, 2018). These findings clearly demonstrate that the relationship between DNA methylation and nucleosome occupancy is conserved between animals and plants, and AS is also predominantly regulated at the chromatin level in plants (Cramer *et al.*, 1999; Chodavarapu *et al.*, 2010; Luco *et al.*, 2011).

1.5.2.2 Histone remodelling modulates alternative splicing in plants

Since transcription by RNAPII is affected by chromatin structure, it is unsurprising that stress-induced chromatin modifications can affect co-transcriptional splicing outcomes in plants. To fully understand the influence of chromatin changes on co-transcriptional AS, stress-induced DNA methylation and histone modification should be considered inter-connected processes. Plant responses to environmental stress have been linked to modification of histone *N*-tails (Tsuji *et al.*, 2006; Zong *et al.*, 2013; Pajoro *et al.*, 2017). However, it is important to understand whether transcriptional regulation mediated by histone modifications can also influence AS. Indeed, emerging evidence indicates the role of single or combined histone marks in AS regulation in plants (Pajoro *et al.*, 2017; Wei *et al.*, 2018a). For example, PRMT5 methyltransferase (also known as SKB1) increases H4R3sme2 (histone 4 arginine 3 symmetric demethylation) levels in Arabidopsis to suppress the transcription of *FLC* and a number of stress-responsive genes (Deng *et al.*, 2010; Zhang *et al.*, 2011). Upon salt stress, SKB1 disassociation from chromatin results in a reduction in the cellular levels of H4R3sme2, resulting in the induction of *FLC* and salt stress-responsive genes through higher methylation of the small nuclear ribonucleoprotein Sm-like4 (LSM4) (Zhang *et al.*, 2011). In addition, *skb1* mutants display pre-mRNA splicing defects caused by reduced symmetric dimethylation of arginine in LSM4 (Zhang *et al.*, 2011). These results demonstrate that SKB1 alters the

methylation status of H4R3sme2 and LSM4 to link transcription and pre-mRNA splicing during stress responses. Additionally, PRMT5 also alters AS in the core clock gene *PSEUDO RESPONSE REGULATOR* 9 (*PRR9*), and influences clock functioning in Arabidopsis (Sanchez *et al.*, 2010). Similarly, recent evidence in rice indicates that histone H3K36-specific methyltransferase (SDG725) regulates IR events in many genes (Wei *et al.*, 2018a). These IR events are much more prevalent at the 5' end of gene bodies, and accompanied by high H3K36me2 histone marks, whereas the 3' end of gene bodies are associated with fewer IR events and minimal H3K36me2 accumulation (Wei *et al.*, 2018a). Furthermore, IR shifts along the ends of gene bodies are more significant when both H3K36me2 and H3K36me3 modifications occur simultaneously (Wei *et al.*, 2018a). In Arabidopsis, temperature-induced differentially spliced genes are enriched in H3K36me3 marks to induce flowering (Pajoro *et al.*, 2017). By contrast, depletion of H3k36me3 marks has the opposite effect to temperature-induced AS (Pajoro *et al.*, 2017). It is possible that plants remember temperature variation via H3k36m3 and associated splicing patterns to influence flowering. Taken together, these studies indicate that stress-induced specific changes in histone PTMs may alter the chromatin landscape to mediate AS patterns in plants. A model illustrating how histone PTMs may regulate AS in response to temperature is presented in figure 1.1
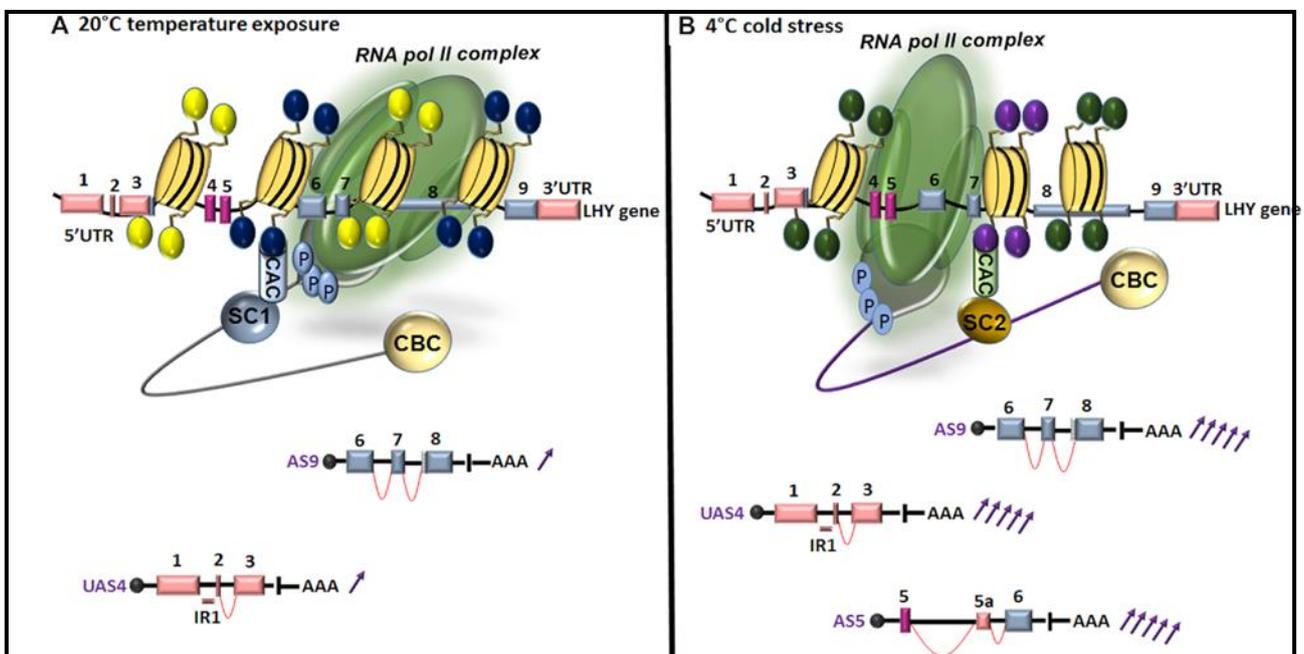
Figure 1.1. Schematic diagram illustrating proposed histone modifications and co-transcriptional splicing mechanisms in response to 22°C (A) and 4°C (B) using the LHY gene as an example. Temperature-dependent alternative splicing of the *LHY* gene generates different transcripts with variable abundance (purple arrows). For clarity, only a part of each splice variant is shown. At 4°C, both splice isoforms (UAS4 and AS9) are elevated from 10% (one arrow) to 50% (five arrows), and a new isoform (AS5) is produced (19). Under different temperatures,

nucleosome (yellow disks) enrichment with single or combined histone marks (yellow, dark blue, green and purple circles) may mediate the RNA RNAPII (green oval) elongation rate and subsequently the differential recruitment of splicing factors complex (SC1/2) through readers and chromatin-adaptor complexes (CACs) to modulate cold-specific splicing. Light blue circles labeled 'P' and the gray teardrop represent phosphorylated CTD. UAS4 represents an intron retention (IR1) event in the 5'-untranslated region (UTR). AS9 removes three nucleotides via an Alt3' in exon 8. AS5 adds an alternative exon 5a of 82 nucleotides via an alternative Alt3' and Alt5'. Exons are displayed as numbered boxes, introns as lines. Myb-encoding exons are purple, exons in the 5'/3'-UTRs and coding sequence are shown in pink and light blue, respectively. Gray circles and AAA represent the 7-methylguanosine cap and poly(A) tail, respectively. Red arcs represent the intervening sequence between 5'ss and 3'ss for different AS events.

1.6 Circular RNAs, R-loops and alternative splicing

Circular RNAs (circRNAs) were discovered more than two decades ago but were largely considered as splicing errors (Ye *et al.*, 2015). CircRNAs are generated by the so-called "backsplicing" of pre-mRNAs where a donor site is joined to an upstream splice acceptor site and could be derived from introns, exons or both regions (Memczak *et al.*, 2013; Ye *et al.*, 2015). CircRNAs are generated co-transcriptionally and compete with splicing and are strongly associated with ES in humans (Kelly *et al.*, 2015). Although functions of most CircRNAs are not known, some evidence points to their role in regulating the levels of microRNAs (P. Zhang *et al.*, 2017; Wilusz, 2018).

Plant CircRNA formation is mediated by developmental and environmental cues and modulate the expression or splicing patterns of their cognate genes (Zhao *et al.*, 2017; Pan *et al.*, 2018). Recent evidence shows that circRNAs are conserved in the diploid progenitors and modern polyploidy cotton varieties (Zhao *et al.*, 2017) indicating that some of the regulatory mechanisms/chromatin contexts may be similar in these varieties. On the other hand, heat stress alters genome-wide patterns of in Arabidopsis (Pan *et al.*, 2018), suggesting their role in stress response. Interestingly, circRNAs are usually spliced at canonical splice sites and predominantly from the same strand as the pre-mRNA; hence, they lack complementarity with it (Conn *et al.*, 2017). Therefore, it is unlikely that circRNAs derived from the sense strand could physically interact with its pre-mRNA, however, an interaction of circRNAs derived from antisense strands may still be possible but needs to be investigated. Alternatively, circRNAs could make DNA:RNA hybrids with the genomic DNA to make the so-called R-loop (Al-Hadid and Yang, 2016; W. Xu *et al.*, 2017). Indeed, circRNA derived from exon 6 of the SEPALLATA3 (SEP3) gene form an R-loop via direct interaction with the SEP3 locus (Conn *et al.*, 2017). The R-loop formation around exon 6 of the SEP3 gene results in skipping of this exon and affects petal and stamen number in Arabidopsis (Conn *et al.*, 2017). Until recently, R-loops were considered to be rare, but recent studies have shown that they are

abundant in yeast, mammals and Arabidopsis (Petrillo *et al.*, 2011; Skourti-Stathaki, Kamieniarz-Gdula and Proudfoot, 2014; Al-Hadid and Yang, 2016). R-loop formation is largely co-transcriptional (Chédin, 2017) and may affect transcription and splicing dynamics using circular or other RNAs. Due to their co-transcriptional nature, R-loops are associated with open chromatin structure and mostly active histone marks (H3K36me3, H3K4me2/me3 and H3K9Ac) (Chédin, 2017; W. Xu *et al.*, 2017).

CircRNAs and R-loop formation provide an additional regulatory mechanism to orchestrate chromatin changes. Since the chromatin environment is important to mediate transcription and splicing outcomes (Carrillo Oesterreich, Preibisch and Neugebauer, 2010b; Jimeno-González *et al.*, 2015), it is possible that this additional regulatory role in otherwise actively transcribing genes with optimum/normal RNAPII dynamics may serve as an additional switch to mediate splicing changes as demonstrated for the SEP3 gene (Chédin, 2017; Conn *et al.*, 2017; W. Xu *et al.*, 2017).

1.7 The epitranscriptome: a regulator of splicing variation

Chemical modification of RNAs, collectively referred to as the epitranscriptome, adds another layer of complexity to pre-mRNA splicing (Meyer and Jaffrey, 2014; Gilbert, Bell and Schaening, 2016). In mammals and plants, m^6A is the most abundant RNA modification, and is involved in the regulation of RNA processing (Zhong *et al.*, 2008; Luo *et al.*, 2014; Annita Louloupi, Evgenia Ntini, Thomas Conrad, 2018). In mammals, co-transcriptional m^6A deposition near splice sites promotes high splicing kinetics. However, high m^6A levels in introns are associated with slow RNAPII processivity and AS of nascent RNA transcripts (Annita Louloupi, Evgenia Ntini, Thomas Conrad, 2018). M^6A is also considered a post-transcriptional regulator of pre-mRNA splicing (Roundtree and He, 2016). In mammals, m^6A recruits the mRNA methylation reader YTHDC, which in turn recruits SR proteins to their corresponding binding sites (Roundtree and He, 2016). Additionally, m^6A facilitates recruitment of hnRNP C, a key player in pre-mRNA splicing, to regulate levels of alternatively spliced transcripts (Roundtree and He, 2016). In another study, the presence of TATA boxes was found to enhance the RNAPII elongation rate in humans (Slobodin *et al.*, 2017). This decreases the time window for recruitment and physical attachment of RNA N6-adenosine-methyltransferase-like 3 (METTL3; an enzyme that methylates adenosine residues of some RNAs) to RNAPII CTD, lowering m^6A modification of mRNAs (Slobodin *et al.*, 2017).

Interestingly, mRNAs with low m^{6}A levels displayed increased translation efficiency, which was not the case for m^{6}A-rich transcripts (Slobodin *et al.*, 2017).

In Arabidopsis, high-throughput annotation of modified ribonucleotides (HAMR) revealed that chemical modification of RNA differentially marks the vicinity around splice donor/acceptor sites of alternatively spliced introns within stable mRNAs (i.e. 3-methylcytidine) (Vandivier *et al.*, 2015). Recent 5'GRO-seq data from Arabidopsis showed that most gene promoters are strongly enriched in AT nucleotides, implying a role for TATA box-mediated transcription (Hetzel *et al.*, 2016). Although transcriptional regulation at the level of initiation is beneficial for plants by facilitating rapid responses under variable environmental conditions, additional control via RNA modification may be employed to dynamically control the fate of a given transcript. Therefore, it is tempting to speculate that co-transcriptional RNA modifications (m^{6}A or other marks), which are highly prevalent in plant mRNAs (Cui *et al.*, 2017; Vandivier and Gregory, 2018), may play a role in regulating splicing outcomes and the translational fate of different transcripts in plants (Figure 1.2). However, more robust methods and tissue/condition-specific profiling are needed to illuminate the mechanisms by which epitranscriptomic changes regulate splicing and the translational outcomes of fully spliced and AS transcripts.
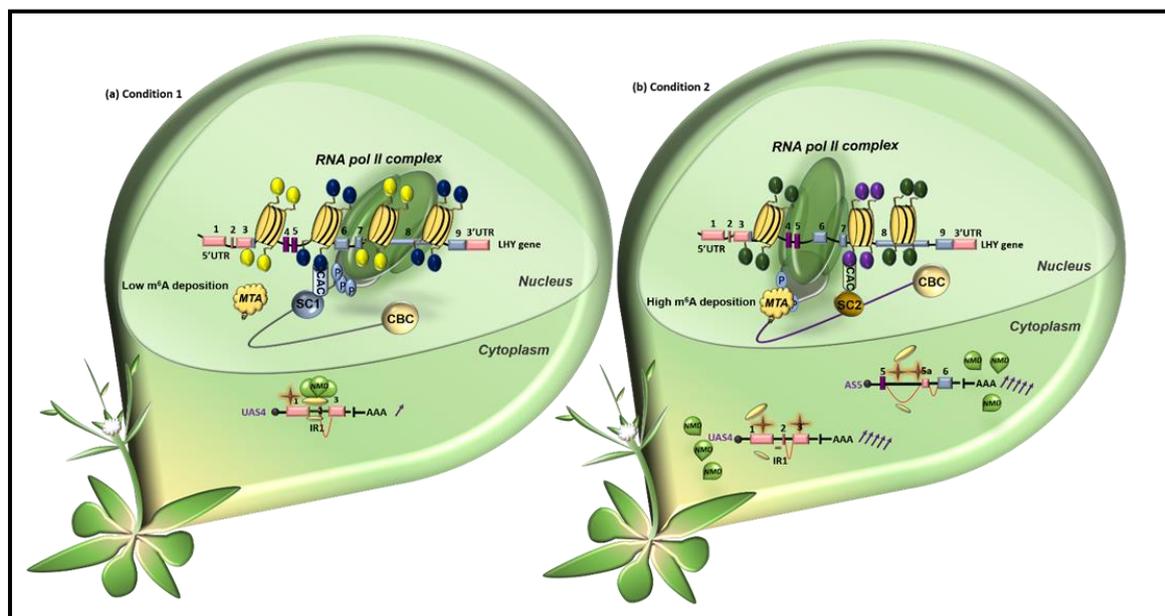
Figure 1.2. Model illustrating how condition-specific epigenetic marks may affect the rate of RNA RNAPII elongation, RNA base modification(s) and the fate of splice isoforms. Two NMD-sensitive splice isoforms of the *LHY* gene are used as hypothetical examples here. A fast RNAPII elongation rate disables methyltransferase (MTA) recruitment, resulting in low m^{6}A deposition (brown stars) over UAS4 (**A**). Slow RNA RNAPII elongation enables MTA recruitment and mediates high m^{6}A deposition over UAS4 and AS5 (**B**). Low m^{6}A deposition allows efficient ribosome (gold spheres) loading and facilitates NMD recruitment (A), whereas the opposite is true for

USA4 and AS5 in condition (B). Hence, condition-specific histone modifications (shown as yellow, dark blue, green and purple circles) and differential nucleosome occupancy (yellow disks) may regulate the RNA RNAPII elongation rate to assist NMD-sensitive transcripts (UAS4 and AS5) escape degradation. *LHY* splice variants UAS4 and AS5 display sensitivity to NMD only under certain conditions (19). The abundance of each transcript under different conditions and relative to each transcript within the same condition is denoted with purple arrows. For labels explanation, see Figure 1.1 legend.

1.8 Functions of Alternative splicing

AS regulates gene expression on different levels. It either affects mRNA stability, transport, and translatability or generates different protein isoforms with altered functions (F. Ding *et al.*, 2014; Y. J. Kwon *et al.*, 2014; S. A. Filichkin *et al.*, 2015; Yu *et al.*, 2016). In this section, the mechanisms by which AS affects gene expression at the mRNA and protein levels are described.

1.8.1 Alternative splicing regulates mRNA fate through NMD

In section 1.2, the various fates of alternatively spliced transcripts has been described, one of which is the RNA degradation system through NMD. NMD is a cytoplasmic RNA degradation system, which occurs on the first round of translation and through which AS regulates the abundance of alternatively spliced transcripts (Sato, Hosoda and Maquat, 2008). In plants up-frameshift (UPF)1, UPF2, UPF3, and SMG-7 orthologs proteins are the factors of NMD machinery that trigger transcript decay if one of the following features take place as a result of AS; (1) IR in the 3'UTR causing long 3'UTRs , (2) Splicing of 3'UTR introns can trigger NMD by creating a splice junction downstream of the stop codon (3) the presence of introns in the 3'UTR, (4) presence of PTC more than 50 to 55 nucleotides upstream of splice junction and upstream open reading frames (uORFs) (Kalyna *et al.*, 2012). In general, AS in the 5'UTR may cause loss of AUG or introduce uORFs thereby, resulting in transcripts sensitive or resistant to NMD (Kalyna *et al.*, 2012; Reddy *et al.*, 2013). Interestingly, in Arabidopsis, alternatively spliced transcripts generated by IR events and possessing NMD features were immune to NMD which may result in truncated proteins whereas, transcripts from the same gene with other type of AS events were substrate for NMD (Kalyna *et al.*, 2012). In Arabidopsis, 13-18% of protein-coding genes undergo AS-NMD and the abundance of NMD transcripts dramatically increases under some stress conditions indicating a functional importance in responding to various signals (Kalyna *et al.*, 2012). Overall, these findings show the importance of AS coupled to NMD in influencing gene expression levels through reducing the level of fully spliced mRNA as a result of the presence of NMD features in the alternatively spliced transcripts.

1.8.2 Alternative splicing regulates mRNA fate through microRNA-mediated mechanisms

The stability of mRNA can also be regulated through microRNAs (miRNAs) mediated mechanisms (Boutz *et al.*, 2007; Meng *et al.*, 2013). In eukaryotes, miRNAs attenuate gene expression post-transcriptionally through their base-pairing with complementary mRNAs for cleavage or translation inhibition (Wahid *et al.*, 2010). In the first instance, primary miRNAs transcripts, which are up to 3000 nucleotides are cropped to smaller transcripts denominated pre-miRNAs. Pre-miRNAs are then exported to the cytoplasm to be incorporated into argonaute 1-containing RNA-induced silencing complex; which are responsible of silencing mRNA targets (Wahid *et al.,* 2010). In plants, AS mediates the regulation of mRNA stability through miRNA in multiple ways (Yu, Jia and Chen, 2017). These include modulating the splicing of proteins involved in the biogenesis of pri-mRNA or pre-mRNA, as well as regulating the generation of splice variants containing or lacking miRNA binding sites. Interestingly in rice, miRNAs associated with AGO4 complexes were also proposed to be involved in DNA methylation in the nucleus (Wu *et al.*, 2010). This findings highlighted the potential role of miRNAs in regulating cleavage of pre-mRNAs intronic sequences. Indeed, a recent study in 40 and 1912 cleavage-based miRNA—intron interactions were detected in rice and Arabidopsis, respectively (Meng *et al.*, 2013). However, alternatively spliced isoforms of some rice genes lacking miRNA binding sites within the introns escaped the regulation by specific miRNA (Meng *et al.*, 2013). Cleaved introns have been shown to be processed into double-stranded RNAs and further processed into 21- and 24-nt phased small RNAs (Meng *et al.*, 2013). This study indicates the novel regulatory role of plant miRNAs in cleaving nuclear-localized, intron-containing pre-mRNA (Meng *et al.*, 2013). In mammals, miR-133 was shown to down-regulate the expression of neuronal homolog nPTB SFs and decreases the inclusion of PTB dependent exons during muscle development which establishes a role for microRNAs in the control of developmentally-specific splicing patterns (Boutz *et al.*, 2007).

Currently, deep sequencing technologies, miRNA microarrays, and quantitative real-time PCR analyses revealed that plants exposed to abiotic stresses such as drought, salinity, and temperature changes display altered expression of miRNAs implicated in plant growth and development in a stress-, tissue-, and genotype-dependent manner (Barciszewska-Pacak *et al.*, 2015). This is a critical step in plants gene expression control to repress negative regulators of

stress tolerance and allow the accumulation of stress-resistant proteins. For example, Arabidopsis seedlings treated with 24 hours of cold stress, 5 hours with salt stress, 10 hours of drought stress, and 3 hours of abscisic acid (ABA) treatment showed that miR393 was strongly induced by all four tested stress conditions, whereas miR389a.1 was inhibited by all of the stress treatments (Sunkar, 2004). Conversely, miR319 showed stress-specific responses to cold stress only but not to other treatments (Sunkar, 2004). Other examples of miRNA expression profiles among plant species rice, barley, maize, tobacco, peach show the importance of miRNAs in conferring plant tolerance to abiotic stresses (Detailed examples can be found in (Zhang, 2015)). The implication of microRNAs in regulating pre-mRNA splicing, DNA methylation, and stress tolerance point towards their importance as part of the splicing code. However, the mechanisms by which plants integrate microRNAs to refine their epigenetic splicing regulation under environmental stress are still unknown. Coupling of AS to NMD and miRNA control mRNA stability and abundance, and subsequently affect protein expression levels and functionality of genes involved in plant growth, development, and stress responses.

1.8.3 Regulation of Proteome Complexity by Alternative Splicing

As sessile organisms, plants exert a tight control over their gene expression patterns under normal and stress conditions to maximise carbon fixation and resource allocation efficiency to promote growth and fitness in the short and long term (Zhu, 2016). AS adds another layer of complexity to modulate transcriptome diversity (F. Ding *et al.*, 2014; Y.-J. Kwon *et al.*, 2014; S. A. Filichkin *et al.*, 2015) and potentially proteome complexity in a tissue- and condition-dependent manner (Marquez *et al.*, 2015; Yu *et al.*, 2016). It is well established that AS often allows fine-tuning of gene expression by changing the ratios of productive and unproductive variants (Reddy *et al.*, 2013; Hartmann *et al.*, 2016). However, limited data is available on the contribution of AS to protein diversity in plants (Yu *et al.*, 2016). Recent transcriptome and translatome data from humans suggest a significant contribution of AS towards protein diversity (Sterne-Weiler *et al.*, 2013; Floor and Doudna, 2016; Weatheritt, Sterne-Weiler and Blencowe, 2016; Yang *et al.*, 2016; Y. Liu *et al.*, 2017; Kahles *et al.*, 2018). However, relatively few alternative isoforms have been discovered in various proteomic studies that encode different proteins (Tress *et al.*, 2007, 2008; Brosch *et al.*, 2011; Abascal *et al.*, 2015; Tress, Abascal and Valencia, 2017). The scientific community is divided on this issue and some argue that poor sensitivity of Mass-Spectrometry (MS) techniques is a major limitation to detect changes in protein isoforms as a result of AS (See section 1.7) (Abascal *et al.*, 2015). On the other hand, it is also proposed that not all alternative isoforms are biologically

important, because alternative transcripts are generally a recent evolutionary innovation and under neutral selection (Tress *et al.*, 2008). Since limited proteomic data is available in plants, it is paramount to perform comprehensive proteomic studies in different tissues and in response to diverse stresses to illuminate the contribution of AS towards protein diversity and/or increasing regulatory capacity in plants. In addition, global analysis of translation patterns of splice isoforms needs to be studied in different tissues and stresses at multiple time points throughout the diurnal cycle.

Transcription and translation are energetically expensive (Gibon *et al.*, 2009), nonetheless plants exhibit a higher level of AS under stressful conditions (S. Filichkin *et al.*, 2015). This scenario poses potential problems, for example, if the aim is to diversify the proteome then why plants should invest in translation when photosynthetic capacity declines in stress conditions? Moreover, AS frequently generates PTC+ transcripts, which are degraded by the NMD pathway (Filichkin *et al.*, 2010; Filichkin and Mockler, 2012; Marquez *et al.*, 2012). NMD is a cytoplasmic mRNA quality control mechanism that targets newly synthesised capped transcripts harbouring NMD+ features during the pioneer round of translation (Lejeune, Ranganathan and Maquat, 2004; Maquat, Tarn and Isken, 2010). Interestingly, evidence from humans suggests that NMD is not restricted to the pioneer round of translation and could also be triggered for already translating mRNAs as a result of change in the cellular environment and/or needs (Durand and Lykke-Andersen, 2013; Rufener and Mühlemann, 2013). Among all AS events, IR is the most prevalent event in plants (Filichkin *et al.*, 2010; Marquez *et al.*, 2012). Most IR transcripts are predominantly sequestered in the nucleus under a particular stress or developmental stage for further processing upon cell requirement or degraded by the NMD pathway (Sun *et al.*, 2010; Gohring, Jacak and Barta, 2014; Hartmann, Wießner and Wachter, 2018; Wei *et al.*, 2018b). Some IR transcripts carry introns with features of protein-coding exons, which are termed as exitrons, and splicing of these exitrons affects protein functionality (Marquez *et al.*, 2015; Staiger and Simpson, 2015). Exitrons and other types of splice variants can often lead to the formation of Intrinsically disordered proteins or regions (IDPs/IDRs) (Johnson *et al.*, 2007; Marquez *et al.*, 2015). IDPs and IDRs lack fixed three-dimensional structure due to their amino acid composition, which prevents appropriate hydrophobic region formation (Oldfield and Dunker, 2014). Importantly, variation in the three-dimensional structure of proteins, as a result of AS and PTMs, results in the diversification of substrate specificity and enhanced regulatory capacity (Buljan *et al.*, 2012; Niklas *et al.*, 2015; Strom *et*

al., 2017; Niklas, Dunker and Yruela, 2018).

Although AS coupled to NMD plays a major role in regulating the Arabidopsis transcriptome (Drechsel *et al.*, 2013) and potentially protein levels, however, most of the PTC+ transcripts (IR and others) if translated, would produce truncated proteins (Figure 1.3) and create a very toxic environment to carry out the normal activity of the cell (Brogna, McLeod and Petric, 2016). The efficiency of NMD during and after the pioneer round of translation is robust and most PTC+ transcripts are rapidly degraded upon their arrival in the cytoplasm (Durand and Lykke-Andersen, 2013; Rufener and Mühlemann, 2013; Trcek *et al.*, 2013). Intriguingly, NMD responses are dampened in both mammals and plants under stress conditions and this strategy may facilitate an appropriate response via translating some of the stress-responsive genes and splice variants (Shaul, 2015).
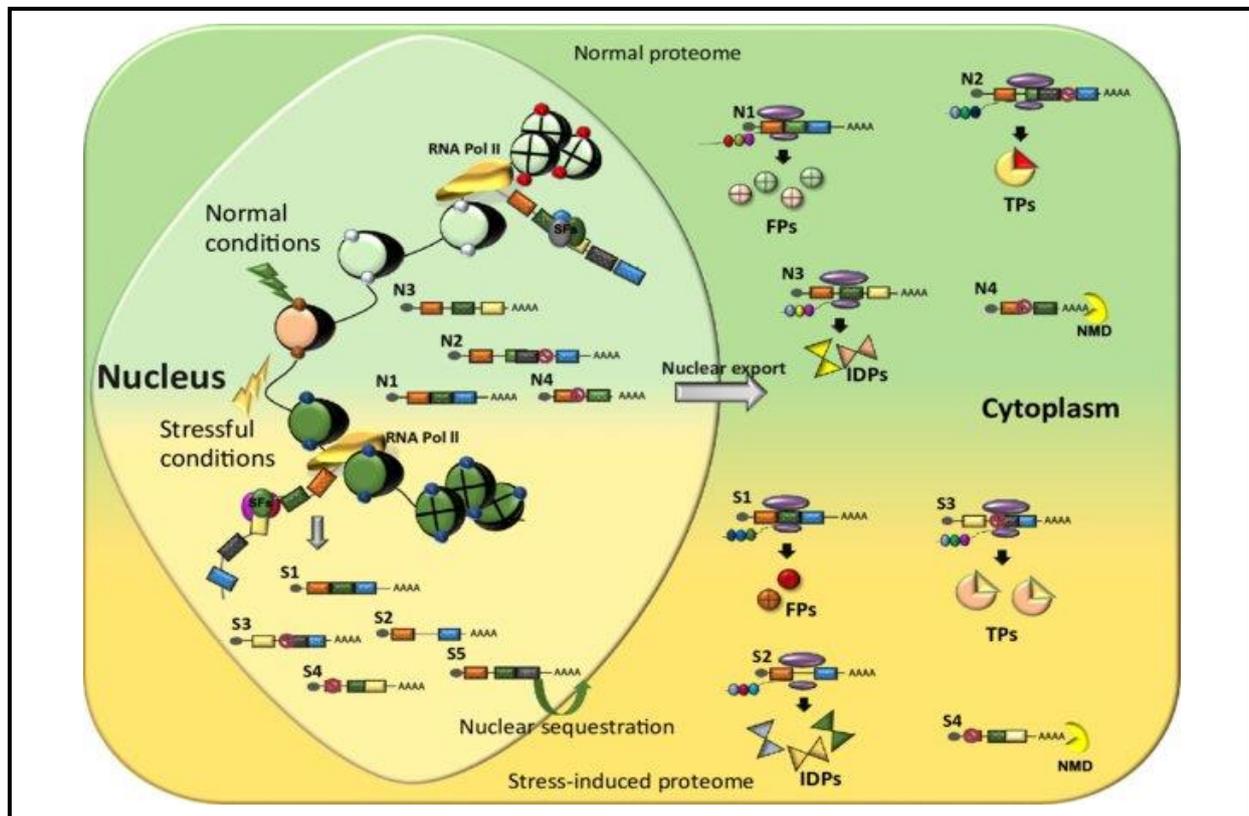
Figure 1.3. Hypothetical schematic diagram showing fates of alternatively spliced transcripts under normal and stress conditions in plants. Alternative splicing generates multiple transcripts under normal (N1–N4) as well as stress (S1–S5) conditions. Constitutively spliced transcripts (N1 and S1) and alternatively spliced PTC transcripts (N3 and S2) are translated into functional protein isoforms (FPs) and intrinsically disordered proteins (IDPs). Alternatively spliced PTC + transcripts (N2, N4, S3, and S4) are either degraded via the nonsense-mediated decay (NMD) pathway (N4 and S4) or escape NMD (S3) to generate truncated proteins (TPs). Although present in both conditions, FPs are more abundant under normal conditions, whereas TPs and IDPS constitute the majority of stress-induced proteome.

The proposed theory here is that under initial episodes of stress conditions, plants buffer against normal protein synthesis level via AS to decrease translation of a significant proportion of the transcriptome and produce the protein isoforms needed for adaptation to stresses. This strategy may allow plants to reduce their metabolic cost but also maintain a sufficient level of regulatory capacity via inclusion of alternative and disordered domains in stress-responsive proteins through AS. Although mechanistic details of such a process are not available in any organism at the moment, however, supporting evidence has just emerged from yeast. Two independent studies using yeast as a model have revealed that introns mediate fitness under stress conditions (nutrient starvation) by repressing ribosomal protein genes (for details see below) (Morgan, Fink and Bartel, 2018; Parenteau *et al.*, 2019). In addition, AS may not only diversify the regulatory capability of plant genes during initial stress episodes but also mediate crosstalk between a given metabolic state and protein diversity/abundance to cope with stressful conditions in the long term. Epigenetic modifications in plants such as DNA methylation and histone modifications define an epigenetic code that translates environmental stresses into an epigenetic footprint affecting cellular signalling network, and could also be recreated upon a recurring stress in the same or future generations (Lang-Mladek *et al.*, 2010). In this way, AS may also be involved in stress memory mediated by epigenetic codes (Lämke and Bäurle, 2017; Ling *et al.*, 2018) and only after repeated onsets of similar stresses, plants could employ AS to generate more protein diversity or preserve the regulatory control in the long term (Niklas *et al.*, 2015; Niklas, Dunker and Yruela, 2018).

1.8.4 AS and IDPs/IDRs: A Way to Regulate Plants Environmental Fitness

Intrinsically disordered proteins or regions were termed as the junk proteome, however recent evidence shows they control important cellular functions via transcriptional regulation, cell cycle, chaperone formation and enrichment of regulatory capacity especially under stress conditions (Figure 1.3) (Dunker *et al.*, 2013). Interestingly, highly conserved enzymes are normally not enriched in IDRs, whereas multifunctional enzymes contain disproportionately long IDRs (Niklas, Dunker and Yruela, 2018). Additionally, most eukaryotic proteins involved in transcription and RNA processing exhibit strong enrichment in IDRs that function in the formation of membraneless organelles in cells such as nuclear speckles, heterochromatin domains, stress granules and processing bodies (Minezaki *et al.*, 2006; Strom *et al.*, 2017; Rai *et al.*, 2018). Interestingly, stress granules can sequester and protect both RNAs and proteins from stress-induced damage (Chavali, Gunnarsson and Babu, 2017; Riback *et al.*, 2017) and

alter signaling pathways during stress as shown for mammalian/mechanistic Target of Rapamycin Complex 1 (mTORC1) (Wippich *et al.*, 2013). Recent data from two yeast studies demonstrate that introns are essential to promote resistance to stress conditions via the nutrient sensing TORC1 pathway (Morgan, Fink and Bartel, 2018; Parenteau *et al.*, 2019). In the first study (Parenteau *et al.*, 2019), introns were found to be essential to downregulate ribosomal protein genes (RPGs) under starvation conditions to promote fitness in the wild type strains. Conversely, intron-deletion strains failed to survive under these conditions due to upregulation of RPGs and respiration-related genes, resulting in uncontrolled growth and starvation (Parenteau *et al.*, 2019). Intriguingly, excised introns, which are rapidly degraded under nutrient-rich conditions, accumulate as linear RNAs under stress conditions (Morgan, Fink and Bartel, 2018). In the second study, deletion of these unusual spliceosomal introns via the CRISPR-Cas9 system resulted in higher growth via TORC1 mediated stress response as well (Morgan, Fink and Bartel, 2018). The presence of intron-mediated regulation of growth response in a eukaryote (yeast) is remarkable and it is tempting to speculate that similar mechanism exists in higher eukaryotes like plants, for at least, a subset of growth and stress-responsive genes.

Biased distribution of nucleotides at splice junctions (SJs) is important for spliceosome recognition, however, most nucleotides at SJs and among *cis*-regulatory elements, code for disorder-promoting amino acids (Lysine, Glutamic acid and Arginine) (Smithers, Oates and Gough, 2015). Interestingly, exonic splicing enhancers are more prevalent in exons encoding disordered protein regions compared to exons associated with structured regions in many taxa including plants (Smithers, Oates and Gough, 2015). Since most protein segments affected by AS are often intrinsically disordered, these likely confer additional regulatory capacity by not only changing the three-dimensional structure but also their PTMs to further diversify their function and substrate specificity in different cells under biotic and abiotic stress conditions in plants (Buljan *et al.*, 2012; Niklas *et al.*, 2015; Strom *et al.*, 2017; Niklas, Dunker and Yruela, 2018). In general, the human proteome is more disordered, however genes involved in environmental responses are significantly more disordered in Arabidopsis (Pietrosemoli *et al.*, 2013). It is possible that the scheme of regulation via IDPs-AS-PTM is more relevant in plant species due to the prevalence of AS under stress conditions where a fine balance between photosynthesis, resource allocation, and acclimation response needs to be generated for adaptive responses and survival (Bah and Forman-Kay, 2016; Niklas, Dunker and Yruela, 2018). Under stress, plants display re-arrangement of their chromatin structure, which might

also affect co-transcriptional splicing outcomes and differential splice site selection and increase AS diversity (Ullah *et al.*, 2018). Recently, it has been shown that in addition to a regulatory role, IDPs play a central role in organisation and assembly of many macromolecular membraneless organelles including speckles, processing bodies, stress granules and chromatin domains (Pietrosemoli *et al.*, 2013; Oldfield and Dunker, 2014; Rai *et al.*, 2018). Consequently, IDPs might be a result of this stress-dependent chromatin modulation to help plants adapt in the short term. Stress- and stage-dependent IDPs can explain how the environment is capable of modulating the three-dimensional structure and PTMs of their proteins via AS. Hence, it is possible that IDPs provide condition-specific and enhanced regulatory network of transcriptional, splicing and translational regulators, and chaperones required for fine-tuning gene expression and refining the proteome in a given tissue under stressful conditions (Figure 1.3). It has been proposed IDPs with AS and PTMs significantly contribute to the diversification of protein function and may also buffer against undesirable changes (Niklas, Dunker and Yruela, 2018). Furthermore, the presence of disordered regions in non-structural domains can aid neo-functionalization by evading the selection pressure that a protein with an altered structural domain would experience (Niklas *et al.*, 2015; Niklas, Dunker and Yruela, 2018).

Plants employ their internal, 24-hour timer, the "circadian clock", to synchronize daily activities to predictable changes in the environment (Millar, 2016), which provides a competitive advantage and maximizes productivity (Seo and Mas, 2015). Evidence from previous studies shows that photosynthesis and starch synthesis rates during the day and resource mobilization to fuel growth during the night are tuned by the plant clock but are also dependent on the length of the photoperiod and growth in the previous night (Graf *et al.*, 2010). A prominent mechanism for clock control of physiological pathways is via the rhythmic regulation of RNA accumulation (Millar, 2016), including regulated AS of RNA synthesis late
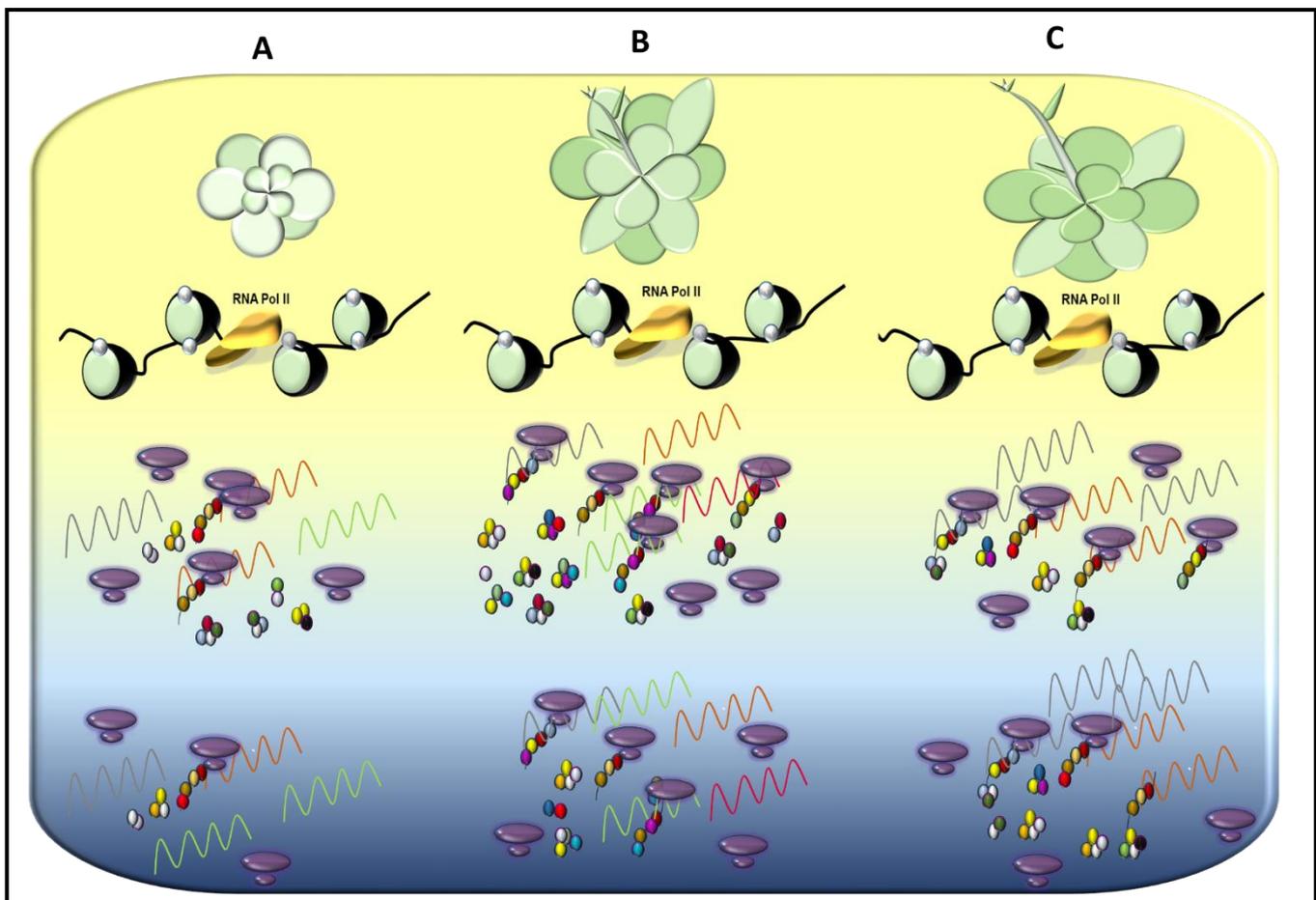
Figure 1.4. Translational coincidence upon photoperiod length and long-term changes. Under long photoperiods (day-time represented by yellow colour), plants translate a higher proportion of their transcriptome to produce more proteins, to support a higher degree of metabolic activity. However, under a short photoperiod (evening and night-time represented by light and dark blue colour, respectively), ribosome loading and translational efficiency are reduced as a result of lower demand. In this way, plants may modulate their proteome using the same transcriptomic pool upon varied physiological needs. Moreover, during different growth stages (A-B-C), the relationship between transcript abundance and protein diversity may not be linear to maintain desirable cost to benefit ratio and regulatory capacity.

in the day (Figure 1.4) (Allan B James, Syed, Bordage, *et al.*, 2012; Filichkin and Mockler, 2012; Schmal, Reimann and Staiger, 2013; S. Filichkin *et al.*, 2015; Seaton *et al.*, 2018). Thousands of plant genes show rhythmic expression, with peaks across the day and night. These RNA rhythms (for mostly higher metabolic activity genes associated with photosynthesis, primary/secondary metabolism and pigment biosynthesis) interact with the photoperiod, where translation rate is higher during the light interval than in darkness (Piques *et al.*, 2009; Seaton *et al.*, 2018). Plants combine transcript rhythms and translational regulation to tune protein expression in different photoperiods, *via* a mechanism called "Translational coincidence". For RNAs peaking late in the photoperiod, the higher ribosome loading in the light interval only coincides with high mRNA levels during longer photoperiods. If the photoperiod ends before the RNA level rises, daily protein synthesis might, therefore, be lower. One way to increase levels of a protein under long photoperiods, as in summer, is to time a rhythmic peak of RNA synthesis late in the day (Figure 1.4) (Seaton *et al.*, 2018). Arabidopsis proteome analysis in different photoperiods revealed that enzymes involved in primary/secondary metabolism and photosynthesis were more abundant and plants show higher metabolic activity under longer photoperiods (Seaton *et al.*, 2018). Hundreds of proteins with rhythmic RNAs peak late in the day were present at higher levels in these long photoperiod conditions, whereas proteins with morning-peaking RNAs were more abundant in short photoperiods.

Since the timing of expression of a particular gene can influence its translation patterns, it is logical to ask whether the same relationship holds true for alternatively spliced transcripts. Indeed, light conditions regulate AS of SR30 pre-mRNA, which encodes a serine/arginine-rich protein involved in RNA splicing in Arabidopsis, and influence their translation patterns (Hartmann, Wießner and Wachter, 2018). One of the splice variants of SR30 (SR30.1) is rapidly generated upon exposure to light and exported to the cytoplasm for translation as evident from the abundance of SR30.1 protein (Hartmann, Wießner and Wachter, 2018). In contrast, another splice variant, SR30.2 only appears in dark-grown seedlings and is enriched in nuclear fractions with poor representation among ribosome-associated transcripts. Interestingly, global analysis of AS in Arabidopsis etiolated seedlings exposed to different wavelengths of light revealed that many events switch from probably unproductive variants in darkness to productive variants in light during seedling photomorphogenesis (Hartmann *et al.*, 2016). Similarly, RS31 gene encoding another serine/arginine-rich splicing factor in Arabidopsis produces three isoforms under light conditions (Petrillo, Godoy Herz, Fuchs, *et al.*, 2014). Of these, mRNA1 codes for

the full-length protein and mRNA2 and mRNA3 are retained in the nucleus (Petrillo, Godoy Herz, Fuchs, *et al.*, 2014). Interestingly, mRNA1 abundance considerably decreases under dark conditions without a significant drop in RS31 transcripts. Transgenic lines overexpressing mRNA1 show no phenotype under 16 and 8 hours of light and dark conditions, respectively, however result in yellowish and small seedlings under dark or low light intensity compared with WT or RS31 mutants as a result of lower levels of chlorophylls a and b (Petrillo, Godoy Herz, Fuchs, *et al.*, 2014). Interestingly, plants treated with a drug that blocks electron transfer from photosystem II to the plastoquinone pool, mimics the effect of darkness on RS31 AS, indicating that a retrograde signal travels from the chloroplast to the nucleus. These data suggest that down-regulation of mRNA1 under dark conditions via AS is crucial for normal growth and development of Arabidopsis plants under changing light conditions. Importantly, signals from chloroplast controlling nuclear events and a complex mechanism like AS is intriguing and indicates that environmental condition can influence gene regulatory mechanisms to confer plant fitness. However, it is notable that such crosstalk may take a long time to develop, considering the evolutionary history of chloroplasts and photosynthetic systems (Xiong and Bauer, 2002; Baena-González *et al.*, 2007). Alternative splicing of SR30 and RS31 genes can serve as a powerful model to understand why some splice variants appear only under variable environmental conditions and translated or retained in the nucleus. Additionally, these results support the notion that the metabolic state of a plant is closely regulated under different photoperiods and/or stress conditions, in part by altering which fraction of the transcriptome would be translated. Since AS transcripts are more abundant under stress condition, plants must tightly control what mRNA species will be translated to keep the metabolic cost of protein synthesis down (Piques *et al.*, 2009; Ishihara *et al.*, 2017). It is therefore not surprising that a significant proportion of AS transcripts (IR) are either sequestered in the nucleus or degraded via the NMD pathway. Furthermore, since plants exhibit more protein translation under longer photoperiod (optimum energy supply) (Seaton *et al.*, 2018), hence, fewer proteins (mostly IDPs) derived via AS under stress (limited energy supply) may become a preferred choice to maintain essential regulatory control with minimum energy cost. Clearly, further work using ribosomal foot-printing and/or Mass Spec (See section 1.7) techniques needs to be done to illuminate this phenomenon (Mustroph *et al.*, 2009; Juntawong *et al.*, 2014).

1.9 Limitations to Detect Alternative Isoforms at the Proteome Level

In the shotgun proteomic analysis, proteins are first digested proteolytically into smaller peptides using trypsin and subsequently analysed by LC-MS/MS (Olsen, Ong and Mann, 2004). Trypsin, the most common enzyme used in Mass Spec cleaves at the C-terminus of lysine or arginine to produce peptides with optimal length and charge (Olsen, Ong and Mann, 2004). Peptides spanning exon-exon junctions provide direct evidence of splice variants at the protein level. Interestingly, lysine and arginine are the most enriched amino acids at exon-ending or exon-exon junctions of transcripts (X. Wang *et al.*, 2018). Exon-exon junctions are preferred sites for trypsin digestion, hindering detection of junction-specific peptides and identification of novel alternative splicing peptides in the proteo-genomics analysis (Ning and Nesvizhskii, 2010; Sheynkman *et al.*, 2013; Wang, Zhang and Wren, 2013). To overcome trypsin digestion limitations, specificity of five proteases including Lys-C, Glu-C, chymotrypsin, Asp-N, and Arg-C was evaluated recently (X. Wang *et al.*, 2018). Among these five enzymes, the highest number of detectable junctions including exon-ending and exon-exon junctions were observed in chymotrypsin digestion, making it a protease of choice in LC-MS/MS studies, especially to predict RNA splicing derived peptides (X. Wang *et al.*, 2018). Since different protein isoforms of the same gene may be localized in different tissues conferring diverse physiological outcomes, it would be useful to improve the sensitivity of current proteomic analysis methods. Alternatively, ribosome profiling/foot-printing along with next-generation sequencing (NGS) (Ribo-Seq), can be employed as an alternative strategy to use ribosome bound transcripts as a proxy for translation (Juntawong *et al.*, 2014; Ingolia, 2016). However, foot-printing data should be treated with caution as ribosome bound transcripts may not be translated as a result of ribosomal scrutiny during the pioneer round of translation (Inada, 2017). In the future, quantitative Ribo-Seq and proteomic data from multiple tissues in the context of RNA-metabolism, degradation, and other features may help to improve the efficiency to detect translated transcripts.

1.10 conclusion

A growing body of evidence acquired in recent years suggest that co-transcriptional splicing regulation mediated by epigenetic mechanisms occurs in both animals and plants. In particular, RNAPII initiation and elongation speed mediate the co-transcriptional processing of pre-mRNAs, and modulate the abundance of constitutive and AS transcripts in animals and plants. In plants, DNA methylation and epigenetic modifications regulate splicing patterns of pre-mRNAs of some genes. Although a direct link between epigenetic modifications and AS in plants is yet to be established, emerging epigenetic engineering approaches should address this in the future. Further work is needed to illuminate the complex regulatory mechanisms controlling splice isoform ratios in a cell-type and condition-specific manner (Figure 1.5).
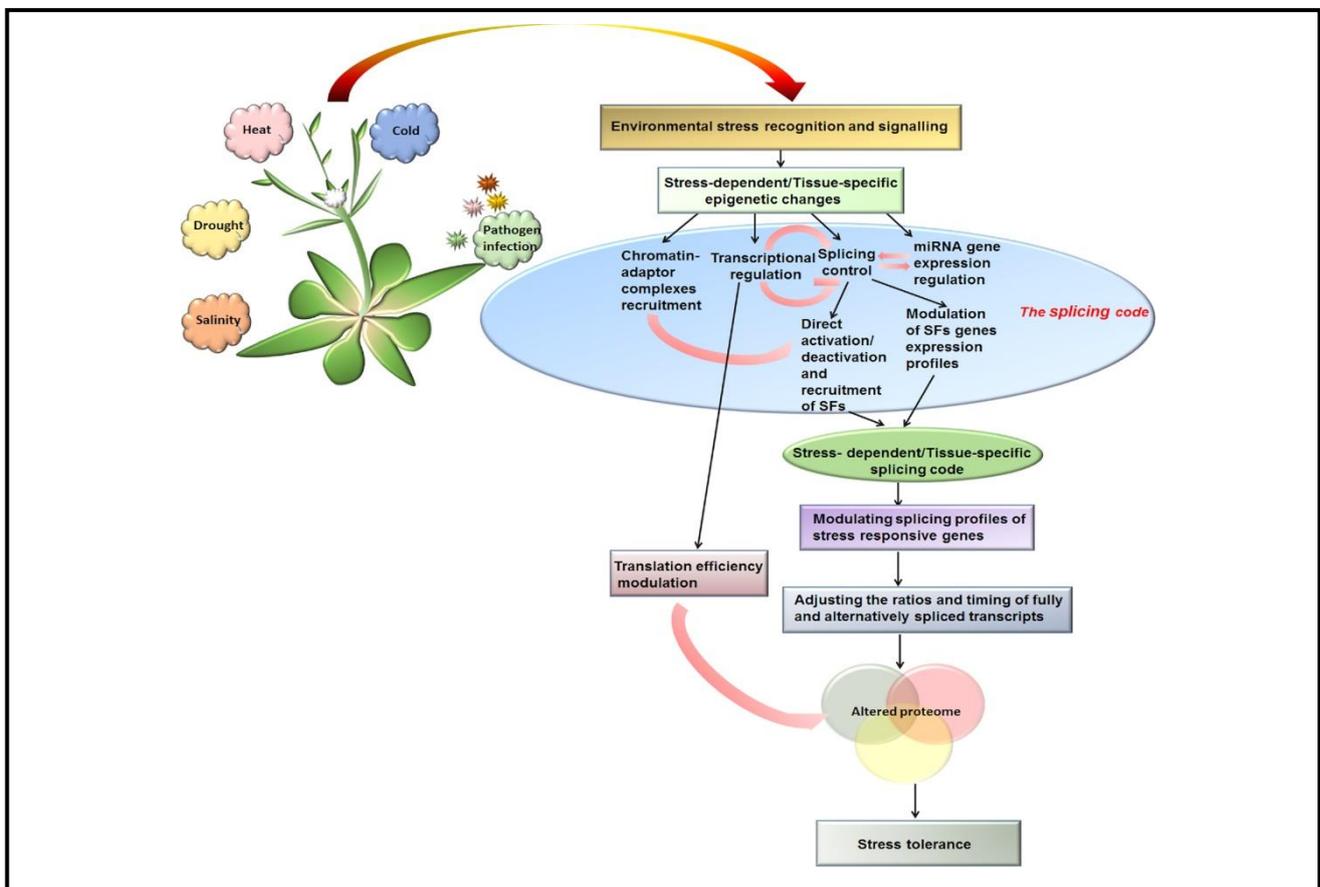
Figure 1.5. Schematic diagram showing how the stress-induced splicing code may promote stress tolerance. Variable environmental conditions alter chromatin structure, regulating transcriptional and splicing dynamics and modulating the expression of stress-responsive genes. Stress-induced epigenetic modifications result in a condition-specific splicing code through the differential recruitment of chromatin-adaptor complexes and/or micro RNA (miRNA) regulation. The stress-specific splicing code can fine-tune the expression of target genes by adjusting transcript ratios and timing, triggering appropriate changes in transcriptome and proteome composition, thereby conferring adaptive responses under stress conditions.

The next steps are to determine how the splicing code is 'built' from epigenetic and epitranscriptomic modifications, and reveal how it can modulate (i) the timing required to

process different pre-mRNAs in an RNAPII speed-dependent manner and (ii) the ratios of fully and alternatively spliced transcripts to produce the desirable transcriptome under different conditions. To help answer these and other questions, the translation efficiency of alternatively spliced transcripts must be determined, and how plants fine-tune their proteome at co/post-transcriptional levels must be revealed, as well as translational/post-translational levels, by directing their transcripts to NMD or nuclear retention. It would also be useful to investigate how RNA methylation patterns are established and preserved after pre-mRNA synthesis and maturation into mRNAs in plants. Addressing these questions will undoubtedly expand our understanding of the chromatin code in plants.

All life forms need to orchestrate their transcriptome patterns to produce an appropriate response under normal and stress conditions. However, plant transcriptomes need to promote efficient carbon fixation and its utilization during the diurnal cycle at different growth and developmental stages. Therefore, it is intriguing that plants generate more splicing variation under stress conditions to fine-tune their gene expression patterns. It is therefore unlikely that plants would produce more proteins under limited energy supply (Walley *et al.*, 2016; G. Xu *et al.*, 2017). Additionally, AS transcripts can produce nonsense transcripts and would result in truncated proteins if translated (Sato, Hosoda and Maquat, 2008; Palusa and Reddy, 2010; Kalyna *et al.*, 2012; Trcek *et al.*, 2013)**.** Similarly, most IR transcripts, if translated, would produce proteins with IDRs and may not confer any specific function. However, most IR transcripts are trapped in the nucleus and thus remain untranslated (Gohring, Jacak and Barta, 2014). Therefore, plants employ AS to not only alter their transcriptional response but also to influence proteome composition via sequestration of intron-containing RNAs and other alternatively spliced transcripts. It is also possible that similar to yeast (Parenteau *et al.*, 2019), plant spliceosomal introns also play regulatory roles under stress conditions, however further work is needed to illuminate this phenomenon. Alternatively, plants may generate additional regulatory capacity via translating some of the AS transcripts that harbour IDRs in different TFs including clock genes, and splicing factors to confer enhanced regulatory capacity to interact with multiple partners, enzymes and their substrates (Dunker *et al.*, 2013; Pietrosemoli *et al.*, 2013; Niklas *et al.*, 2015; Niklas, Dunker and Yruela, 2018). This is reminiscent of *Down syndrome cell adhesion molecule* (*Dscam*) protein, which is required for neuronal connections in drosophila (Wojtowicz *et al.*, 2004). *Dscam* gene can generate thousands of splice isoforms. Although, all splice isoforms share the same domain, variable amino acids with in the immunoglobulin (Ig) domains confer binding specificity and contribute to complex neuronal

wiring (Wojtowicz *et al.*, 2004; Hattori *et al.*, 2007). In this way, isoform diversity provides each neuron with a unique identity to facilitate self-recognition, which is essential for neuronal wiring in drosophila (Wojtowicz *et al.*, 2004; Hattori *et al.*, 2007).

It logical to postulate that AS may increase regulatory capacity in the short term but only contributes to protein diversity in the long term when different combinations have been tried over many generations and purifying selection has taken its course (Kovacs *et al.*, 2010; Smithers, Oates and Gough, 2015; Niklas, Dunker and Yruela, 2018). A recent study showed that plants possess splicing memory for heat stress and only previously primed plants with heat stress show a predicted AS response to the same stress again (Ling *et al.*, 2018). This short-term AS memory may be engendered through specific chromatin marks that in turn give birth to long-term adaptations mediated by chromatin landscape. This strategy provides spatiotemporal order and reproduction of a specific AS pattern under a similar condition, tissue and/or developmental stage (Lämke and Bäurle, 2017). Since chromatin state also mediates transcription and splicing dynamics (Luco *et al.*, 2011; H. Liu *et al.*, 2018; Ullah *et al.*, 2018), chromatin environment may not only mediate specific AS outcomes but could also serve as an epigenetic footprint to trigger a comparable response in the event of a similar stress in the future (Lämke and Bäurle, 2017; H. Liu *et al.*, 2018; Ling *et al.*, 2018). Understanding the transcriptional and translational dynamics of different AS transcripts in concert with associated chromatin marks, in different photoperiods and environmental conditions will be fruitful to understand the impact of AS on the alternative proteome. To fully appreciate the role of AS in gene regulation and protein diversity, not only the chromatin context in which different AS patterns appear in the short and long term need to be understood but also look at their partners by using yeast hybrid system and modified MS and LC-MS techniques in a tissue and condition-specific manner among diverse populations and under different conditions.

1.11 Glossary

Alternative Splicing: A gene regulatory mechanism that produces different messenger-RNAs (mRNAs) from a single gene via inclusion and/or exclusion of exons or introns fully or partially in different transcripts.

Mass-spectrometry (MS): An analytical technique to identify small molecules and macromolecules (including proteins) on the basis of mass to charge ratio.

Liquid chromatography-MS (LC-MS): A technique that combines the power of liquid chromatography for sample ionization/physical separation with MS.

Intron Retention: An alternative splicing event that retains an intron in the transcript.

Intrinsically Disordered Proteins: Proteins that lack well-defined globular three-dimensional structures and frequently interact with or function as hubs in protein interaction networks.

Intrinsically Disordered Region: Some proteins completely disordered, whereas others only harbour disordered sequences, referred to as intrinsically disordered regions (IDRs).

Translational Coincidence: Differences in the rates of protein synthesis across photoperiods that explain the changes in the coincidence of rhythmic RNA expression with light resulting in higher rates of translation.

Photosystem II: First protein complex located in the thylakoid membrane of chloroplasts that uses energy from sunlight to extract electrons from water molecules.

Plastoquinone: Carriers of electrons in Photosystem II that establish the electron transport chain during photosynthesis.

GRO-seq: Global run-on sequencing is a technique in which actively transcribing nascent RNAs are sequenced using next-generation sequencing platforms.

pNET-seq: Plant native elongating transcript sequencing is a technique that involves isolation of the 3' ends of actively transcribing genes via immunoprecipitation of the RNA polymerase II complex, to precisely map RNAPII position and is followed by next-generation sequencing.

Ser2(5)P CTD: CTD of the RNA polymerase II is dynamically phosphorylated during transcription via different phosphorylation patterns that help recruit required mRNA processing

and histone modifying factors. Serines 2 (Ser2) and Ser5 are major phosphorylation sites in the CTD domain.

CRISPR-Cas9 system: CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9) is a naturally occurring bacterial derived genome editing system. CRISPR-Cas9 system allows insertion and deletion of genomic regions with greater precision than previously available methods.

1.12 Aims and objectives of this study

Previous data have shown that (1) AS is largely co-transcriptional in Arabidopsis, (2) epigenetic modifications (i.e: DNA methylation and nucleosome occupancy) have a strong influence in regulating AS events in both plants and mammals (Listerman, Sapra and Neugebauer, 2006; Khodor *et al.*, 2011; Luco *et al.*, 2011; Wang *et al.*, 2016; Ullah *et al.*, 2018; Jabre *et al.*, 2019), and (3) both AS and epigenetic modifications can modulate the expression of genes involved in stress responses. The hypothesis of this study is that dynamic changes in the chromatin landscape in response to stress (cold in my case) may provide a scaffold around which gene expression and the AS patterns are orchestrated. Many studies in Arabidopsis have shown the role of AS in modulating the transcriptome under cold stress or the effect of environmental stresses on chromatin re-arrangement to modulate gene expression. However, no studies have shown how changes in the chromatin landscape upon cold stress could modulate AS profiles. Therefore, I wanted to develop a system in which the effect of the epigenetic context could be evaluated on AS variation in an identical genetic background to remove the cofounding effects that may be associated with the sequence variation. I employed bisulphite sequencing, nucleosome occupancy and RNA-seq analyses to understand the relationship between DNA methylation, nucleosome occupancy and splicing variation. To the best of my knowledge, this is the first study in Arabidopsis that investigates the influence of epigenetic mechanisms on AS under cold stress in Arabidopsis plants having different epigenetic landscapes but identical DNA sequence. The importance of understanding the role of epigenetic features in gene expression and AS regulation in Arabidopsis has three facades: (1) Understand the epigenetic mechanisms by which plants adapt to their stressful environment, (2) Development of better adapted crop via epigenetic mean rather than creating genome changes, (3) Attract other scientist to investigate how these epigenetic mechanisms changes in responses to other environmental stress.

Publications

1. **Jabre I**, Reddy ASN, Kalyna M, Chaudhary S, Khokhar W, Byrne LJ, Wilson CM and Syed NH. Does co-transcriptional regulation of alternative splicing mediate plant stress responses? Nucleic Acids Research 47: 2716-2726.

2. Chaudhary S, **Jabre I**, Reddy ASN, Staiger D and Syed NH. Perspective on alternative splicing and proteome complexity in plants. Trends in Plant Science. 214: 496:506 (First Joint author)

3. Chaudhary S, Khokhar W, **Jabre I**, Reddy ASN, Wilson CM and Syed NH. Alternative splicing and proteome diversity: humans versus plants. Frontiers in Plant Science 2019. DOIoi: 10.3389/fpls.2019.00708 (First Joint author)

4. Khokhar W, Hassan M, Reddy ASN, Chaudhary S, **Jabre I**, Syed NH. Comprehensive identification of splicing QTLs in diverse ecotypes of Arabidopsis thaliana. Frontiers in Plant Science DOI:10.3389/fpls.2019.01160 (In press).

5. **Ibtissam Jabre**[1], Waqas Khokhar[1], Wenbin Guo[2], Maria Kalyna[4], Anireddy S N Reddy[5], Ezequiel Petrillo[6], Weizhong Chen[3], Saurabh Chaudhary[1], Runxuan Zhang[2], Lee J Byrne[1], Carol Trim[1], Cornelia M Wilson[1] and Naeem H Syed[1]. Epigenetic differences in genetically identical *Arabidopsis thaliana* plants modulate alternative splicing (manuscript submitted to The Plant Cell)

References

Abascal, F. *et al.* (2015) 'Alternatively Spliced Homologous Exons Have Ancient Origins and Are Highly Expressed at the Protein Level', *PLoS Computational Biology*, 11(6), p. e1004325. doi: 10.1371/journal.pcbi.1004325.

Al-Hadid, Q. and Yang, Y. (2016) 'R-loop: An emerging regulator of chromatin dynamics', *Acta Biochimica et Biophysica Sinica*, 48(7), pp. 623–631. doi: 10.1093/abbs/gmw052.

Alexander, R. D. *et al.* (2010) 'Splicing-Dependent RNA polymerase pausing in yeast', *Molecular Cell*, 40(4), pp. 582–593. doi: 10.1016/j.molcel.2010.11.005.

Anamika, K. *et al.* (2012) 'RNA Polymerase II Pausing Downstream of Core Histone Genes Is Different from Genes Producing Polyadenylated Transcripts', *PloS one*, 11(6), p. e1004325. doi: 10.1371/journal.pone.0038769.

Louloupi, A., Ntini, E. and Conrad, T. (2018) 'Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency', *Cell Reports*, 23(12), pp. 3419–3698.

Arciga-Reyes, L. *et al.* (2006) 'UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in Arabidopsis', *Plant Journal*, 47(3), pp. 480–489. doi: 10.1111/j.1365-313X.2006.02802.x.

Baena-González, E. *et al.* (2007) 'A central integrator of transcription networks in plant stress and energy signalling', *Nature*, 448(7), pp. 938–42. doi: 10.1038/nature06069.

Bah, A. and Forman-Kay, J. D. (2016) 'Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications.', *The Journal of biological chemistry*, 291(13), pp. 6696–705. doi: 10.1074/jbc.R115.695056.

Barash, Y. *et al.* (2010) 'Deciphering the splicing code', *Nature*, 465(7294), pp. 53–59. doi: 10.1038/nature09000.

Barciszewska-Pacak, M. *et al.* (2015) 'Arabidopsis microRNA expression regulation in a wide range of abiotic stress responses', *Frontiers in Plant Science*, 6, p. 410. doi: 10.3389/fpls.2015.00410.

Boutz, P. L. *et al.* (2007) 'MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development.', *Genes & development*, 21(1), pp. 71–84. doi: 10.1101/gad.1500707.

Brody, Y. *et al.* (2011) 'The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing', *PLoS Biology*, 9(1), p. e1000573. doi: 10.1371/journal.pbio.1000573.

Brogna, S., McLeod, T. and Petric, M. (2016) 'The Meaning of NMD: Translate or Perish', *Trends in Genetics*, 32(7), pp. 395–407. doi: 10.1016/j.tig.2016.04.007.

Brosch, M. *et al.* (2011) 'Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome', *Genome Research*, 21(5), pp. 756–767. doi: 10.1101/gr.114272.110.

Buljan, M. *et al.* (2012) 'Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks', *Molecular Cell*, 46(6), pp. 871–883. doi: 10.1016/j.molcel.2012.05.039.

Calixto, C. P. G. *et al.* (2018) 'Rapid and dynamic alternative splicing impacts the Arabidopsis cold response transcriptome.', *The Plant cell*, 30(7), pp. 1424–1444. doi: 10.1105/tpc.18.00177.

Carrillo Oesterreich, F., Preibisch, S. and Neugebauer, K. M. (2010a) 'Global analysis of nascent rna reveals transcriptional pausing in terminal exons', *Molecular Cell*, 40(4), pp. 571–581. doi: 10.1016/j.molcel.2010.11.004.

Carrillo Oesterreich, F., Preibisch, S. and Neugebauer, K. M. (2010b) 'Global analysis of nascent rna reveals transcriptional pausing in terminal exons', *Molecular Cell*, 40, pp. 571–581. doi: 10.1016/j.molcel.2010.11.004.

Chambon, M. (1978) 'Summary: the molecular biology of the eukaryotic genome is coming of age', *Cold Spring Harbor Symposia on Quantitative Biology, 2(42),* p. 1209-34.

Chao, J. A., Yoon, Y. J. and Singer, R. H. (2012) 'Imaging translation in single cells using fluorescent microscopy', *Cold Spring Harbor Perspectives in Biology*, 4(11), p. a012310. doi: 10.1101/cshperspect.a012310.

Chaudhary, S. *et al.* (2019) 'Perspective on Alternative Splicing and Proteome Complexity in Plants', *Trends in Plant Science*, 24(6), pp. 496–506. doi: 10.1016/j.tplants.2019.02.006.

Chavali, S., Gunnarsson, A. and Babu, M. M. (2017) 'Intrinsically Disordered Proteins Adaptively Reorganize Cellular Matter During Stress', *Trends in Biochemical Sciences*, 42(6), pp. 410–412. doi: 10.1016/j.tibs.2017.04.007.

Chédin, F. (2017) 'Hybrids in the chromatin', *Nature Plants*, 3(9), pp. 692–693. doi: 10.1038/s41477-017-0011-y.

Chen, W., Luo, L. and Zhang, L. (2010) 'The organization of nucleosomes around splice sites', *Nucleic Acids Research*, 38(9), pp. 2788–2798. doi: 10.1093/nar/gkq007.

Cheng, T. L. *et al.* (2017) 'Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain', *Scientific Reports*, 7, p. 42790. doi: 10.1038/srep42790.

Chodavarapu, R. K. *et al.* (2010) 'Relationship between nucleosome positioning and DNA methylation', *Nature*, 466(7304), pp. 388–392. doi: nature09147 [pii]\n10.1038/nature09147.

Churchman, L. S. and Weissman, J. S. (2011) 'Nascent transcript sequencing visualizes transcription at nucleotide resolution', *Nature*, 469(7330), pp. 368–373. doi: 10.1038/nature09652.

Chwialkowska, K. *et al.* (2016) 'Water-deficiency conditions differently modulate the methylome of roots and leaves in barley (Hordeum vulgare L.)', *Journal of Experimental Botany*, 67(4), pp. 1109–1121. doi: 10.1093/jxb/erv552.

Cokus, S. J. *et al.* (2008) 'Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning', *Nature*, 452(7184), pp. 215–219. doi: 10.1038/nature06745.

Conn, V. M. *et al.* (2017) 'A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation', *Nature Plants*, 3, p. 17053. doi: 10.1038/nplants.2017.53.

Core, L. J., Waterfall, J. J. and Lis, J. T. (2008) 'Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters', *Science*, 322(5909), pp. 1845–1848. doi: 10.1126/science.1162228.

Cowling, V. H. and Cole, M. D. (2010) 'Myc regulation of mRNA cap methylation', *Genes*

and Cancer, 1(6), pp. 576–579. doi: 10.1177/1947601910378025.

Cramer, P. *et al.* (1999) 'Coupling of Transcription with Alternative Splicing', *Molecular Cell*, 4(2), pp. 251–258. doi: 10.1016/S1097-2765(00)80372-X.

Cui, J. *et al.* (2013) 'Cytoplasmic polyadenylation is a major mRNA regulator during oogenesis and egg activation in Drosophila', *Developmental Biology*, 383(1), pp. 121–131. doi: 10.1016/j.ydbio.2013.08.013.

Cui, X. *et al.* (2017) '5-Methylcytosine RNA Methylation in Arabidopsis Thaliana', *Molecular Plant*, 10(11), pp. 1387–1399. doi: 10.1016/j.molp.2017.09.013.

Cutter, A. R. and Hayes, J. J. (2015) 'A brief review of nucleosome structure', *FEBS Letters*, 589(20 pT a), pp. 2914–2922. doi: 10.1016/j.febslet.2015.05.016.

Darnell, J. E. (2013) 'Reflections on the history of pre-mRNA processing and highlights of current knowledge: A unified picture', *RNA*, 19(4), pp. 443–60. doi: 10.1261/rna.038596.113.

Day, I. S. *et al.* (2012) 'Interactions of SR45, an SR-like protein, with spliceosomal proteins and an intronic sequence: Insights into regulated splicing', *Plant Journal*, 71(6), pp. 936–947. doi: 10.1111/j.1365-313X.2012.05042.x.

Deng, X. *et al.* (2010) 'Arginine methylation mediated by the Arabidopsis homolog of PRMT5 is essential for proper pre-mRNA splicing', *Proceedings of the National Academy of Sciences*, 107(44), pp. 19114–19119. doi: 10.1073/pnas.1009669107.

Dever, T. E. and Green, R. (2012) 'The elongation, termination, and recycling phases of translation in eukaryotes', *Cold Spring Harbor Perspectives in Biology*, 4(7), p. a013706. doi: 10.1101/cshperspect.a013706.

Ding, F. *et al.* (2014) 'Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis', *BMC Genomics*, 15(1), pp. 1–14. doi: 10.1186/1471-2164-15-431.

Ding, Y. *et al.* (2014) 'In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features', *Nature*, 505(7485), pp. 696–700. doi: 10.1038/nature12756.

Dolata, J. *et al.* (2015) 'NTR1 is required for transcription elongation checkpoints at alternative exons in Arabidopsis', *The EMBO Journal*, 34(4), pp. 544–558. doi: 10.15252/embj.201489478.

Drechsel, G. *et al.* (2013) 'Nonsense-Mediated Decay of Alternative Precursor mRNA Splicing Variants Is a Major Determinant of the Arabidopsis Steady State Transcriptome', *The Plant Cell*, 25(10), pp. 3726–3742. doi: 10.1105/tpc.113.115485.

Dubin, M. J. *et al.* (2015) 'DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation', *eLife*, 4, p. e05255. doi: 10.7554/eLife.05255.

Dunker, A. K. *et al.* (2013) 'What's in a name? Why these proteins are intrinsically disordered', *Intrinsically Disordered Proteins*, 1(1), p. e24157. doi: 10.4161/idp.24157.

Duque, P. (2011) 'A role for SR proteins in plant stress responses', *Plant Signaling and Behavior*, 6(1), pp. 49–54. doi: 10.4161/psb.V.I.14063.

Durand, S. and Lykke-Andersen, J. (2013) 'Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells', *Nature Structural & Molecular Biology*, 20(6), pp. 702–9. doi: 10.1038/nsmb.2575.

Dvinge, H. (2018) 'Regulation of alternative mRNA splicing: old players and new perspectives.', *FEBS letters*, 592(17), pp. 2987–3006. doi: 10.1002/1873-3468.13119.

Ehrlich, M. *et al.* (1982) 'Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells', *Nucleic Acids Research*, 10(8), pp. 2709–2721. doi: 10.1093/nar/10.8.2709.

Feng, J. *et al.* (2015) 'SKIP Confers Osmotic Tolerance during Salt Stress by Controlling Alternative Gene Splicing in Arabidopsis', *Molecular Plant*, 8(7), pp. 1038–1052. doi: 10.1016/j.molp.2015.01.011.

Filichkin, S. *et al.* (2015) 'Alternative splicing in plants: Directing traffic at the crossroads of adaptation and environmental stress', *Current Opinion in Plant Biology*, 24, pp. 125–135. doi: 10.1016/j.pbi.2015.02.008.

Filichkin, S. A. *et al.* (2010) 'Genome-wide mapping of alternative splicing in Arabidopsis thaliana', *Genome Research*, 20(1), pp. 45–58. doi: 10.1101/gr.093302.109.

Filichkin, S. A. *et al.* (2015) 'Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis', *Molecular Plant*, 8(2), pp. 207–227. doi: 10.1016/j.molp.2014.10.011.

Filichkin, S. A. *et al.* (2018) 'Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching', *Frontiers in Plant Science*, 9, pp. 1–17. doi: 10.3389/fpls.2018.00005.

Filichkin, S. A. and Mockler, T. C. (2012) 'Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes', *Biology Direct*, 7(1), p. 20. doi: 10.1186/1745-6150-7-20.

Floor, S. N. and Doudna, J. A. (2016) 'Tunable protein synthesis by transcript isoforms in human cells', *eLife*, 5, p. e10921. doi: 10.7554/eLife.10921.001.

Fong, N. *et al.* (2014) 'Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate', *Genes and Development*, 28(23), pp. 2663–2676. doi: 10.1101/gad.252106.114.

Fusby, B. *et al.* (2015) 'Coordination of RNA Polymerase II Pausing and 3' end processing factor recruitment with alternative polyadenylation', *Molecular and Cellular Biology*, 36(2), pp. 295–303. doi: 10.1128/MCB.00898-15.

García, A., González, S. and Antequera, F. (2017) 'Nucleosomal organization and DNA base composition patterns', *Nucleus*, 8(5), pp. 469–474. doi: 10.1080/19491034.2017.1337611.

Gasch, A. *et al.* (2006) 'The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains', *Journal of Biological Chemistry*, 281(1), pp. 356–364. doi: 10.1074/jbc.M508047200.

Gebauer, F., Preiss, T. and Hentze, M. W. (2012) 'From cis-regulatory elements to complex RNPs and back', *Cold Spring Harbor Perspectives in Biology*, 4(7), p. a012245. doi: 10.1101/cshperspect.a012245.

Gelfman, S. *et al.* (2013) 'DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure', *Genome Research*, 23(5), pp. 789–799. doi: 10.1101/gr.143503.112.

Gibon, Y. *et al.* (2009) 'Adjustment of growth, starch turnover, protein content and central metabolism to a decrease of the carbon supply when Arabidopsis is grown in very short photoperiods', *Plant, Cell and Environment*, 32(7), pp. 859–874. doi: 10.1111/j.1365-3040.2009.01965.x.

Gibcus J. H. and Dekker J (2012) 'The context of gene expression regulation', F1000 Biology Reports, 4, p.8.

Gilbert, W. V., Bell, T. A. and Schaening, C. (2016) 'Messenger RNA modifications: Form, distribution, and function', *Science*, 352(6292), pp. 1408–1412. doi: 10.1126/science.aad8711.

Godoy Herz, M. A. *et al.* (2014) 'Shedding light on the chloroplast as a remote control of nuclear gene expression', *Plant Signaling and Behavior*, 9(11), p. e976150. doi: 10.4161/15592324.2014.976150.

Godoy Herz, M. A. *et al.* (2019) 'Light Regulates Plant Alternative Splicing through the Control of Transcriptional Elongation', *Molecular Cell*, 73(5), p. 1066–1074.e3. doi: 10.1016/j.molcel.2018.12.005.

Gohring, J., Jacak, J. and Barta, A. (2014) 'Imaging of Endogenous Messenger RNA Splice Variants in Living Cells Reveals Nuclear Retention of Transcripts Inaccessible to Nonsense-Mediated Decay in Arabidopsis', *The Plant Cell*, 26(2), pp. 754–764. doi: 10.1105/tpc.113.118075.

Gonzalez, I. *et al.* (2015) 'A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature', *Nature Structural and Molecular Biology*, 22(5), pp. 370–376. doi: 10.1038/nsmb.3005.

Graf, A. *et al.* (2010) 'Circadian control of carbohydrate availability for growth in Arabidopsis plants at night.', *Proceedings of the National Academy of Sciences*, 107(20), pp. 9458–9463. doi: 10.1073/pnas.0914299107.

Gib

Hahn, S. (2004) 'Structure and mechanism of the RNA polymerase II transcription machinery', *Nature Structural and Molecular Biology*, 11(5), pp. 394–403. doi: 10.1038/nsmb763.

Hajheidari, M., Koncz, C. and Eick, D. (2013) 'Emerging roles for RNA polymerase II CTD in Arabidopsis', *Trends in Plant Science*, 18(11), pp. 633–643. doi: 10.1016/j.tplants.2013.07.001.

Han, S. P., Tang, Y. H. and Smith, R. (2010) 'Functional diversity of the hnRNPs: past, present and perspectives', *Biochemical Journal*, 430(3), pp. 379–392. doi: 10.1042/BJ20100396.

Hartmann, L. *et al.* (2016) 'Alternative Splicing Substantially Diversifies the Transcriptome during Early Photomorphogenesis and Correlates with the Energy Availability in Arabidopsis', *The Plant Cell*, 28(11), pp. 2715–2734. doi: 10.1105/tpc.16.00508.

Hartmann, L., Wießner, T. and Wachter, A. (2018) 'Subcellular Compartmentation of Alternatively-Spliced Transcripts Defines SERINE/ARGININE-RICH PROTEIN 30 Expression', *Plant Physiology*, 176(4), pp. 2886–2903. doi: 10.1104/pp.17.01260.

Hattori, D. *et al.* (2007) 'Dscam diversity is essential for neuronal wiring and self-

recognition', *Nature*, 449(7159), pp. 223–7. doi: 10.1038/nature06099.

Heintzen, C. *et al.* (1997) 'AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in Arabidopsis thaliana', *Proceedings of the National Academy of Sciences*, 94(16), pp. 8515–8520. doi: 10.1073/pnas.94.16.8515.

Hetzel, J. *et al.* (2016) 'Nascent RNA sequencing reveals distinct features in plant transcription', *Proceedings of the National Academy of Sciences*, 113(43), pp. 12316–12321. doi: 10.1073/pnas.1603217113.

Hinnebusch, A. G. and Lorsch, J. R. (2012) 'The mechanism of eukaryotic translation initiation: New insights and challenges', *Cold Spring Harbor Perspectives in Biology*, 4(10), p. a011544. doi: 10.1101/cshperspect.a011544.

Hirose, Y. and Manley, J. L. (2000) 'RNA polymerase II and the integration of nuclear events', *Genes and Development*, 14(12), pp. 1415–1429.

Hori, K. and Watanabe, Y. (2005) 'UPF3 suppresses aberrant spliced mRNA in Arabidopsis', *Plant Journal*, 43(4), pp. 530–540. doi: 10.1111/j.1365-313X.2005.02473.x.

Hori, K. and Watanabe, Y. (2007) 'Context analysis of termination codons in mRNA that are recognized by plant NMD', *Plant and Cell Physiology*, 48(7), pp. 1072–1078. doi: 10.1093/pcp/pcm075.

Hossain, M. S. *et al.* (2017) 'Divergent cytosine DNA methylation patterns in single-cell, soybean root hairs', *New Phytologist*, 214(2), pp. 808–819. doi: 10.1111/nph.14421.

Huang, J. *et al.* (2017) 'An oomycete plant pathogen reprograms host pre-mRNA splicing to subvert immunity', *Nature Communications*, 8(1), p. 2051. doi: 10.1038/s41467-017-02233-5.

Inada, T. (2017) 'The Ribosome as a Platform for mRNA and Nascent Polypeptide Quality Control', *Trends in Biochemical Sciences*, 42(1), pp. 5–15. doi: 10.1016/j.tibs.2016.09.005.

Ingolia, N. T. (2016) 'Ribosome Footprint Profiling of Translation throughout the Genome', *Cell*, 165(1), pp. 22–33. doi: 10.1016/j.cell.2016.02.066.

Irimia, M. *et al.* (2014) 'A highly conserved program of neuronal microexons is misregulated in autistic brains', *Cell*, 159(7), pp. 1511–1523. doi: 10.1016/j.cell.2014.11.035.

Ishihara, H. *et al.* (2017) 'Growth rate correlates negatively with protein turnover in Arabidopsis accessions', *Plant Journal*, 91(3), pp. 416–429. doi: 10.1111/tpj.13576.

Jabre, I. *et al.* (2019) 'Does co-transcriptional regulation of alternative splicing mediate plant stress responses?', *Nucleic Acids Research*. Oxford University Press, 47(6), pp. 2716–2726. doi: 10.1093/nar/gkz121.

Jackson, R. J., Hellen, C. U. T. and Pestova, T. V. (2010) 'The mechanism of eukaryotic translation initiation and principles of its regulation', *Nature Reviews Molecular Cell Biology*, 11(2), pp. 113–27. doi: 10.1038/nrm2838.

James, A. B., Syed, N. H., Bordage, S., *et al.* (2012) 'Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes.', *The Plant cell*, 24(3), pp. 961–81. doi: 10.1105/tpc.111.093948.

James, A. B., Syed, N. H., Brown, J. W. S., *et al.* (2012) 'Thermoplasticity in the plant circadian clock: how plants tell the time-perature.', *Plant signaling & behavior*, 7(10), pp.

1219–23. doi: 10.4161/psb.21491.

James AB, Sullivan S, N. H. (2018) 'Global spatial analysis of Arabidopsis natural variants implicates 5'UTR splicing of LATE ELONGATED HYPOCOTYL in responses to temperature.', *Plant Cell Environ*, 41(7), pp. 1524–1538.

James AB et al. (2018) 'Isoform switching of splicing factors regulates splicing of LATE ELONGATED HYPOCOTYL (LHY).', *Plant Cell Environ*, 41(7), pp. 1539–1550.

Jiang, J. *et al.* (2017) 'Integrating Omics and Alternative Splicing Reveals Insights into Grape Response to High Temperature', *Plant Physiology*, 173(2), pp. 1502–1518. doi: 10.1104/pp.16.01305.

Jimeno-González, S. *et al.* (2015) 'Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing', *Proceedings of the National Academy of Sciences*, 112(48), pp. 14840–14845. doi: 10.1073/pnas.1506760112.

Johnson, D. S. *et al.* (2007) 'Genome-wide mapping of in vivo protein-DNA interactions', *Science*, 316(5830), pp. 1497–1502. doi: 10.1126/science.1141319.

Juntawong, P. *et al.* (2014) 'Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), pp. E203-12. doi: 10.1073/pnas.1317811111.

Kahles, A. *et al.* (2018) 'Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients', *Cancer Cell*, 34(2), p. 211–224.e6. doi: 10.1016/j.ccell.2018.07.001.

Kalyna, M. *et al.* (2012) 'Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis', *Nucleic Acids Research*, 40(6), pp. 2454–69. doi: 10.1093/nar/gkr932.

Kalyna, M. and Barta, A. (2004) 'A plethora of plant serine/arginine-rich proteins: Redundancy or evolution of novel gene functions?', *Biochemical Society Transactions*, 32(4), pp. 561–564. doi: 10.1042/BST0320561.

Lelli, K *et al.* (2012) 'Disentangling the Many Layers of Eukaryotic Transcriptional Regulation', *Annual Review of Genetics,* 46, p.43-68.

Kawakatsu, T. *et al.* (2017) 'Dynamic DNA methylation reconfiguration during seed development and germination', *Genome Biology*, 18(1), p. 171. doi: 10.1186/s13059-017-1251-x.

Kelly, S. *et al.* (2015) 'Exon Skipping Is Correlated with Exon Circularization', *Journal of Molecular Biology*, 427(15), pp. 2414–2417. doi: 10.1016/j.jmb.2015.02.018.

Kerényi, Z. *et al.* (2008) 'Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay', *The EMBO Journal*, 27(11), pp. 1585–1595. doi: 10.1038/emboj.2008.88.

Khodor, Y. L. *et al.* (2011) 'Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila', *Genes and Development*, 25(23), pp. 2502–2512. doi: 10.1101/gad.178962.111.

Kim, E., Magen, A. and Ast, G. (2007) 'Different levels of alternative splicing among eukaryotes', *Nucleic Acids Research*, 35(1), pp. 125–131. doi: 10.1093/nar/gkl924.

Kovacs, E. *et al.* (2010) 'Dual coding in alternative reading frames correlates with intrinsic protein disorder', *Proceedings of the National Academy of Sciences*, 107(12), pp. 5429–5434.

doi: 10.1073/pnas.0907841107.

Kozak, M. (1989) 'The scanning model for translation: An update', *Journal of Cell Biology*, 108(2), pp. 229–41. doi: 10.1083/jcb.108.2.229.

Kozak, M. (1992) 'Regulation of Translation in Eukaryotic Systems', *Annual Review of Cell and Developmental Biology*, 8, pp. 197–225. doi: 10.1146/annurev.cellbio.8.1.197.

Kwon, Y.-J. *et al.* (2014) 'Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis', *BMC Plant Biology*, 14(1), p. 136. doi: 10.1186/1471-2229-14-136.

Kwon, Y. J. *et al.* (2014) 'Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis', *BMC Plant Biol*, 14, p. 136. doi: 10.1186/1471-2229-14-136.

Lämke, J. and Bäurle, I. (2017) 'Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants', *Genome Biology*, 18(1), p. 124. doi: 10.1186/s13059-017-1263-6.

Lang-Mladek, C. *et al.* (2010) 'Transgenerational inheritance and resetting of stress-induced loss of epigenetic gene silencing in arabidopsis', *Molecular Plant*, 3(3), pp. 594–602. doi: 10.1093/mp/ssq014.

Lasko, P. (2012) 'mRNA localization and translational control in Drosophila oogenesis', *Cold Spring Harbor Perspectives in Biology*, 4(10), p. a012294. doi: 10.1101/cshperspect.a012294.

Lee, A. *et al.* (2016) 'The OsCYP19-4 gene is expressed as multiple alternatively spliced transcripts encoding Isoforms with distinct cellular localizations and PPIase activities under cold stress', *International Journal of Molecular Sciences*, 17(7), p. 1154. doi: 10.3390/ijms17071154.

Lee, Y. and Rio, D. C. (2015) 'Mechanisms and Regulation of Alternative Pre-mRNA Splicing', *Annual Review of Biochemistry*, 84(1), pp. 291–323. doi: 10.1146/annurev-biochem-060614-034316.

Lejeune, F., Ranganathan, A. C. and Maquat, L. E. (2004) 'eIF4G is required for the pioneer round of translation in mammalian cells', *Nature Structural and Molecular Biology*, 11(10), pp. 992–1000. doi: 10.1038/nsmb824.

Lemay, J. F. and Bachand, F. (2015) 'Fail-safe transcription termination: Because one is never enough', *RNA Biology*, 12(9), pp. 927–32. doi: 10.1080/15476286.2015.1073433.

Lenasi, T. and Barboric, M. (2010) 'P-TEFb stimulates transcription elongation and pre-mRNA splicing through multilateral mechanisms', *RNA Biology*, 7(2), pp. 145–150. doi: 10.4161/rna.7.2.11057.

Li-Byarlay, H., Li, Y. and Stroud, H. (2013) 'RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee', *Proceedings of the National Academy of Sciences of the United States of America*, 110(31), pp. 12750–12755. doi: 10.1073/pnas.1310735110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1310735110.

Li, B., Carey, M. and Workman, J. L. (2007) 'The Role of Chromatin during Transcription', *Cell*, 128(4), pp. 707–19. doi: 10.1016/j.cell.2007.01.015.

Lim, J. *et al.* (2016) 'MTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development', *Genes and Development*, 30, pp. 1671–1682. doi: 10.1101/gad.284802.116.

Ling, Y. *et al.* (2018) 'Thermopriming triggers splicing memory in Arabidopsis', *Journal of Experimental Botany*, 69(10), pp. 2659–2675. doi: 10.1093/jxb/ery062.

Lister, R. *et al.* (2008) 'Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis', *Cell*, 133(3), pp. 523–536. doi: 10.1016/j.cell.2008.03.029.

Listerman, I., Sapra, A. K. and Neugebauer, K. M. (2006) 'Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells', *Nature Structural and Molecular Biology*, 13(9), pp. 815–822. doi: 10.1038/nsmb1135.

Liu, H. *et al.* (2018) 'Distinct heat shock factors and chromatin modifications mediate the organ-autonomous transcriptional memory of heat stress', *The Plant Journal*, 95(3), pp. 401–413. doi: 10.1111/tpj.13958.

Liu, J. *et al.* (2013) 'An Autoregulatory Loop Controlling Arabidopsis HsfA2 Expression: Role of Heat Shock-Induced Alternative Splicing', *PLANT PHYSIOLOGY*, 162(1), pp. 512–521. doi: 10.1104/pp.112.205864.

Liu, J. *et al.* (2016) 'Alternative splicing of rice WRKY62 and WRKY76 transcription factor genes in pathogen defense', *Plant Physiology*, 171(2), pp. 1427–42. doi: 10.1104/pp.15.01921.

Liu, M.-J. *et al.* (2015) 'Determinants of nucleosome positioning and their influence on plant gene expression', *Genome Research*, 25(8), pp. 1182–1195. doi: 10.1101/gr.188680.114.

Liu, Y. *et al.* (2017) 'Impact of Alternative Splicing on the Human Proteome', *Cell Reports*, 20(5), pp. 1229–1241. doi: 10.1016/j.celrep.2017.07.025.

Liu, Z. *et al.* (2017) 'Identified of a novel cis-element regulating the alternative splicing of LcDREB2', *Scientific Reports*, 7, p. 46106. doi: 10.1038/srep46106.

Liu, Z. *et al.* (2018) 'Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (Triticum aestivum L.)', *Plant Biotechnology Journal*, 16(3), pp. 714–726. doi: 10.1111/pbi.12822.

Lorković, Z. J. *et al.* (2000) 'Pre-mRNA splicing in higher plants', *Trends in Plant Science*, 5(4), pp. 160–167. doi: doi:10.1016/S1360-1385(00)01595-8.

Lothrop, A. P., Torres, M. P. and Fuchs, S. M. (2013) 'Deciphering post-translational modification codes', *FEBS Letters*, 587(8), pp. 1247–57. doi: 10.1016/j.febslet.2013.01.047.

Lu, X. *et al.* (2017) 'Single-base resolution methylomes of upland cotton (Gossypium hirsutum L.) reveal epigenome modifications in response to drought stress', *BMC Genomics*, 18(1), p. 297. doi: 10.1186/s12864-017-3681-y.

Luco, R. F. *et al.* (2010) 'Regulation of alternative splicing by histone modifications.', *Science (New York, N.Y.)*, 327(5968), pp. 996–1000. doi: 10.1126/science.1184208.

Luco, R. F. *et al.* (2011) 'Epigenetics in alternative pre-mRNA splicing', *Cell*, 144(1), pp. 16–26. doi: 10.1016/j.cell.2010.11.056.

Luo, G.-Z. *et al.* (2014) 'Unique features of the m6A methylome in Arabidopsis thaliana.', *Nature communications*, 5, p. 5630. doi: 10.1038/ncomms6630.

Lyko, F. *et al.* (2010) 'The honey bee epigenomes: Differential methylation of brain DNA in queens and workers', *PLoS Biology*, 8(11), p. e1000506. doi: 10.1371/journal.pbio.1000506.

Mahrez, W. *et al.* (2016) 'BRR2a Affects Flowering Time via FLC Splicing', *PLoS Genetics*, 12(4), p. e1005924. doi: 10.1371/journal.pgen.1005924.

Malapeira, J., Khaitova, L. C. and Mas, P. (2012) 'Ordered changes in histone modifications at the core of the Arabidopsis circadian clock.', *Proceedings of the National Academy of Sciences of the United States of America*, 109(52), pp. 21540–5. doi: 10.1073/pnas.1217022110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1217022110.

Malapeira, J. and Mas, P. (2013) 'A chromatin-dependent mechanism regulates gene expression at the core of the Arabidopsis circadian clock', *Plant Signal Behav*, 8(5), p. e24079. doi: 10.4161/psb.24079.

Maquat, L. E., Tarn, W. Y. and Isken, O. (2010) 'The pioneer round of translation: Features and functions', *Cell*, 142(3), pp. 368–374. doi: 10.1016/j.cell.2010.07.022.

Marquez, Y. *et al.* (2012) 'Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis', *Genome Research*, 22(6), pp. 1184–1195. doi: 10.1101/gr.134106.111.

Marquez, Y. *et al.* (2015) 'Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity', *Genome Research*, 25(7), pp. 995–1007. doi: 10.1101/gr.186585.114.

Mastrangelo, A. M. *et al.* (2012) 'Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity', *Plant Science*, 185–186, pp. 40–49. doi: 10.1016/j.plantsci.2011.09.006.

Matlin, A. J., Clark, F. and Smith, C. W. J. (2005) 'Understanding alternative splicing: towards a cellular code.', *Nature reviews. Molecular cell biology*, 6(5), pp. 386–98. doi: 10.1038/nrm1645.

Mavrich, T. N. *et al.* (2008) 'Nucleosome organization in the Drosophila genome', *Nature*, 453(7193), pp. 358–362. doi: 10.1038/nature06929.

Memczak, S. *et al.* (2013) 'Circular RNAs are a large class of animal RNAs with regulatory potency', *Nature*, 495(7441), pp. 333–338. doi: 10.1038/nature11928.

Meng, Y. *et al.* (2013) 'Introns targeted by plant microRNAs: A possible novel mechanism of gene regulation', *Rice*, 6(1), pp. 1–10. doi: 10.1186/1939-8433-6-1.

Meyer, K. D. and Jaffrey, S. R. (2014) 'The dynamic epitranscriptome: N6-methyladenosine and gene expression control.', *Nature reviews. Molecular cell biology*, 15(5), pp. 313–26. doi: 10.1038/nrm3785.

Millar, A. J. (2016) 'The Intracellular Dynamics of Circadian Clocks Reach for the Light of Ecology and Evolution', *Annual Review of Plant Biology*, 67(1), pp. 595–618. doi: 10.1146/annurev-arplant-043014-115619.

Minezaki, Y. *et al.* (2006) 'Human Transcription Factors Contain a High Fraction of Intrinsically Disordered Regions Essential for Transcriptional Regulation', *Journal of Molecular Biology*, 359(4), pp. 1137–1149. doi: 10.1016/j.jmb.2006.04.016.

Mischo, H. E. and Proudfoot, N. J. (2013) 'Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(1), pp. 174–85. doi: 10.1016/j.bbagrm.2012.10.003.

Morgan, J. T., Fink, G. R. and Bartel, D. P. (2018) 'Excised linear introns regulate growth in yeast', *Nature*. Springer US, 565(7741), pp. 606–611. doi: 10.1038/s41586-018-0828-1.

Mustroph, A. *et al.* (2009) 'Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis', *Proceedings of the National Academy of Sciences*, 106(44), pp. 18843–8. doi: 10.1073/pnas.0906131106.

Nahkuri, S., Taft, R. J. and Mattick, J. S. (2009) 'Nucleosomes are preferentially positioned at exons in somatic and sperm cells', *Cell Cycle*, 8(20), pp. 3420–3424. doi: 10.4161/cc.8.20.9916.

Nechaev, S. *et al.* (2010) 'Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila', *Science*, 327(5963), pp. 335–338. doi: 10.1126/science.1181421.

Niklas, K. J. *et al.* (2015) 'Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications', *Frontiers in Cell and Developmental Biology*, 3, pp. 1–13. doi: 10.3389/fcell.2015.00008.

Niklas, K. J., Dunker, A. K. and Yruela, I. (2018) 'The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins', *Journal of Experimental Botany*, 69(7), pp. 1437–1446. doi: 10.1093/jxb/erx493.

Ning, K. and Nesvizhskii, A. I. (2010) 'The utility of mass spectrometry-based proteomic data for validation of novel alternative splice forms reconstructed from RNA-Seq data: A preliminary assessment', *BMC Bioinformatics*, 11(11), p. S14. doi: 10.1186/1471-2105-11-S11-S14.

Nojima, T. *et al.* (2015) 'Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing', *Cell*, 161(3), pp. 526–540. doi: 10.1016/j.cell.2015.03.027.

Nyikó, T. *et al.* (2009) 'Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner', *Plant Molecular Biology*, 71(4–5), pp. 367–378. doi: 10.1007/s11103-009-9528-4.

Oldfield, C. J. and Dunker, A. K. (2014) 'Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions', *Annual Review of Biochemistry*, 83(1), pp. 553–584. doi: 10.1146/annurev-biochem-072711-164947.

Olsen, J. V., Ong, S.-E. and Mann, M. (2004) 'Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues', *Molecular & Cellular Proteomics*, 3(6), pp. 608–614. doi: 10.1074/mcp.T400003-MCP200.

Pajoro, A. *et al.* (2017) 'Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants', *Genome Biology*, 18(1), p. 102. doi: 10.1186/s13059-017-1235-x.

Palusa, S. G., Ali, G. S. and Reddy, A. S. N. (2007) 'Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: Regulation by hormones and stresses', *Plant Journal*, 49(6), pp. 1091–1107. doi: 10.1111/j.1365-313X.2006.03020.x.

Palusa, S. G. and Reddy, A. S. N. (2010) 'Extensive coupling of alternative splicing of pre-

mRNAs of serine/arginine (SR) genes with nonsense-mediated decay', *New Phytologist*, 185(1), pp. 83–89. doi: 10.1111/j.1469-8137.2009.03065.x.

Palusa, S. G. and Reddy, A. S. N. (2015) 'Differential recruitment of splice variants from SR Pre-mRNAs to polysomes during development and in response to stresses', *Plant and Cell Physiology*, 56(3), pp. 421–427. doi: 10.1093/pcp/pcv010.

Pan, T. *et al.* (2018) 'Heat stress alters genome-wide profiles of circular RNAs in Arabidopsis', *Plant Molecular Biology*, 96(3), pp. 217–229. doi: 10.1007/s11103-017-0684-7.

Parenteau, J. *et al.* (2019) 'Introns are mediators of cell response to starvation', *Nature*, 565(7741), pp. 612–617. doi: 10.1038/s41586-018-0859-7.

Pavitt, G. D. and Ron, D. (2012) 'New insights into translational regulation in the endoplasmic reticulum unfolded protein response', *Cold Spring Harbor Perspectives in Biology*, 4(6), p. a012278. doi: 10.1101/cshperspect.a012278.

Penfield, S., Josse, E. M. and Halliday, K. J. (2010) 'A role for an alternative splice variant of PIF6 in the control of Arabidopsis primary seed dormancy', *Plant Molecular Biology*, 73(1–2), pp. 89–95. doi: 10.1007/s11103-009-9571-1.

Petrillo, E. *et al.* (2011) 'Alternative splicing adds a new loop to the circadian clock', *Communicative & Integrative Biology*, 4(3), pp. 284–286. doi: 10.4161/cib.4.3.14777.

Petrillo, E., Godoy Herz, M. A., Fuchs, A., *et al.* (2014) 'A chloroplast retrograde signal regulates nuclear alternative splicing', *Science*, 344(6182), pp. 427–430. doi: 10.1126/science.1250322.

Petrillo, E., Godoy Herz, M. A., Barta, A., *et al.* (2014) 'Let there be light: regulation of gene expression in plants', *RNA biology*, 11(10), pp. 1215–1220. doi: 10.4161/15476286.2014.972852.

Pietrosemoli, N. *et al.* (2013) 'Genome-Wide Analysis of Protein Disorder in Arabidopsis thaliana: Implications for Plant Environmental Adaptation', *PLoS ONE*, 8(2), p. e55524. doi: 10.1371/journal.pone.0055524.

Piques, M. *et al.* (2009) 'Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis', *Molecular Systems Biology*, 5, p. 314. doi: 10.1038/msb.2009.68.

Proudfoot, N. J. (2011) 'Ending the message: Poly(A) signals then and now', *Genes and Development*, 25(17), pp. 1770–82. doi: 10.1101/gad.17268411.

Rai, A. K. *et al.* (2018) 'Kinase-controlled phase transition of membraneless organelles in mitosis', *Nature*, 559(7713), pp. 211–216. doi: 10.1038/s41586-018-0279-8.

Reddy, A. S. N. *et al.* (2012) 'Deciphering the Plant Splicing Code: Experimental and Computational Approaches for Predicting Alternative Splicing and Splicing Regulatory Elements', *Frontiers in Plant Science*, 3, p. 18. doi: 10.3389/fpls.2012.00018.

Reddy, A. S. N. *et al.* (2013) 'Complexity of the alternative splicing landscape in plants.', *The Plant cell*, 25(10), pp. 3657–83. doi: 10.1105/tpc.113.117523.

Remy, E. *et al.* (2013) 'A Major Facilitator Superfamily Transporter Plays a Dual Role in Polar Auxin Transport and Drought Stress Tolerance in Arabidopsis', *The Plant Cell*, 25(3),

pp. 901–926. doi: 10.1105/tpc.113.110353.

Riback, J. A. *et al.* (2017) 'Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response', *Cell*, 168(6), pp. 1028–1040. doi: 10.1016/j.cell.2017.02.027.

Richardson, D. N. *et al.* (2011) 'Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: Insights into sub-family classification and extent of alternative splicing', *PLoS ONE*, 6(9), p. e24542. doi: 10.1371/journal.pone.0024542.

Roberts, G. C. *et al.* (1998) 'Co-transcriptional commitment to alternative splice site selection.', *Nucleic acids research*, 26(24), pp. 5568–72. doi: 10.1093/nar/26.24.5568.

Roundtree, I. A. and He, C. (2016) 'Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing', *Trends in Genetics*, 32(6), pp. 320–321. doi: 10.1016/j.tig.2016.03.006.

Roux, P. P. and Topisirovic, I. (2012) 'Regulation of mRNA translation by signaling pathways', *Cold Spring Harbor Perspectives in Biology*, 4(11), p. a012252. doi: 10.1101/cshperspect.a012252.

Rufener, S. C. and Mühlemann, O. (2013) 'EIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells', *Nature Structural and Molecular Biology*, 20(6), pp. 710–7. doi: 10.1038/nsmb.2576.

Rühl, C. *et al.* (2012) 'Polypyrimidine tract binding protein homologs from Arabidopsis are key regulators of alternative splicing with implications in fundamental developmental processes.', *The Plant cell*, 24(11), pp. 4360–75. doi: 10.1105/tpc.112.103622.

Sanchez, S. E. *et al.* (2010) 'A methyl transferase links the circadian clock to the regulation of alternative splicing', *Nature*, 468(7320), pp. 112–116. doi: 10.1038/nature09470.

Sato, H., Hosoda, N. and Maquat, L. E. (2008) 'Efficiency of the Pioneer Round of Translation Affects the Cellular Site of Nonsense-Mediated mRNA Decay', *Molecular Cell*, 29(2), pp. 255–262. doi: 10.1016/j.molcel.2007.12.009.

Schmal, C., Reimann, P. and Staiger, D. (2013) 'A Circadian Clock-Regulated Toggle Switch Explains AtGRP7 and AtGRP8 Oscillations in Arabidopsis thaliana', *PLoS Computational Biology*, 9(3), p. e1002986. doi: 10.1371/journal.pcbi.1002986.

Schwartz, A. M. *et al.* (2006) 'Stability of plant mRNAs depends on the length of the 3′-untranslated region', *Biochemistry (Moscow)*, 71(12), pp. 1377–1384. doi: 10.1134/S0006297906120145.

Schwartz, S., Meshorer, E. and Ast, G. (2009) 'Chromatin organization marks exon-intron structure.', *Nature Structural & Molecular Biology*, 16(9), pp. 990–995. doi: 10.1038/nsmb.1659.

Seaton, D. D. *et al.* (2018) 'Photoperiodic control of the *Arabidopsis* proteome reveals a translational coincidence mechanism', *Molecular Systems Biology*, 14(3), p. e7962. doi: 10.15252/msb.20177962.

Seo, P. J. and Mas, P. (2015) 'STRESSing the role of the plant circadian clock', *Trends in Plant Science*, 20(4), pp. 230–237. doi: 10.1016/j.tplants.2015.01.001.

Shaul, O. (2015) 'Unique Aspects of Plant Nonsense-Mediated mRNA Decay', *Trends in Plant Science*, 20(11), pp. 767–779. doi: 10.1016/j.tplants.2015.08.011.

Shayevitch, R. *et al.* (2018) 'The Importance of DNA Methylation of Exons on Alternative

Splicing', *Rna*, 24(10), pp. 1351–1362. doi: 10.1261/rna.064865.117.

Shen, H., Kan, J. L. C. and Green, M. R. (2004) 'Arginine-Serine-Rich Domains Bound at Splicing Enhancers Contact the Branchpoint to Promote Prespliceosome Assembly', *Molecular Cell*, 13(3), pp. 367–376. doi: 10.1016/S1097-2765(04)00025-5.

Sheynkman, G. M. *et al.* (2013) 'Discovery and mass spectrometric analysis of novel splice-junction peptides using RNA-Seq.', *Molecular & cellular proteomics : MCP*, 12, pp. 2341–53. doi: 10.1074/mcp.O113.028142.

Shukla, S. *et al.* (2011) 'CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.', *Nature*, 479(7371), pp. 74–9. doi: 10.1038/nature10442.

Shukla, S. and Oberdoerffer, S. (2012) 'Co-transcriptional regulation of alternative pre-mRNA splicing', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(7), pp. 673–683. doi: 10.1016/j.bbagrm.2012.01.014.

Skourti-Stathaki, K., Kamieniarz-Gdula, K. and Proudfoot, N. J. (2014) 'R-loops induce repressive chromatin marks over mammalian gene terminators', *Nature*, 516(7531), pp. 436–439. doi: 10.1038/nature13787.

Slobodin, B. *et al.* (2017) 'Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation', *Cell*, 169(2), p. 326–337.e12. doi: 10.1016/j.cell.2017.03.031.

Smithers, B., Oates, M. E. and Gough, J. (2015) 'Splice junctions are constrained by protein disorder', *Nucleic Acids Research*, 43(10), pp. 4814–4822. doi: 10.1093/nar/gkv407.

Sonenberg, N. and Hinnebusch, A. G. (2009) 'Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets', *Cell*, 136(4), pp. 731–745. doi: 10.1016/j.cell.2009.01.042.

Staiger, D. *et al.* (2003) 'The circadian clock regulated RNA- binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre- mRNA', *The Plant*, 33(2), pp. 361–71.

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

Sterne-Weiler, T. *et al.* (2013) 'Frac-seq reveals isoform-specific recruitment to polyribosomes', *Genome Research*, 23(10), pp. 1615–1623. doi: 10.1101/gr.148585.112.

Strom, A. R. *et al.* (2017) 'Phase separation drives heterochromatin domain formation', *Nature*, 547(7662), pp. 241–245. doi: 10.1038/nature22989.

Strumillo, M. and Beltrao, P. (2015) 'Towards the computational design of protein post-translational regulation', *Bioorganic and Medicinal Chemistry*, 23(12), pp. 2877–82. doi: 10.1016/j.bmc.2015.04.056.

Sun, S. *et al.* (2010) 'SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control', *Nature Structural and Molecular Biology*, 17(3), pp. 306–312. doi: 10.1038/nsmb.1750.

Sunkar, R. (2004) 'Novel and Stress-Regulated MicroRNAs and Other Small RNAs from Arabidopsis', *Plant Cell.*, 16(8), pp. 2001–2019. doi: 10.1105/tpc.104.022830.

Svejstrup, J. Q. (2004) 'The RNA polymerase II transcription cycle: Cycling through

chromatin', *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1677(1–3), pp. 64–73. doi: 10.1016/j.bbaexp.2003.10.012.

Swinburne, I. A. *et al.* (2006) 'Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription', *Genome Research*, 16(7), pp. 912–921. doi: 10.1101/gr.5211806.

Syed, N. H. *et al.* (2012) 'Alternative splicing in plants - coming of age', *Trends in Plant Science*, 17(10), pp. 616–623. doi: 10.1016/j.tplants.2012.06.001.

Tanabe, N. *et al.* (2007) 'Differential expression of alternatively spliced mRNAs of Arabidopsis SR protein homologs, atSR30 and atSR45a, in response to environmental stress', *Plant and Cell Physiology*, 48(7), pp. 1036–1049. doi: 10.1093/pcp/pcm069.

Tilgner, H. *et al.* (2009) 'Nucleosome positioning as a determinant of exon recognition', *Nature Structural and Molecular Biology*, 16(9), pp. 996–1001. doi: 10.1038/nsmb.1658.

Trcek, T. *et al.* (2013) 'Temporal and spatial characterization of nonsense-mediated mRNA decay.', *Genes & development*, 27(5), pp. 541–51. doi: 10.1101/gad.209635.112.

Tress, M. L. *et al.* (2007) 'The implications of alternative splicing in the ENCODE protein complement', *Proceedings of the National Academy of Sciences*, 104(13), pp. 5495–5500. doi: 10.1073/pnas.0700800104.

Tress, M. L. *et al.* (2008) 'Proteomics studies confirm the presence of alternative protein isoforms on a large scale', *Genome Biology*, 9(11), p. R162. doi: 10.1186/gb-2008-9-11-r162.

Tress, M. L., Abascal, F. and Valencia, A. (2017) 'Alternative Splicing May Not Be the Key to Proteome Complexity', *Trends in Biochemical Sciences*, 42(2), pp. 98–110. doi: 10.1016/j.tibs.2016.08.008.

Tsuji, H. *et al.* (2006) 'Dynamic and reversible changes in histone H3-Lys4 methylation and H3 acetylation occurring at submergence-inducible genes in rice', *Plant and Cell Physiology*, 47(7), pp. 995–1003. doi: 10.1093/pcp/pcj072.

Ullah, F. *et al.* (2018) 'Exploring the relationship between intron retention and chromatin accessibility in plants', *BMC Genomics*, 19(1), p. 21. doi: 10.1186/s12864-017-4393-z.

Vandivier, L. E. *et al.* (2015) 'Chemical Modifications Mark Alternatively Spliced and Uncapped Messenger RNAs in Arabidopsis', *The Plant Cell*, 27(11), pp. 3024–3037. doi: 10.1105/tpc.15.00591.

Vandivier, L. E. and Gregory, B. D. (2018) 'New insights into the plant epitranscriptome', *Journal of Experimental Botany*, 69(20), pp. 4659–4665. doi: 10.1093/jxb/ery262.

Wachter, A., Rühl, C. and Stauffer, E. (2012) 'The Role of Polypyrimidine Tract-Binding Proteins and Other hnRNP Proteins in Plant Splicing Regulation.', *Frontiers in Plant Science*, 3, p. 81. doi: 10.3389/fpls.2012.00081.

Wahid, F. *et al.* (2010) 'MicroRNAs: Synthesis, mechanism, function, and recent clinical trials', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1803(11), pp. 1231–1243. doi: 10.1016/j.bbamcr.2010.06.013.

Wahl, M. C., Will, C. L. and Lührmann, R. (2009) 'The Spliceosome: Design Principles of a Dynamic RNP Machine', *Cell*, 136(4), pp. 701–718. doi: 10.1016/j.cell.2009.02.009.

Walley, J. W. *et al.* (2016) 'Integration of omic networks in a developmental atlas of maize',

Science, 353(6301), pp. 814–818. doi: 10.1126/science.aag1125.

Wan, J. *et al.* (2013) 'Integrative analysis of tissue-specific methylation and alternative splicing identifies conserved transcription factor binding motifs', *Nucleic Acids Research*, 41(18), pp. 8503–8514. doi: 10.1093/nar/gkt652.

Wang, M. *et al.* (2018) 'A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation', *New Phytologist*, 217(1), pp. 163–178. doi: 10.1111/nph.14762.

Wang, X. *et al.* (2016) 'DNA Methylation Affects Gene Alternative Splicing in Plants: An Example from Rice', *Molecular Plant*, 9(2), pp. 305–307. doi: 10.1016/j.molp.2015.09.016.

Wang, X. *et al.* (2018) 'Detection of Proteome Diversity Resulted from Alternative Splicing is Limited by Trypsin Cleavage Specificity', *Molecular & Cellular Proteomics*, 17(3), pp. 422–430. doi: 10.1074/mcp.RA117.000155.

Wang, X., Zhang, B. and Wren, J. (2013) 'CustomProDB: An R package to generate customized protein databases from RNA-Seq data for proteomics search', *Bioinformatics*, 29(24), pp. 3235–3237. doi: 10.1093/bioinformatics/btt543.

Weatheritt, R. J., Sterne-Weiler, T. and Blencowe, B. J. (2016) 'The ribosome-engaged landscape of alternative splicing', *Nature Structural and Molecular Biology*, 23(12), pp. 1117–1123. doi: 10.1038/nsmb.3317.

Wei, G. *et al.* (2018a) 'Position-specific intron retention is mediated by the histone methyltransferase SDG725', *BMC Biology*, 16(1), p. 44. doi: 10.1186/s12915-018-0513-8.

Will, C. L. and Lührmann, R. (2011) 'Spliceosome structure and function', *Cold Spring Harbor Perspectives in Biology*, 3(7), pp. 1–2. doi: 10.1101/cshperspect.a003707.

Wilusz, J. E. (2018) 'A 360° view of circular RNAs: From biogenesis to functions', *Wiley Interdisciplinary Reviews: RNA, 9*, p. e1478. doi: 10.1002/wrna.1478.

Wippich, F. *et al.* (2013) 'Dual specificity kinase DYRK3 couples stress granule condensation/ dissolution to mTORC1 signaling', *Cell*. Elsevier Inc., 152(4), pp. 791–805. doi: 10.1016/j.cell.2013.01.033.

Wojtowicz, W. M. *et al.* (2004) 'Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding', *Cell*, 118(5), pp. 619–33. doi: 10.1016/j.cell.2004.08.021.

Wu, L. *et al.* (2010) 'DNA Methylation Mediated by a MicroRNA Pathway', *Molecular Cell*, 38(3), pp. 465–475. doi: 10.1016/j.molcel.2010.03.008.

Wu, Z. *et al.* (2016) 'Quantitative regulation of *FLC* via coordinated transcriptional initiation and elongation', *Proceedings of the National Academy of Sciences*, 113(1), pp. 218–223. doi: 10.1073/pnas.1518369112.

Xiong, J. and Bauer, C. E. (2002) 'Complex evolution of photosynthesis', *Annual Review of Plant Biology*, 53, pp. 503–21. doi: 10.1146/annurev.arplant.53.100301.135212.

Xu, G. *et al.* (2017) 'Global translational reprogramming is a fundamental layer of immune regulation in plants', *Nature*, 545(7655), pp. 487–490. doi: 10.1038/nature22371.

Xu, W. *et al.* (2017) 'The R-loop is a common chromatin feature of the Arabidopsis genome', *Nature Plants*, 3(9), pp. 704–714. doi: 10.1038/s41477-017-0004-x.

Yang, W., Wightman, R. and Meyerowitz, E. M. (2017) 'Cell Cycle Control by Nuclear Sequestration of CDC20 and CDH1 mRNA in Plant Stem Cells', *Molecular Cell*, 68(6), p. 1108–1119.e3. doi: 10.1016/j.molcel.2017.11.008.

Yang, X. *et al.* (2016) 'Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing', *Cell*, 164(4), pp. 805–817. doi: 10.1016/j.cell.2016.01.029.

Ye, C. Y. *et al.* (2015) 'Widespread noncoding circular RNAs in plants', *New Phytologist*, 208(1), pp. 88–95. doi: 10.1111/nph.13585.

Yu, H. *et al.* (2016) 'Transcriptome Survey of the Contribution of Alternative Splicing to Proteome Diversity in Arabidopsis thaliana', *Molecular Plant*, 9(5), pp. 749–752. doi: 10.1016/j.molp.2015.12.018.

Zhang, B. (2015) 'MicroRNA: A new target for improving plant tolerance to abiotic stress', *Journal of Experimental Botany*, 66(7), pp. 1749–1761. doi: 10.1093/jxb/erv013.

Zhang, P. *et al.* (2013) 'Alterations of alternative splicing patterns of Ser/Arg-Rich (SR) genes in response to hormones and stresses treatments in different ecotypes of rice (Oryza sativa)', *Journal of Integrative Agriculture*, 12(5), pp. 737–748. doi: 10.1016/S2095-3119(13)60260-9.

Zhang, P. *et al.* (2017) 'PlantCircNet: a database for plant circRNA–miRNA–mRNA regulatory networks', *Database*, 2017, pp. 89–1. doi: 10.1093/database/bax089.

Zhang, Z. *et al.* (2011) '*Arabidopsis* Floral Initiator SKB1 Confers High Salt Tolerance by Regulating Transcription and Pre-mRNA Splicing through Altering Histone H4R3 and Small Nuclear Ribonucleoprotein LSM4 Methylation', *The Plant Cell*, 23(1), pp. 396–411. doi: 10.1105/tpc.110.081356.

Zhao, T. *et al.* (2017) 'Characterization of conserved circular RNA in polyploid Gossypium species and their ancestors', *FEBS Letters*, 591(21), pp. 3660–3669. doi: 10.1002/1873-3468.12868.

Zhong, S. *et al.* (2008) 'MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor.', *The Plant cell*, 20(5), pp. 1278–88. doi: 10.1105/tpc.108.058883.

Zhu, J.-K. (2016) 'Abiotic Stress Signaling and Responses in Plants', *Cell*, 167(2), pp. 313–324. doi: 10.1016/j.cell.2016.08.029.

Zhu, J. *et al.* (2018) 'RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis', *Nature Plants*. Springer US, 4(12), pp. 1112–1123. doi: 10.1038/s41477-018-0280-0.

Zong, W. *et al.* (2013) 'Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress', *Plant Molecular Biology*, 81(1–2), pp. 175–188. doi: 10.1007/s11103-012-9990-2.

Zuo, Y. C. and Li, Q. Z. (2011) 'Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility', *Genomics*, 97(2), pp. 112–120. doi: 10.1016/j.ygeno.2010.11.002.

Chapter 2. Materials and Methods

In this chapter, the general and common experimental procedures for each of the results chapters are described here. All laboratory experiments were conducted in Naeem Syed's lab. Materials and methods specific to each of the results chapters are described in Chapters 3, 4.

2.1 Plant material

In this study, Arabidopsis thaliana, Columbia (Col-0) ecotype was used. Plant material used for RNA-seq and MNase-seq were wild type Col-0 plants and Col-0 plants treated with 5-aza-2′-deoxycytosine (5-aza-dC, Sigma cat # A3656) whereas, for WGBS only Col-0 plants treated 5-aza-dc plants were used. Col-0 wild type seeds were obtained from Nottingham Arabidopsis Stock Centre (NASC) and used as controls in RNA-seq and MNase-seq experiments.

2.2 Seeds Sterilisation

For RNA-Seq, MNase-Seq and WGBS wild type seeds of Col-0 plants were surface sterilized in one batch as follows. Each 1 millilitre (ml) of seeds were covered with 2 ml of sterilization solution consisting of 20% household bleach and 0.05% Tween 20 (Sigma cat# 93773), then vortexed at maximal speed for 30 seconds. Vortexed seeds were then placed for 10 minutes at room temperature with occasional vortex every 2 minutes. Afterwards, seeds were spun at 8,000 revolutions per minute (rpm) for 5 seconds. Then, supernatants were removed and 1 ml of distilled water was added to the seeds, followed by vortexing to suspend the seeds. Next, seeds were spun down at 8,000 rpm for 5 seconds followed by discarding the supernatant. Washing with 1 ml distilled water followed by vortexing and spinning down at 8,000 rotation per minute for 5 seconds were repeated 7 times. Surface sterilized seeds were then transferred to a small petri dishes (Size 100 mm × 20 nm, Sigma cat # P5481-500EA) with a wet Whatman filter paper (Sigma cat# WHA2200185) on its bottom.

2.3 Growth agar medium plates preparation and 5-aza-2′-deoxycytosine treatment

Sterilized seeds were grown on agar medium with or without 5-aza-dC as follows. 4.31 gram (g) of Murashige and Skoog (MS) basal salt mixture and 0.5 g of 2-(N-Morpholino) ethanesulfonic acid (MES) were mixed in 0.8 litre (L) of autoclaved distilled water. After stirring, pH was then adjusted to 5.7 using 1M KOH followed by adding 10 g of agar to a final volume of 1L. Then, prepared media was autoclaved with a magnetic stirrer for 1 hour and 10 minutes by maintaining a temperature of 121°C at 15 psi of pressure. Afterwards, autoclaved media was placed under the hood to cool down while stirring. Then, 1 g of sucrose was added to a final volume of 1 L of media. Finally, 500 ml of the prepared media was poured into sterile petri dishes containing 20 ml of media each to obtain agar medium plates, while the remaining of the media was kept to prepare plates with 5-aza-dC treatment.

To obtain agar medium with 5-aza-dC treatment, 5 mg of 5-aza-dC was dissolved at 50 mg/ml in distilled water and was added to the remaining media to a final concentration of 4 µg/ml. After stirring, media containing 5-aza-dC was poured into petri dishes containing 20 ml of media each.

Agar medium plates with and without 5-aza-dC treatment were kept under the hood for nearly 1 hour to consolidate. Afterwards, 200 µl of surfaced sterilized seeds suspended in autoclaved distilled water were pipetted onto both types of agar medium plates. After spreading the seeds
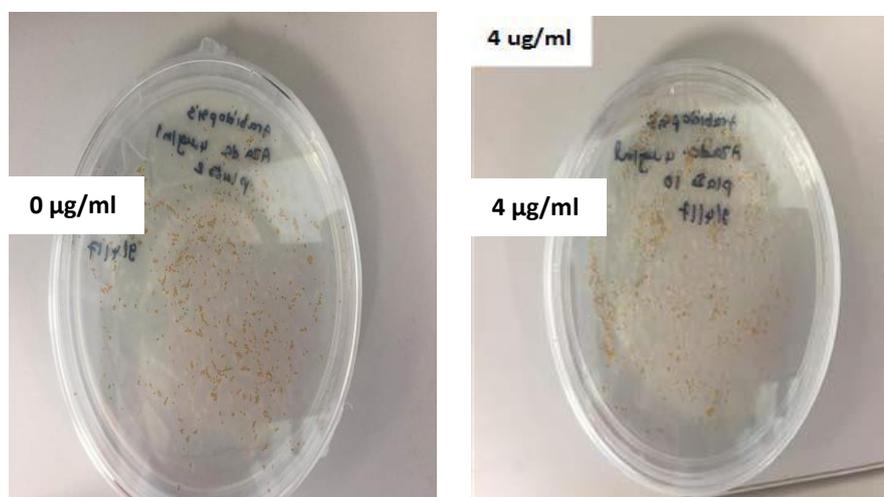
equally over the plate, water residues were removed from the plate using a pipette. Plates were then wrapped with Parafilm® and placed at 4ºC in the dark for 4 days to allow seeds stratification (Figure 2.1)

Figure 2.1. Surface sterilized Arabidopsis Col-0 seeds plated on agar medium plates without (0 µg/ml) and with (4 µg/ml) 5-aza-dC treatment.

Afterwards, stratified seeds were grown in environment controlled cabinets (Fitotron®, England) at 22ºC under a 16 hours light (130 μE.m^{-2}.s^{-1}) and 8 hours dark period, in an environment of 50% relative humidity.

2.4 Soil-transferred Arabidopsis plants

One week after germination (Figure 2.2), Arabidopsis seedlings treated and non-treated with 5-aza-dC were transferred to separate pots containing compost with perlite (Table 2.1) to provide non-stressful environment and space to get sufficient amount of tissue required during the harvesting stage. Soil-grown Arabidopsis seedlings were then grown for 3 weeks in environment controlled cabinets (Fitotron®, England) at 22ºC under a 16 hours light (130 μE.m^{-2}.s^{-1}) and 8 hours dark period, in an environment of 50% relative humidity. Arabidopsis plants were watered every 2-3 days depending on the soil moisture.

0 µg/ml **4 µg/ml**
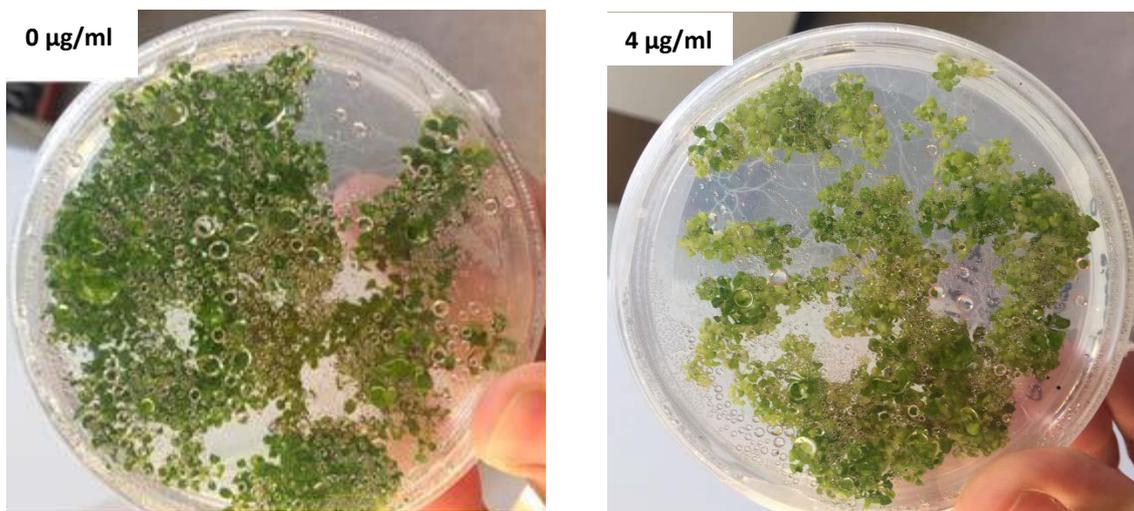
Figure 2.2. Arabidopsis seedlings after one week of growth on agar medium without (0 µg/ml) and with (4 µg/ml) 5-aza-dC treatment. Compared with their respective control, 5-aza-dC treated plants displayed relatively frequent abnormalities. The most common abnormality was a semidwarf phenotype that displayed many secondary inflorescences. Other examples of phenotypic abnormalities included petals with less chlorophyll content, dwarfism, and reduced roots size. In certain cases, the severity of abnormalities changed in intensity along growth. In contrast, untreated plants displayed the normal growth phenotype **(Figure 2.1)**. The observed phenotypic differences suggest that 5-aza-dC compromised mechanisms of epigenetic gene regulation which resulted in the development of altered morphologies.

Table 2.1. The composition of the compost used for growing Arabidopsis plants

Mixture	Volume
Levington M2 (peat)	450 L
Grit	25 Kg
Intercept	190 kg

2.5. Cold treatment and tissue harvesting

After 3 weeks of growth on soil as described in section 2.3, shoots of half of the pots sown with Arabidopsis plants treated and non-treated with 5-azadC were harvested four hours after subjective dawn at 22°C. At the end of sampling, the temperature was reduced from 22°C to 4°C. Harvesting continued for the remaining samples after 24 hours from cold treatment. For both temperature treatment (22°C and 4°C), three biological replicates were collected at the same time to avoid variation in light and temperature. Collected tissues were flash frozen in liquid nitrogen and stored in -80°C until use for total RNA and genomic DNA extraction as well as nuclei isolation as indicated in chapter 3 and 4. A summary of the experimental model followed in this study is schematized in figure 2.3.
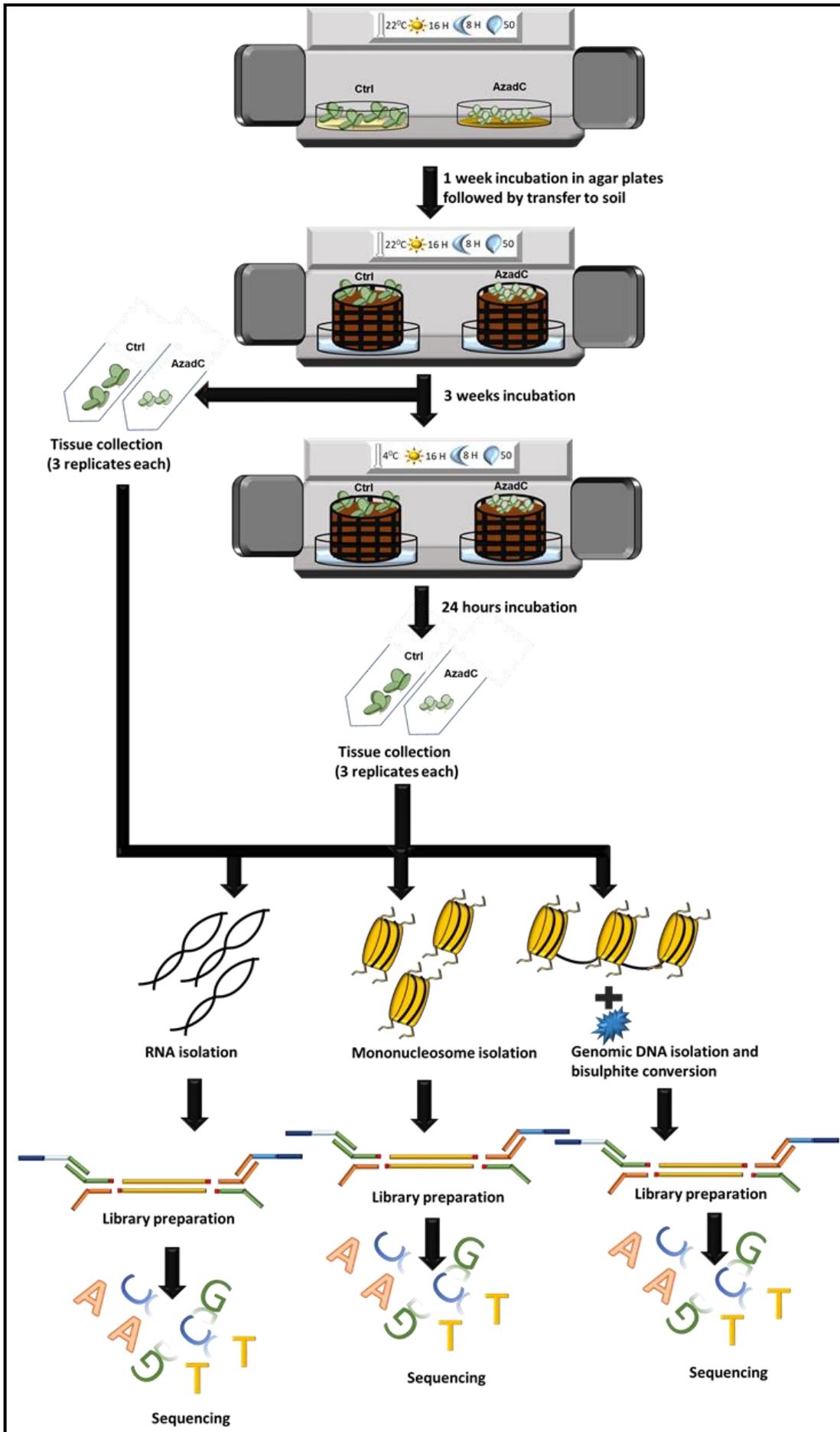
Figure 2.3. Scheme summarizing the experimental procedure. Arabidopsis seeds were grown on media without (control plants) and with chromatin-modifying agent (AzadC plants). After 3 weeks incubation at 22°C, leaf tissues (3 replicates) were collected from each sample and then plants were shifted to 4°C. Leaf tissues (3 replicates) were collected again after 24 hours cold treatment. RNA, mononucleosomes, and total genomic DNA were then isolated for library preparation and sequencing. The white panel of the incubator represent growth conditions. The thermometer indicates the growth temperature, the sun and moon indicates the light and dark cycle lengths respectively, and the drop indicates the humidity. Ctrl and AzadC are Arabidopsis plants without and with 5-AzadC treatment respectively.

Chapter 3. Identification of differentially expressed and alternatively spliced genes in epigenetically different Arabidopsis plants with identical genetic background

3.1 Introduction

Plants employ different genetic and epigenetic strategies to fine-tune their transcriptional responses during daily cycles and under stress to support life and confer adaptive responses (J. Liu *et al.*, 2015; Yu *et al.*, 2016). Emerging evidence shows that modulating co-transcriptional alternative splicing (AS) may be a key gene regulatory mechanism in plants (Syed *et al.*, 2012; Reddy *et al.*, 2013; Jabre *et al.*, 2019). Alternative splicing of pre-mRNA uses alternative splice sites to generate multiple transcripts from a single gene. In plants, the majority of intron-containing genes (up to 70%) are alternatively spliced (Zhang *et al.*, 2010; Marquez *et al.*, 2012; Chamala *et al.*, 2015) and contribute towards transcriptome diversity and potentially proteome complexity, in response to abiotic and biotic stresses such as cold, drought, heat, and pathogen infection in a tissue- and cell-specific manner, and during different development stages (Mastrangelo *et al.*, 2012; Calixto *et al.*, 2018; Filichkin *et al.*, 2018; Chaudhary *et al.*, 2019). Alternative splicing also plays an important role in regulating the transcript isoform levels of key circadian clock genes in Arabidopsis (A. B. James *et al.*, 2012; S. A. Filichkin *et al.*, 2015).

Regulation of AS is orchestrated by the abundance of different splicing factors (SFs) recognizing various *cis*-regulatory elements in pre-mRNA in a cell type- and condition-dependent manner (Shen, Julie L C Kan and Green, 2004; Y. Ding *et al.*, 2014). Variation in DNA sequence (*cis*-regulatory elements) can impact splicing outcomes however, emerging evidence shows that the chromatin environment such as DNA methylation and nucleosome occupancy also has a strong bearing on the splicing process by modulating RNA polymerase II (RNAPII) processivity and SFs recruitment (Hirose and Manley, 2000; Gasch *et al.*, 2006; Lenasi and Barboric, 2010; Hajheidari, Koncz and Eick, 2013). In eukaryotes, DNA methylation occurs in symmetric CG and CHG (H = A, T or C) and asymmetric CHH contexts (Ehrlich *et al.*, 1982). However, DNA methylation is largely dependent on the CpG context in plants. In the Arabidopsis genome, 24% of CG sites are methylated, compared with only 6.7% of CHG and 1.7% of CHH sites (Cokus *et al.*, 2008; Lister *et al.*, 2008). Constitutive exons have higher CG methylation content and nucleosome occupancy levels compared to introns

and alternative exons in animals and plants (Mavrich *et al.*, 2008; Nahkuri, Taft and Mattick, 2009; Schwartz, Meshorer and Ast, 2009; Tilgner *et al.*, 2009; Chen, Luo and Zhang, 2010; Chodavarapu *et al.*, 2010; Gelfman *et al.*, 2013; M.-J. Liu *et al.*, 2015). DNA methylation is also higher in nucleosome bound DNA in both humans and Arabidopsis affecting chromatin compaction/remodelling (Chodavarapu *et al.*, 2010; Huff and Zilberman, 2014; M.-J. Liu *et al.*, 2015). It is not surprising that AS is emerging largely as a co-transcriptional process since RNAPII speed is affected by the chromatin state that in turn affects splicing outcomes (Alexander *et al.*, 2010; Ullah *et al.*, 2018; Zhu *et al.*, 2018). Native elongating transcript NET-seq and GRO-seq data from Arabidopsis show that phosphorylation of RNAPII CTD mediates interactions with the spliceosome and that RNAPII accumulation is associated with different chromatin states (Zhu *et al.*, 2018).

Plants exhibit stable as well as dynamic DNA methylation patterns under different growth and stress conditions that provide the template through which gene expression and AS are modulated in a condition-specific manner (Steward *et al.*, 2002; Dowen *et al.*, 2012; Garg *et al.*, 2015; Secco *et al.*, 2015; Lu *et al.*, 2017). However, the relationship between DNA methylation and nucleosome occupancy under different stresses, generations and growth conditions is not clear. Recent evidence shows that stress-induced DNA methylation can influence transgenerational epigenetic memories in plants (Luna *et al.*, 2012; Rasmann *et al.*, 2012). However, it remains to be seen whether underlying DNA methylation patterns could affect chromatin architecture (i.e. nucleosome occupancy) and provide a reproducible context through which AS patterns can be modulated in a wide range of physiological processes including stress responses (Jabre *et al.*, 2019). In this way, a dynamic but reproducible chromatin environment could modulate transcription and AS to mediate appropriate growth and stress responses. Such a scheme could also be part of a stress- and condition-dependent splicing memory that provides a dynamic yet reproducible response as and when required. Recent data shows that plants indeed possess splicing memory and display a reproducible splicing pattern under normal conditions and high temperature stress (Ling *et al.*, 2018).

Regulation of AS is dependent on the genetic as well as the epigenetic context (Reddy *et al.*, 2013); however, it is unclear to which extent genetic and epigenetic differences mediate AS outcomes. To answer this, plants with identical DNA sequence but differential DNA methylation and nucleosome occupancy has been used to reveal how differences in epigenetic landscapes could influence splicing dynamics without the confounding effects of sequence variation. Towards this goal, Arabidopsis seedlings of Columbia-0 (Col-0) ecotype have been

treated with 5-aza-2′-deoxycytosine (5-aza-dC) to reduce DNA methylation levels. 5-aza-dC is a nucleoside analogue of cytosine that inhibits DNA methyltransferases, resulting in hypomethylation and gene activation through uncoiling of constitutive heterochromatin (Christman, 2002). 5-aza-dC has been shown to create heritable hypomethylation and phenotypic trait variation in rice (Sano *et al.*, 1990; Kumpatla *et al.*, 1997), flax (Fieldes, 1994), tobacco (Vyskot *et al.*, 1995), *Brassica* (King, 1995), *Melandrium album* (Janoušek, Široký and Vyskot, 1996), triticale (Amado *et al.*, 1997), Arabidopsis (BURN *et al.*, 1993), *Fragaria vesca* (Xu *et al.*, 2016) , and *Solanum ruiz-lealii* (Marfil, Asurmendi and Masuelli, 2012).

In Arabidopsis, cold induces a cascade of gene expression reprogramming to modulate their transcriptome and proteome (Thomashow, 2010; Knight and Knight, 2012; Barrero-Gil and Salinas, 2013). Collective data show that AS is the hub of cold-stress responses in plants (Palusa, Ali and Reddy, 2007; Calixto *et al.*, 2018; Filichkin *et al.*, 2018). For example, in Arabidopsis, cold-dependent AS of the *LATE ELONGATED HYPOCOTYL (LHY)* gene generates different transcripts with variable abundance (A. B. James *et al.*, 2012) and recently, co-transcriptional regulation of *LHY* pre-mRNA splicing under cold stress has been proposed to be regulated by the chromatin structure (Jabre *et al.*, 2019). Cold-induced DNA methylation and nucleosome occupancy changes are relatively rapid epigenetic regulators that mediate environmental cues and provide flexible cold responses in rice, Arabidopsis and maize (Steward *et al.*, 2002; Kumar and Wigge, 2010; McClung and Davis, 2010; Pan *et al.*, 2011; Roy *et al.*, 2014). Therefore, cold treatment has been used as a system of choice to understand how epigenetic differences could influence AS outcomes under normal (22^0C) and cold (4^0C) growth conditions.

 In this chapter, differential gene expression at the gene and transcript level have been examined from RNA-seq data of wild type (Ctrl) and 5-aza-dC treated (AzadC) plants under normal growth conditions and cold stress. The major findings in this chapter are that epigenetic features are likely to be involved in regulating gene expression and AS profiles in Arabidopsis upon cold temperature, which results in reprogramming plants transcriptome to adapt environmental changes. Importantly, the results of the RNA-seq analysis show that epigenetic differences alone are sufficient to modulate global variation in gene expression at the gene and splicing level in plants with identical DNA sequence.

3.2 Materials and Methods

3.2.1 Total RNA extraction

Arabidopsis leaf tissues collected in three biological replicates from Ctrl and AzadC plants under normal and cold stress (as described in chapter 2 section 4) were finely ground with liquid nitrogen using mortar and pestles. Then, total RNA was extracted from 100 milligrams (mg) of ground frozen leaf tissue using the RNeasy Plant Mini Kit (Qiagen), then on-column DNase treatment was applied to remove DNA contamination according to manufacturer's instructions. Initial quality control (QC) of the RNA extracted from 12 samples (3 biological replicates for Ctrl and AzadC plants grown at 22ºC and 4ºC for 24 hours) was performed at the Earlham institute and involved RNA concentration measurement using Qubit RNA (Life technologies Q32852) assays, as well as a quality check using the Bioanalyser with the Nano kit (Agilent 5067-1511).

3.2.2 Library preparation

RNA-Seq libraries were prepared at the Earlham institute as follows using the TruSeq RNA protocol–with amendments (Illumina 15026495 Rev.F). After passing initial QC as described in 3.2.1, poly(A) biotin beads have been used to pull down mRNA from 1 micrograms (µg) of RNA purified from each sample. Then, mRNAs were fragmented at 94°C for 6 minutes followed by first strand cDNA synthesis using the following program: 25°C for 10 minutes, 42°C for 50 minutes, 70°C for 15 minutes, and final hold at 4°C. This process consists of transcribing the cleaved RNA fragments primed with random hexamers into first strand cDNA using reverse transcriptase and random primers.

To generate double stranded cDNA, the RNA template has been removed by incubating the product from the first step at 16°C for 1 hour followed by a second strand cDNA synthesis. cDNA is then purified using a 0.8x clean up using Beckman Coulter XP beads (Beckman Coulter A63880). The ends of the samples were repaired using the 3' to 5' exonuclease activity to remove the 3' overhangs and the polymerase activity to fill in the 5' overhangs creating blunt ends by incubating at 30°C for 30 minutes.

To prevent fragments ligation to each other during the adaptor ligation process, a single 'A' nucleotide was added to the 3' ends of the blunt fragments. For that, reaction samples are incubated for 30 minutes at 37ºC, followed by 5 minutes at 70ºC. Afterwards, corresponding single 'T' nucleotides have been added to the 3' end of the adapter to provide a complementary overhang for ligating the adapter to the fragment. This strategy ensured a low rate of chimera

formation. For the adapter ligation reaction step, Bio LT adapters (Newmarket Scientific 514103), have been diluted from their stock concentration at 25 micromolars (µM) to 6 µM to be added to the ends of the DNA fragments by incubation at 30°C for 10 minutes which prepared them for hybridisation onto a flow cell.

To remove the majority of un-ligated adapters, as well as any adapters that may have ligated to one another, the ligated products were subjected to a 0.8x bead based size selection using Beckman Coulter XP beads (Beckman Coulter A63880).

Prior to hybridisation to the flow cell, PCR enrichment for samples has been performed using a PCR primer cocktail that annealed to the ends of the adapter using the following program: 98°C for 30 seconds, 10 cycles (98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds), followed by 72°C for 5 minutes and final hold at 4°C.

Following bead clean-up (0.8x) the final libraries were resuspended in 30 microliter (µl) RSB. The insert size of the libraries was verified by running an aliquot of the library on the Agilent bioanalyser using the High Sensitivity chip (Agilent 5067-4626) and the concentration was determined by using a High Sensitivity Qubit assay (ThermoFisher Q32854). A 16-plex equimolar pool was prepared and checked by quantitative polymerase chain reaction (q-PCR), before preparing and loading for sequencing on the Hiseq 4000 (Illumina) using 150 paired end reads across 3 lanes.

3.2.3 RNA sequencing procedure

The constructed RNA libraries were normalised and equimolar pooled, the final pool was quantified using a KAPA Library Quant Kit (Roche Diagnostics Limited) and found to be 9.73 nanomolars (nM). The library pool was diluted to 3 nM and spiked with 1% PhiX Control V3 (Illumina FC-110-3001). Then, libraries were denatured with NaOH and neutralised with Tris before addition of Illumina's ExAmp mix and loading onto the Illumina cBot, to give a final loading concentration of 300 pM. The flow cell was clustered using a HiSeq 4000 PE Cluster Kit (Illumina, PE-410-1001), utilising the Illumina HiSeq_3000_4000_HD_Exclusion_Amp_v1.0 method on the Illumina cBot. Following clustering, the patterned flow cell was loaded onto the Illumina HiSeq4000 instrument following the manufacturer's instructions. Each paired sequencing read was 151bp long. The sequencing chemistry used was HiSeq 4000 SBS Kit (Illumina, FC-410-1003) with HiSeq

Control Software 3.3.52 and RTA 2.7.3. Reads in binary base call (bcl) format were converted to FASTQ format by bcl2fastq2 (Illumina). Below a table summarising the number of reads generated from each sample (Table 3.1).

Table 3.1. RNA sequencing reads information generated from all samples

Sample name	Number of reads	Mean Q30 to base Read 1	Mean Q30 to base Read 2	Lane
Ctrl_22°C_R1	20812626	151	109	7
Ctrl_22°C_R1	21,289,259	151	109	8
Ctrl_22°C_R2	17663509	151	119	7
Ctrl_22°C_R2	17,951,723	151	119	8
Ctrl_22°C_R3	22,650,793	150	149	7
Ctrl_22°C_R3	19,572,266	150	150	8
Ctrl_4°C_R1	25,453,951	151	114	7
Ctrl_4°C_R1	25,848,035	151	114	8
Ctrl_4°C_R2	20287545	151	114	7
Ctrl_4°C_R2	20,602,886	151	114	8
Ctrl_4°C_R3	26,876,386	150	150	7
Ctrl_4°C_R3	22,021,724	150	150	8
AzadC_22°C_R1	19497848	151	114	7
AzadC_22°C_R1	19,813,534	151	114	8
AzadC_4°C_R2	22116346	151	119	7
AzadC_22°C_R2	22,434,922	151	114	8
AzadC_22°C_R3	21,139,402	150	150	7
AzadC_22°C_R3	23,887,951	150	150	8
AzadC_4°C_R1	20593696	151	119	7

AzadC_4°C_R1	20,955,626	151	114	8
AzadC_4°C_R2	22116346	151	119	7
AzadC_4°C_R2	22,479,617	151	114	8
AzadC_4°C_R3	22,402,905	150	150	7
AzadC_4°C_R3	24,212,770	150	150	8

3.3. Bioinformatics analysis of RNA-Sequencing data

RNA-seq data analysis has been performed at the James Hutton Institute under the supervision of Runxuan Zhang and Wenbin Guo using a combination of Linux command lines and R packages. To begin with, FASTAQ files obtained from Earlham sequencing facility have been QC'd using Fastqc version 0.11.8 and trimmed using trimmomatic version 0.32 with default parameters. Then, Salmon version 0.8.2 has been used for transcript quantification followed by differential gene expression and alternative splicing (DE and DAS respectively) analysis using different R packages. Furthermore, SUPPA version 2 has been used to obtain differentially alternatively spliced genes for local AS events. Henceforth, in this section, the details of the bioinformatics analysis of the RNA-seq data to obtain transcript quantification, DE/DAS, and differential AS events are described.

3.3.1 Transcript quantification

To quantify transcript-level abundances, Salmon requires a reference transcriptome in the form of FASTA file, containing the sequence of a transcript in each entry. For that, the quasi mapping mode has been used to build an auxiliary k-mer of length 31 (–type quasi –k 31) using Arabidopsis thaliana reference transcriptome dataset version 2-QUASI (AtRTDv2-QUASI). Quasi-mapping technique refers to lightweight-alignment and pseudoalignment, that allows rapid and accurate mapping of sequenced reads to the reference transcriptome to find the *best* mappings (targets and positions) for each read, and does so (approximately) by finding minimal collections of dynamically sized, right-maximal, matching contexts between target and query positions. AtRTD2-QUASI has been used as a modified version of AtRTD2, which is a high-quality reference transcript data set for Arabidopsis Col-0 containing > 82,000 unique transcripts (R. Zhang *et al.*, 2017). Here, this database has been used since it was designed specifically for the accurate quantification of individual transcript expression for AS analysis.

Transcript quantification has been performed using trimmed RNA-seq reads (FASTAQ files) of each sample and the indexed reference transcriptome. Salmon transcript quantification has been run with the following extra parameters: --useVBOpt (Optimise transcript abundance estimate), --numBootstraps 30 (Assess technical variance in the main abundance estimates produced by Salmon), –seqBias (Correct for the sequence specific bias). Otherwise, all other parameters were on default settings. Once transcript quantification has been completed by Salmon, a file named quant.sf will be generated containing transcript abundance in transcripts per million (TPM).

3.3.2 Differential gene expression and alternative splicing analysis pipeline

To obtain DE and DAS genes, transcript quantification files (quant.sf) obtained by Salmon were processed by different R packages as follows.

3.3.2.1 Transcript and gene read counts generation

For this step, two "csv" (comma delimited) spreadsheets are required in addition to the quant.sf files generated by Salmon. The first "csv" spreadsheet contains the information of experimental design, including treatments, biological replicates, sequencing replicates, and quantification file names and directory. The second one is a "csv" spreadsheet with the first and second column listing transcript names and gene IDs, respectively. The second "csv" spreadsheet will help relating transcript names to gene IDs in order to summarise transcript level quantifications to gene level expression. Once these files are obtained, tximport version 0.99.2 R can be used at the gene and transcript level to convert TPM values (4th column of salmon quant.sf output) to read counts, and with the option "lengthScaledTPM" turned on to correct possible gene length variations across samples.

3.3.2.2 Data pre-processing

3.3.2.2.1 Merging sequencing replicates

Given that RNA-seq has been performed on different lanes for each sample (Sequencing replicates), the first step of data pre-processing consists of merging sequencing replicates for each sample to increase sequencing depth.

3.3.2.2.2 Filtering low expressed transcripts

To remove lowly expressed transcripts and genes, the decreasing trend between means and variances has been analyzed. While read counts follow negative binomial distribution, the

expression of lowly expressed transcripts follows a different distribution. This results in a decrease of the variance of $\log_2$transformed read counts with the increase of mean, and a drop of mean-variance trend towards low values of $\log_2$ read counts. This can be solved by removal of lowly expressed transcripts on the basis of a threshold that an expressed transcript should have a minimum count per million (CPM), *n*, in at least *m* samples. Hence, providing optimal conditions for filtering low expressed transcripts.

In this RNA-seq data, the decreasing trend at the low expression end of the mean-variance plot is removed when removing transcripts that did not have ≥1 CPM in two or more samples out of 12 (Figure 3.1). At the gene level, if any transcript passed the expression level filtering step, the gene was categorized as an expressed gene.
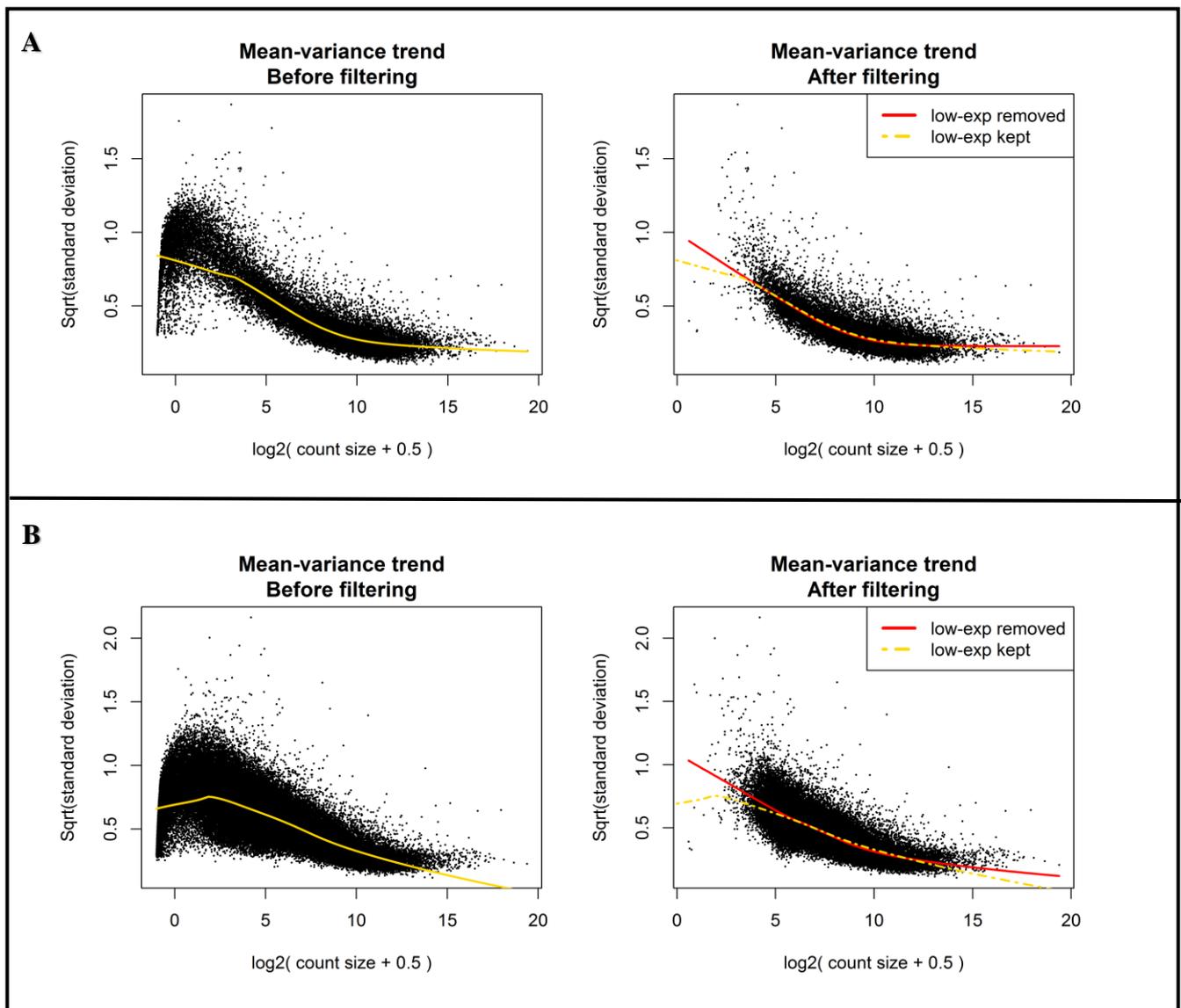
Figure 3.1. Mean-variance trend plot used to filter low expressed transcripts (A) and genes (B) from RNA-seq data. Each black point represents a transcript. The red and yellow curves are the fitted trends of these points. The red circle in plot A) highlights the drop trend of low expressed transcripts. By using a cut-off of cpm=1 and sample=2, the drop observed in the mean-variance trend plot of the raw counts has been removed at the transcript (A) and gene level (B).

3.3.2.2.3 Principal components analysis

To investigate if the RNA-seq data is affected by unwanted variation (batch effects), principal components analysis (PCA) at the gene and transcript level has been performed using FactoMineR version 1.42 and Factoextra version 1.0.5. In this case, two PCA dimensions are used at the transcript and gene level to visualize data variance of two variables; temperature and 5-aza-dC treatment (Figure 3.2). PCA analysis at the gene and transcript levels show that

the RNA-seq data is not biased by batch effect and that the three biological replicates of each condition treatment are highly reproducible.
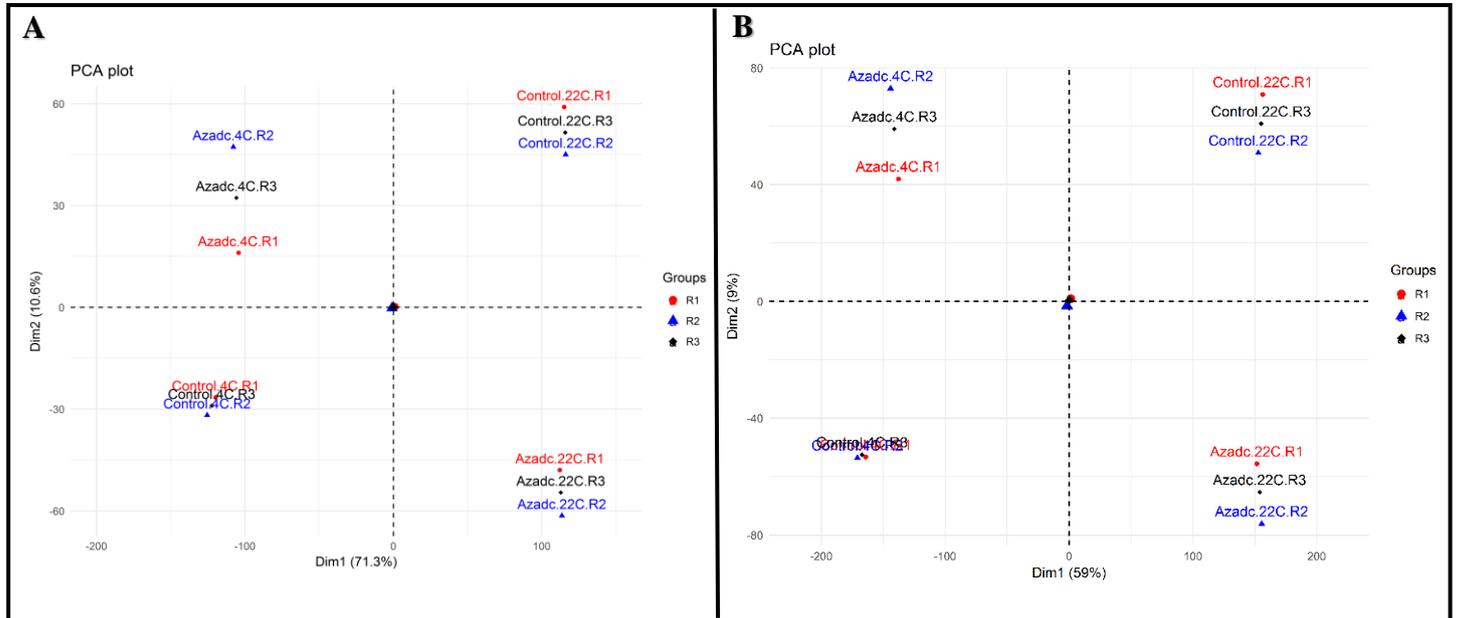
Figure 3.2 PCA plots of transcript (A) and gene (B) expression levels from RNA-seq data. Principal components dimensions (Dim)1 and Dim2 corresponds to temperature and 5-aza-dC treatment, respectively. R= replicate. Control and AzadC are Ctrl and azadC plants respectively.

3.3.2.2.4 Data normalization

The final step before identifying DE and DAS genes, is to normalize the data using CPM to $\log_2$ transformation method to reduce sequencing bias and the false positives for highly abundant transcript outliers. The normalization factor, which accounted for the raw library size, was estimated using the weighted trimmed mean of M values method using edgeR version 3.12.1. Read count distributions before and after normalization are then visualized using FactoMineR version 1.42 and Factoextra version 1.0.5 (Figure 3.3)
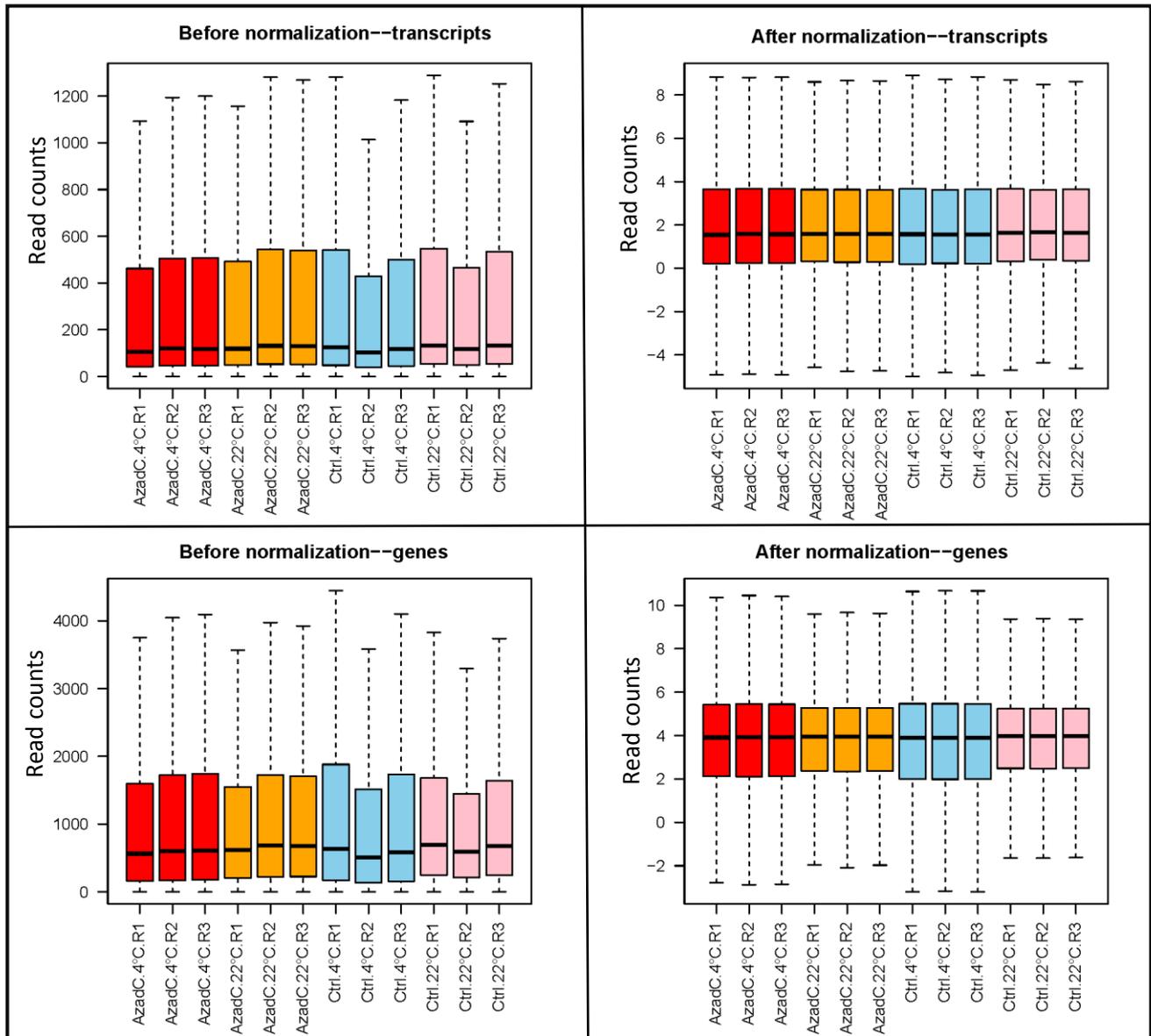
Figure 3.3. Box plots showing read counts distribution from RNA-seq data before and after normalization at the gene and transcript level for each sample. Before normalization at the gene and transcript level, read counts display different distribution between different samples and biological replicates (Different median represented by different levels of black line in the middle of each box plot, and different distribution represented by different upper quartile levels between samples and different quartile groups). However, after normalisation of read counts at the gene and transcript level, all samples present the same median.

3.3.2.3 Identification of differentially expressed and alternatively spliced genes

At both gene and transcript levels, a general linear model was used to determine differential expression using temperature and 5-azadC treatment as factors. Then, six contrast groups are used where Ctrl plants are compared to azadC plants at 22°C or 4°C (AzadC 22°C Vs Ctrl 22°C and AzadC 4°C Vs Ctrl 4°C), Ctrl or AzadC plants grown at 22° were compared to those grown

at 4ºC (Ctrl 4ºC Vs Ctrl 22ºC and AzadC 4ºC Vs AzadC 22ºC), AzadC plants grown at 22ºC were compared to Ctrl plants grown at 4ºC (AzadC 22ºC Vs Ctrl 4ºC), and AzadC plants grown at 4ºC were compared to Ctrl plants grown at 22ºC (Ctrl 22ºC Vs AzadC 4ºC). The first two contrast groups (AzadC 22ºC Vs Ctrl 22ºC and AzadC 4ºC Vs Ctrl 4ºC) aims to detect the effect of DNA methylation inhibitor on gene expression in each temperature conditions. The second two contrast group (AzadC 4ºC Vs AzadC 22ºC and Ctrl 4ºC Vs Ctrl 22ºC) are set to detect the number of genes that are affected by temperature shift from 22ºC to 4ºC in both Ctrl and AzadC plants. While this is informative, two additional comparisons (AzadC 22ºC Vs Ctrl 4ºC and Ctrl 22ºC Vs AzadC 4ºC) are relevant to identify the genes affected by both temperature and DNA methylation changes.

In each contrast group, multiple statistics were used to determine DE and DAS. To determine DE, log2fold change (L_2FC) was used, which is the difference of difference of log2-CPM values in contrast groups. Further, Δ percent spliced (ΔPS), defined as the difference of PS values (the ratios of transcript average abundances divided by the average gene abundances), was used for DAS analysis. Finally, p-values for multiple testing comparisons are adjusted by the Benjamini-Hochberg to control the false discovery rate (FDR) (Benjamini and Yekutieli, 2001).

 A gene was considered significantly DE if the L_2FC of CPM for each contrast group was ≥ 1 and P values < 0.01. To detect DAS genes, the L_2FC of each transcript was compared to the weighted average of log2 fold changes of all transcripts of the gene, which is a proxy of gene level changes. An F-test was carried out to test if the changes for all the transcripts and the gene are the same. A gene was classified as significantly DAS if $p$value < 0.01 and if the difference in the relative abundance of an alternative splice isoform in relation to the total gene expression within a contrast group (ΔPS) ≥ 0.1.

A gene was then classified as DE only gene if the gene and transcript expression levels change significantly but to the same degree such that transcripts do not differ from one another. Further, a gene was classified as DAS only if its expression level does not change significantly but that of at least one transcript does. Finally, a gene is DE and DAS if both gene level expression changes and different changes of at least one transcript.

 Tor run the DE and DAS analysis using the criteria mentioned above, Limma version 3.40.6 with its voom method has been used (Law et al., 2014). Limma voom has proven its fidelity towards RNA-Seq data analysis through stringent control of FDR in addition to simultaneous

analysis of DE and DAS. Additionally, limma uses linear model instead of bootstrapping, which decreases the time and the memory required for the analysis. More importantly, limma allows multiple comparisons where multiple contrast group can be set for the experimental design (Pimentel et al., 2017; Rapaport et al., 2013; Tang et al., 2015).

3.3.2.4 Identification of percent spliced-in and differential splicing of local events

SUPPA version 2 was used to identify PSI of each AS event followed by detecting differential expression of local AS events in each contrast group (same contrast groups used to detect DE and DAS) (Trincado., J. *et al.* 2018) . The relative abundances of the alternative splicing event or splice isoforms detected in RNA-seq data presented in the form of percentage or proportion are termed percent spliced-in (PSI). Whereas, differential splicing of local AS events between conditions is the difference of transcript relative abundances (or the relative abundance of AS event) and is detonated ∆PSI.

The first step in SUPPA is indexing, which consists of generating AS events from an input annotation file (AtRTDv2, GTF format) to output an ioe file, defining the transcripts from the annotation file that define that a particular splicing event. The different local events generated by SUPPA are Skipping Exon (SE), Alternative 5'/3' Splice Sites (A5/A3), Mutually Exclusive Exons (MX), Retained Intron (RI), and Alternative First/Last Exons (AF/AL) (Trincado., J. *et al.* 2018).

A typical ioe file usually contains the following columns

1. **Seqname**: The chromosome name, in this study the seqname are: Chr1, Chr2, Chr3, Chr4, and Chr5 (Corresponds to Arabidopsis five chromosomes)

2. **Gene_id**: Name of the gene as described in the GTF file, in which the event takes file from the GTF file.

3. **event_id**: Name of the event, displayed as **gene_id; transcript_id**

4. **transcript_id**: ID of the transcript that defines the alternative splicing event, for which the relative inclusion (PSI) is calculated.

5. **Total transcripts**: IDs of the all transcripts transcribed from the gene and which are usually detected from the GTF annotation file.

Then, to detect PSI value for each AS event, SUPPA uses the transcript expression files (Salmon quantification files with TPM values for each transcript) and the ioe file generated in

the previous step transcripts. The start (s) and end (e) coordinates for exons involved in the different AS event are given by SUPPA. The external coordinates of the event are only used for the RI, AF and AL events. For more information about SUPPA nomenclature for different AS (as shown in tables 3.3 and 3.4), the reader can refer to SUPPA version 2 manual (Trincado., J. *et al.* 2018)

To calculate ΔPSI for each event between two conditions, SUPPA was used with three biological replicated for each event to infer a statistical significant for the detected change. For that, the statistical significance is calculated by comparing the observed ΔPSI between conditions with the distribution of the ΔPSI between replicates as a function of the expression of the transcripts defining the events (for events) or as a function of the gene expression (for transcripts).

3.4 Results

3.4.1 Changes in DNA methylation can modulate gene expression and alternative splicing

To study the role of DNA methylation in regulating gene expression and AS changes in response to cold temperature, deep Illumina RNA-seq was performed on AzadC and Ctrl Arabidopsis rosettes (3 biological replicates) before and after their shift from 22°C to 4°C for 24 hours. Principal component analysis of the gene and transcript-level expression data across samples showed that temperature (71.3% and 59% of total variance, respectively) and 5-aza-dC treatment (10.6% and 9% of total variance, respectively) are the major contributors to gene expression variation (Figure 3.2). Based on the filtering criteria described in section 3.3.2.3, a total of 912 and 646 DE and DAS genes has been identified respectively showing change in at least one contrast group, of which 883 and 617 are uniquely DE and DAS respectively (Table 3.2).

Table 3.2. Number of genes and transcripts from results of the analysis of differentially expressed (DE), differentially alternatively spliced (DAS). C1 : Contrast group 1 : AzadC 22°C vs Ctrl 22°C, C2 : Contrast group 2 : AzadC 4°C vs Ctrl 4°C, C3 : Contrast group 3 : AzadC 4°C vs AzadC 4°C, C4 : Contrast group 4 : Ctrl 4°C vs Ctrl 22°C , C5 : Contrast group 5 : AzadC 22°C vs Ctrl 4°C, C6 : Contrast group 6 : Ctrl 22°C vs AzadC 4°C.

	Gene	Transcript
AtRTD2 (Zhang et al., 2017)	34,212	82,190
Expressed	18,362	40,494

No or low expression	15,850	41,696
DE and/or DAS	1558	NA
DE	912	NA
DAS	646	NA
DE and DAS	29	NA
DE-only	883	NA
DAS-only	617	NA
DE-C1-only	42	NA
DE-C2-only	19	NA
DE-C3-only	179	NA
DE-C4-only	324	NA
DE-C5-only	199	NA
DE-C6-only	149	NA
DAS-C1-only	7	NA
DAS-C2-only	16	NA
DAS-C3-only	117	NA
DAS-C4-only	199	NA
DAS-C5-only	186	NA
DAS-C6-only	121	NA

At the gene level, RNA-seq data analysis show that differences in DNA methylation between AzadC and Ctrl change the expression of 656 and 835 genes at 22°C and 4°C, respectively (Figure 3.4 A and B). Further, 6377 and 7020 genes whose expression is affected by temperature shift from 22°C to 4°C in both AzadC and Ctrl plants, respectively (Figure 3.4 C and D). Interestingly, although differences in DNA methylation between AzadC and Ctrl

changed the expression of fewer genes under the same temperature conditions (Figure 3.4 A and B), RNA-seq results show that 6533 cold-responsive genes are regulated through DNA methylation changes (Figure 3.4 F). DNA hypomethylation also induced changes in the expression of 6745 genes under normal growth conditions (Figure 3.4 E). While cold stress induced genome-wide down-regulation of genes expression in both AzadC and Ctrl plants; it is clear that the proportion of down-regulated genes were less in AzadC plants compared to Ctrl (52% and 60% of DE genes in group C and D respectively of Figure 3.4); which is due to the relaxed chromatin structure of AzadC plants allowing more gene expression. This is further confirmed by the high proportion of up-regulated genes compared to the down-regulated ones detected in response to cold stress and are regulated by DNA methylation changes (84%, 56%, and 57% of DE genes in group B, E, and F respectively of Figure 3.4).

At the AS level, a gene was considered DAS in a contrast group if the Benjamini-Hochberg (BH) adjusted p-value < 0.01 and at least one of the transcripts had ΔPS (percent spliced) $\geq$ 0.1. Although differences in DNA methylation between AzadC and Ctrl induced few changes in AS at 22ºC (87 genes) and 4ºC (259 genes), a tremendous number of genes change their AS profiles upon cold stress and are regulated by changes in DNA methylation (2481 and 2027 genes upon shift to 22ºC and 4ºC, respectively). Interestingly, DNA hypomethylation also decreases DAS genes in plants subject to cold stress compared to Ctrl plants (Figure 3.4 C and D) (Supplementary table 1-3).
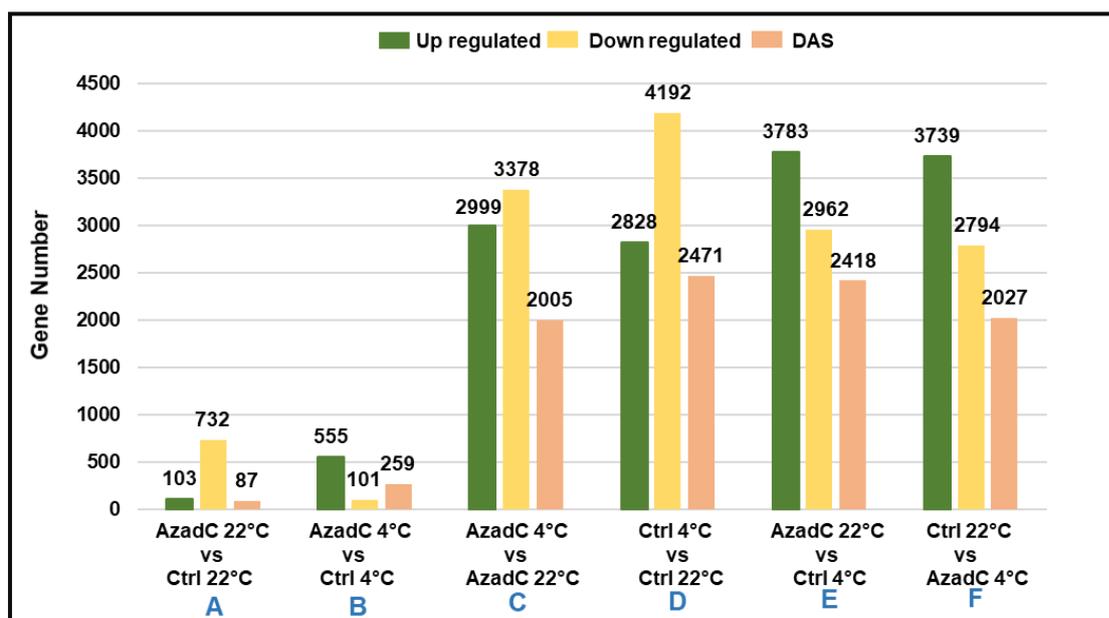
Figure 3.4. Chart displaying the number of differentially expressed (DE) genes (Up- and Down-regulated) and differentially alternatively spliced (DAS) genes in different contrast groups (A-F). The number of up- and down-regulated genes in each contrast group (A-F) represent the total number of DE genes. The x- and y-axis represent

the contrast groups and gene number, respectively. AzadC and Ctrl are Arabidopsis plants treated and untreated with 5-Aza-dC, respectively. Contrast group A and B display the lowest (835 and 656, respectively) number of DE genes compared to contrast groups C (6377), D (7020), E (6745) and F (6533). At the gene level, changes in DNA methylation are more likely to induce genome-wide changes in gene expression upon temperature shift (contrast group E and F) rather than under constant temperature conditions (contrast group A and B). Although the effect of temperature was sufficient to up-regulate genes in AzadC and Ctrl plants (2999 and 2828 in contrast groups C and D, respectively), DNA hypomethylation clearly increased the number of up-regulated genes (3783 and 3739 in contrast groups E and F, respectively). Similarly, at the splicing level, DAS gene number increased from 87 and 259 in the first two contrast groups respectively, to 2418 and 2027 genes in the last two contrast groups. Temperature shifts alone in AzadC and Ctrl plants (contrast groups C and D) were sufficient to increase DAS gene number when compared to group A and B.

In each contrast group, most transcriptional changes are observed for genes that do not display splicing changes. Similarly, most splicing changes occur in genes that are not DE (DE and DAS gene sets are largely different with an overlap of only 0.8-11% (Figure 3.5 A-F).
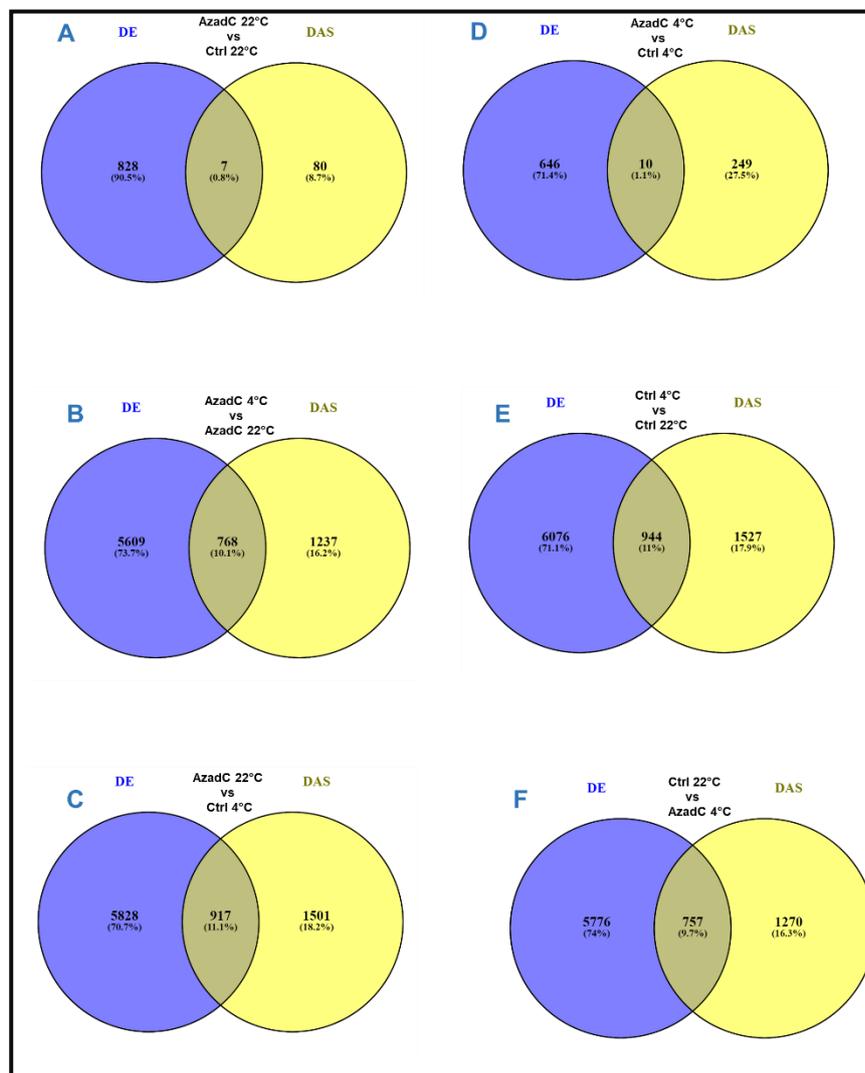
Figure 3.5. Venn diagram of DE and DAS genes. Venn diagram of DE (blue) DAS (yellow) and DE and DAS (intersection) genes in each contrast groups. DE and DAS are differentially expressed and alternatively spliced, respectively. **A and B** show that DE (in A) and DAS (in B) detected in different contrast groups are likely to be either uniquely DE (A) or DAS (B) for each contrast group or common with one or more contrast groups. Both Venn diagrams of DE (A) and DAS (B) show that only 25 and 3 genes are common in DE and DAS genes,

respectively among all contrast groups. Considering the overlap of DE and DAS in each contrast group (C-F), the Venn diagrams show a very low number of genes which are regulated at both the gene and transcript level for all contrast group and that the majority of DE and DAS genes are either regulated at the gene or transcript level.

It has been previously reported that DNA methylation and chromatin features are likely to regulate gene expression and splicing *via* differential recruitment of transcription factors (TFs) and SFs (Goodrich and Tjian, 2010; Reddy *et al.*, 2013; Jabre *et al.*, 2019). In this study, the list of predicted Arabidopsis 798 SF-RBPs and 2534 TFs has been used (Calixto *et al.*, 2018) To investigate this relationship and to interrogate the overlap between DE and DAS of genes encoding splicing factors/RNA-binding proteins (SF-RBPs) and TFs among different contrast groups. Interestingly, large variation between DE and DAS genes belonging to TFs and SF-RBPs between AzadC and Ctrl has been found in the different contrast groups, indicating that changes in DNA methylation may be associated with differential expression of SFs and TFs to regulate genome-wide changes in gene expression at the gene and transcript level (Figure 3.6 and supplementary table 4).
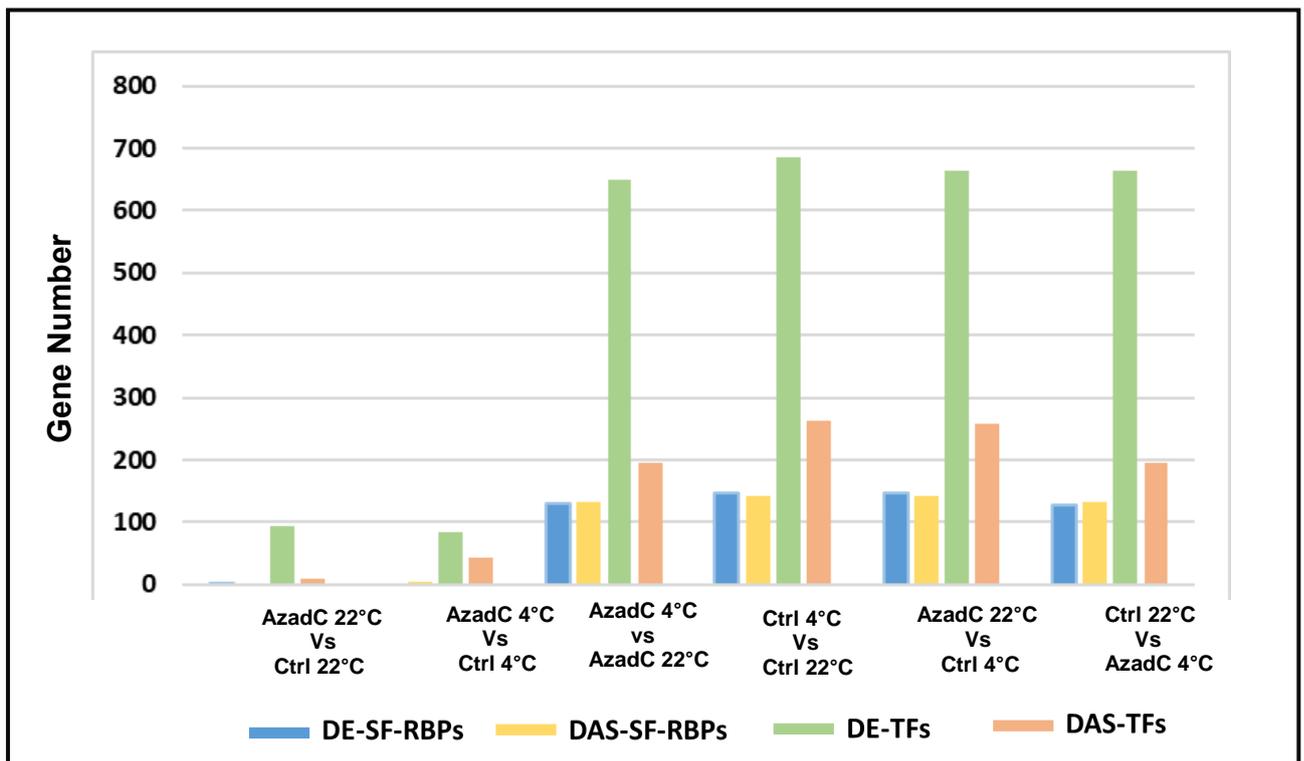
Figure 3.6. Chart displaying the number of differentially expressed genes (DE) and differentially alternatively spliced (DAS) genes in four contrast groups that belong to Arabidopsis transcription factors (TFs) and splicing-related genes (*SRs*). Ctrl and AzadC are Arabidopsis plants without and with 5-AzadC treatment respectively.

Gene ontology (GO) analysis of DE genes from the different contrast groups, including those which are regulated by changes in DNA methylation, showed enrichment in diverse biological processes (circadian rhythm, response to cold, photosynthesis, protein phosphorylation, and hormone-activating signalling pathways), cellular components (plasma membrane, cell

membrane, cell wall, and vacuole), and molecular function (Kinase activity, protein/threonine kinase activity, protein binding, and sequence specific DNA binding activity) (Figure 3.7).

83
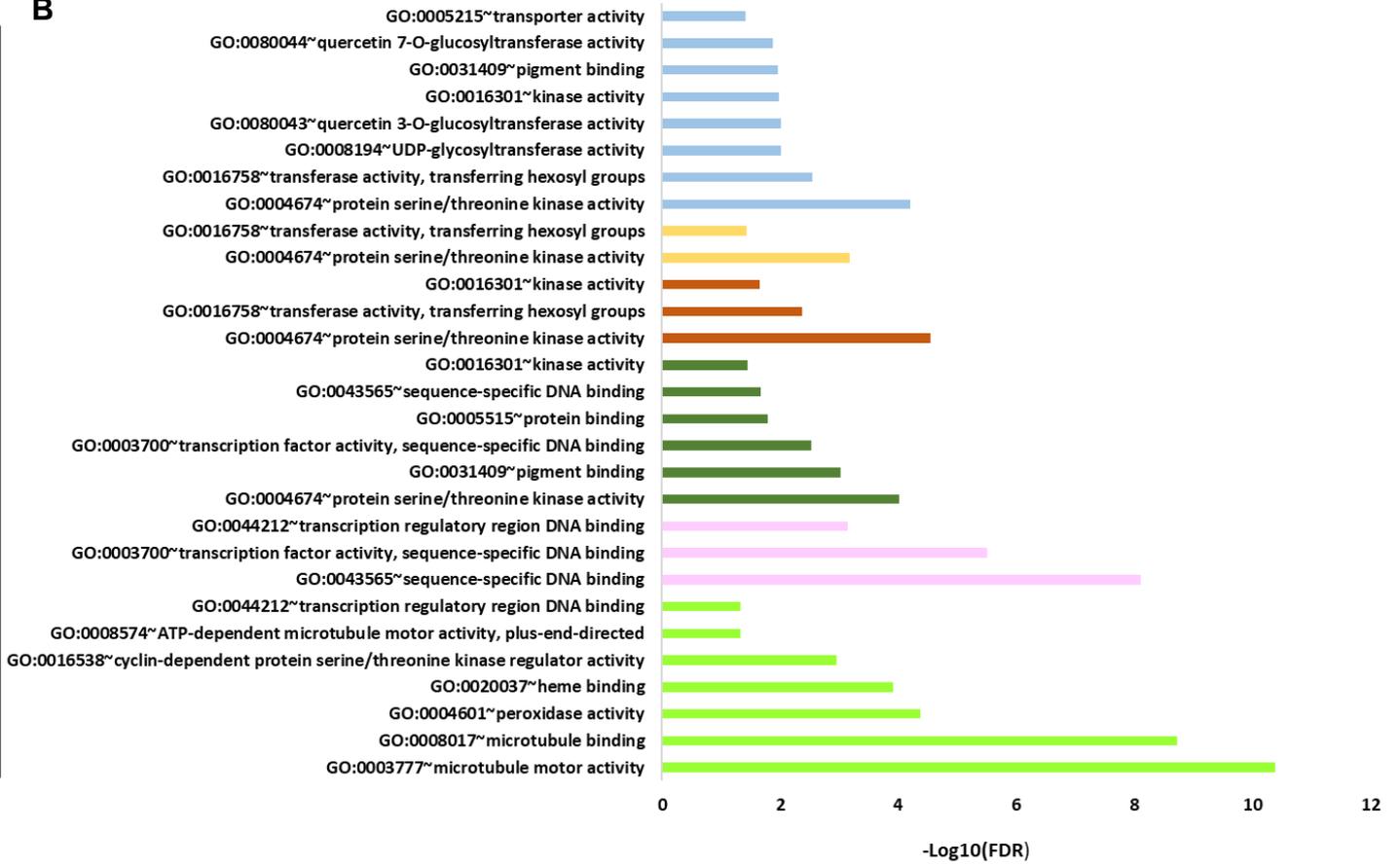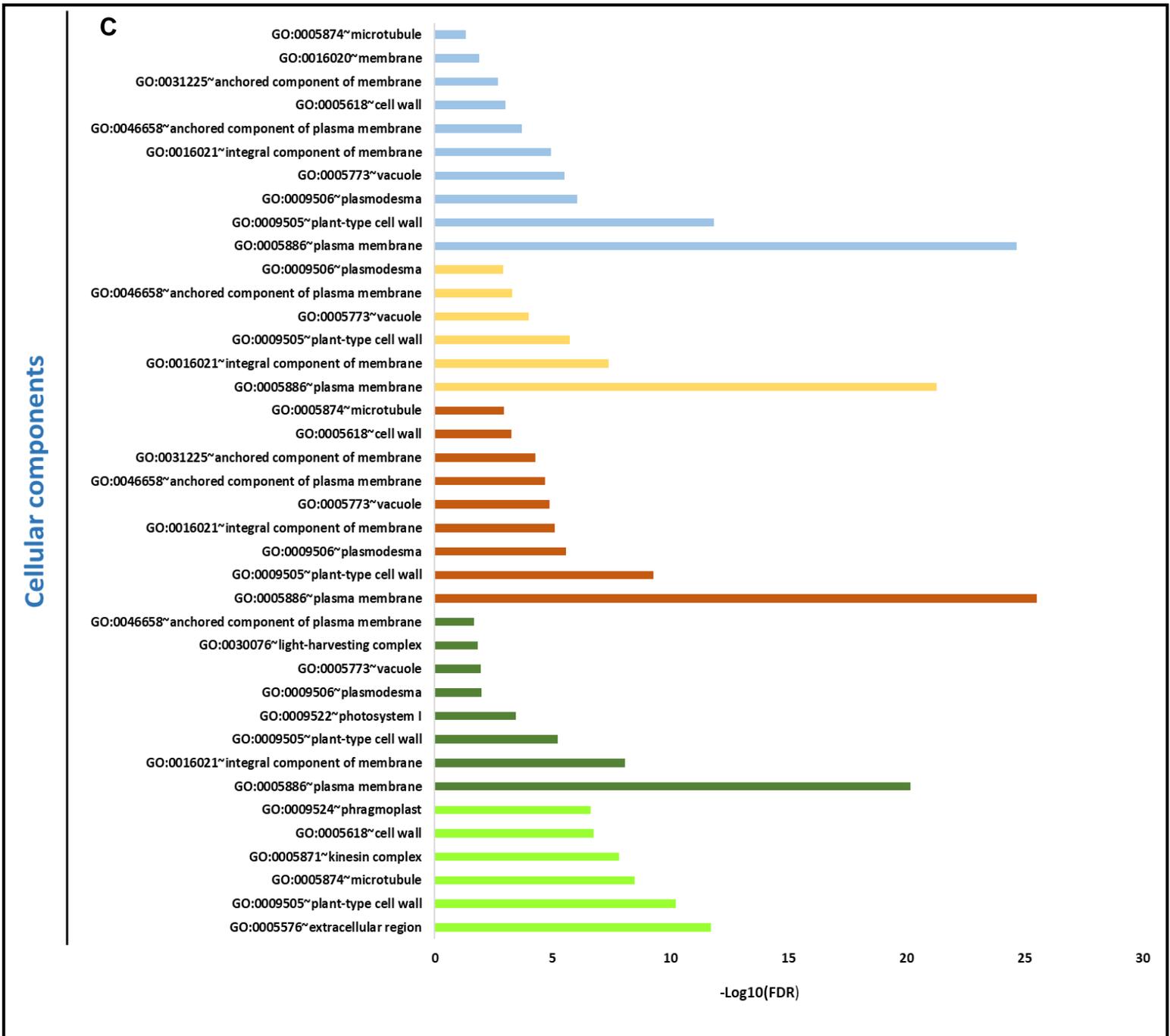
Figure 3.7. Significant (FDR < 0.05) GO term enrichment analysis of differentially expressed genes. The x-axis represents the –log10 FDR value for the GO term; the y-axis represents biological processes, molecular functions and cellular components. No significant GO terms for cellular components were detected for contrast group 2.
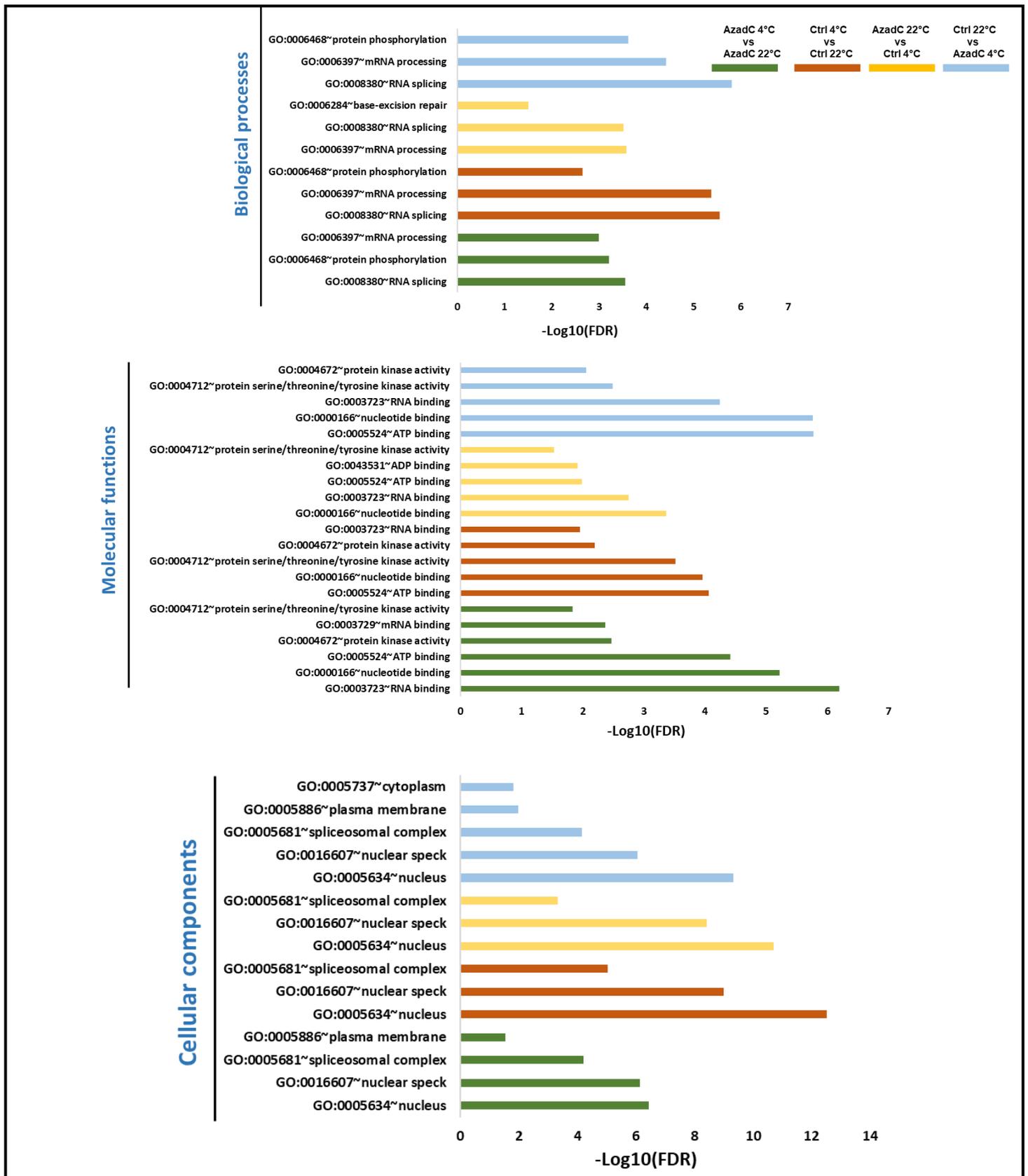
Figure 3.8. Significant (FDR < 0.05) GO term enrichment analysis of differentially alternatively spliced genes.
The x-axis represents the –log10 FDR value for the GO term; the y-axis represents biological processes, molecular functions and cellular components. For group 1 and 2, no significant GO terms were detected for DAS genes, which is due to the low number of DAS genes detected in these groups.

Overall, these results supports the hypothesis that changes in DNA methylation regulate gene expression and AS patterns in response to low temperature of genes, which are involved in regulating cold acclimation as well as multiple physiological processes.

3.4.2 Local splicing events are orchestrated by DNA methylation changes

Since DNA methylation differences induced changes in AS transcript levels of hundreds of genes under normal and cold conditions, the next step is to investigate how plants with DNA methylation differences (AzadC and Ctrl plants) regulate local AS events under both temperature conditions. Towards this goal, local AS events PSI values for a total of 43953 AS events (Table 3.3) and the difference of their distribution (ΔPSI, table 3.4) have been identified by SUPPA as described in section 3.3.2.4 (Supplementary table 5 and the remaining data of table 3.3 and 3.4 can be found in supplementary table 6 and 7, respectively). Comparing AzadC to Ctrl at 22°C and 4°C 267 and 502 significant (pvalue ≤ 0.05) differential AS events have been identified, respectively. Furthermore, AzadC and Ctrl plants shifted from normal to cold stress display 2138 and 2660 significant differential AS events, respectively. Interestingly, the last two contrast groups (AzadC 22°C Vs Ctrl 4°C and Ctrl 22°C Vs AzadC 4°C) display 2726 and 2122, respectively. In line with DAS analysis, the first two contrast groups show the lowest number of DAS events hence, emphasising that changes in DNA methylation under the same temperature conditions are less likely to induce global changes in local AS events. The differential AS events results obtained from AzadC 22°C Vs Ctrl 4°C and Ctrl 22°C Vs AzadC 4°C also support our DAS analysis where AzadC plants showed less differential AS events compared to Ctrl plants after shifting from normal to cold stress. Interestingly, the highest number of differential AS events are detected in plants displaying both temperature shift and changes in DNA methylation.

Table 3.3. The most significant (Top 50) AS events identified by SUPPA in each sample alongside their PSI values. In column 1, AS events are presented as follows: gene_id, event-type, seqname, coordinates-of-the-event, strand: either '+' or '-'.

AS event	PSI-Values			
	AzadC 22°C	AzadC 4°C	Ctrl 22°C	Ctrl 4°C
AT1G18382 IR Chr1 6326366 6326600-6326676 6326699 -	0.833333424	0.999999999	0.999999998	0.955427701
AT1G69252 ES Chr1 26036623-26036798 26037276-26037400 +	0.835323943	0.684158525	0.999999304	0.834274594
AT1G01725 A3'SS Chr1 270090-270622 270085-270622 -	0.999998815	0.997257096	0.99999775	0.999997692
AT2G48121 A3'SS Chr2 19681929-19682026 19681929-19682160 +	0.99903715	0.83136062	0.999997353	0.997137323
AT2G36170 A5'SS Chr2 15172626-15172862 15172592-15172862 +	0.997978548	0.997351056	0.999969662	0.996640709
AT1G29465 IR Chr1 10308574 10308804-10309575 10309798 +	0.999947071	0.999905471	0.999964935	0.999893378
AT2G24040 A3'SS Chr2 10224117-10224394 10224117-10224748 +	0.999983305	0.999966215	0.999957288	0.999937731
AT3G01500 IR Chr3 194746 194891-195179 195337 -	0.999926626	0.999844563	0.999955193	0.999889186
AT5G03240 A3'SS Chr5 772896-773272 772286-773272 -	0.999791944	0.99998637	0.999953293	0.999856004
AT4G21730 IR Chr4 11544933 11545093-11545231 11545527 +	0.98679568	0.896454301	0.999951098	0.945502348
AT4G16695 A5'SS Chr4 9394264-9394421 9394260-9394421 +	0.999934307	0.999897071	0.999919784	0.999897066
AT4G04830 A5'SS Chr4 2445999-2446452 2445969-2446452 +	0.999870108	0.999463474	0.999917745	0.999453468
AT1G06400 A3'SS Chr1 1951518-1952466 1951297-1952466 -	0.998156746	0.997443456	0.999907738	0.999447388
AT1G08830 A3'SS Chr1 2827191-2827680 2827191-2827689 +	0.999597354	0.999739456	0.999899783	0.999569315
AT3G56940 EI Chr3 21076505 21076631-21076709 21077067 +	0.999689019	0.99976004	0.999893664	0.999402806
AT3G47833 A5'SS Chr3 17648774-17648862 17648558-17648862 +	0.999893011	0.999871073	0.999892373	0.999892798

AT3G46385 A3'SS Chr3 17065795-17065996 17065489-17065996 -	0.966583992	0.999805747	0.999891984	0.999770703
AT1G48030 IR Chr1 17717421 17717482-17718581 17718682 -	0.999861116	0.999764584	0.999887299	0.99982421
AT4G24440 A5'SS Chr4 12633588-12633688 12633538-12633688 +	0.998195453	0.998125537	0.999876347	0.99772623
AT1G56200 A5'SS Chr1 21030909-21031092 21030900-21031092 +	0.999600144	0.99984088	0.999864105	0.999858807
AT5G35680 IR Chr5 13857921 13858150-13858203 13858594 -	0.999427028	0.999921005	0.999864034	0.998572012
AT5G13930 EI Chr5 4489041 4489150-4489821 4490264 +	0.999812359	0.999974368	0.999859895	0.999869222
AT5G21920 A5'SS Chr5 7242082-7242460 7242082-7242499 -	0.999826367	0.999447927	0.999820333	0.999538215
AT1G29465 A5'SS Chr1 10308815-10309575 10308804-10309575 +	0.999738204	0.923631097	0.999804925	0.946149318
AT5G02500 EI Chr5 553745 554228-555603 555796 -	0.999939124	0.999908324	0.999803331	0.999808656
AT1G55990 IR Chr1 20942487 20942614-20942660 20942951 -	0.999703559	0.999908708	0.999781716	0.999775022
AT3G60245 A3'SS Chr3 22269590-22269687 22269590-22269696 +	0.999770361	0.999886703	0.999781139	0.999993159
AT1G07920 EI Chr1 2456114 2456148-2456645 2457321 +	0.999836773	0.999798818	0.999780393	0.999863477
AT1G67090 A5'SS Chr1 25048701-25048838 25048701-25048845 -	0.999743777	0.999821613	0.999778707	0.999773589
AT1G21065 A3'SS Chr1 7375001-7375166 7375001-7375184 +	0.998806595	0.999131143	0.999775532	0.999109801
AT3G52590 AF Chr3 19505255 19505285-19505635 19505511 19505550-19505635 +	0.999613637	0.999729756	0.999772309	0.999603698
AT1G77350 A3'SS Chr1 29070875-29071270 29070875-29071282 +	0.999794033	0.998338339	0.999771457	0.99724452
AT1G07940 IR Chr1 2462953 2463706-2464203 2464237 -	0.999785884	0.999685751	0.999770709	0.999851817
AT2G27030 A3'SS Chr2 11533056-11534077 11533056-11534083 +	0.989454154	0.999816954	0.999768457	0.990109075
AT5G60390 EI Chr5 24289788 24290487-24290570 24291020 +	0.999698964	0.999654992	0.999741908	0.999624753

AS event				
AT3G54050 A5'SS Chr3 20017397-20017523 20017120-20017523 +	0.999778846	0.999672516	0.999739452	0.999678151
AT3G54890 IR Chr3 20339504 20339848-20339940 20340228 -	0.99979931	0.999694526	0.999734251	0.999681905
AT3G46040 AF Chr3 16914525 16914640-16914873 16914760 16914775-16914873 +	0.99912846	0.999569192	0.999729553	0.999063433
AT1G20693 A3'SS Chr1 7178376-7178470 7178376-7178473 +	0.999368547	0.999217958	0.999711319	0.999187193
AT2G01250 A3'SS Chr2 133158-133258 133079-133258 -	0.999582117	0.999603699	0.999690873	0.999555569
AT4G05320 EI Chr4 2718559 2718911-2719596 2720308 +	0.999660076	0.999662423	0.99968067	0.999643984
AT2G43780 A5'SS Chr2 18136665-18137410 18136665-18137571 -	0.99975515	0.999781753	0.999674827	0.998091134
AT2G27775 A3'SS Chr2 11842649-11842887 11842645-11842887 -	0.995666979	0.998365846	0.999666837	0.998013524
AT2G39730 IR Chr2 16572500 16572569-16572694 16572816 -	0.999740027	0.999701322	0.99966527	0.999747733
AT5G02450 AF Chr5 533058 533195-533496 533234 533290-533496 +	0.999936855	0.999948826	0.999658535	0.999799502
AT5G45550 A3'SS Chr5 18463165-18463314 18463165-18463323 +	0.996754518	0.998190006	0.999656001	0.997666999
AT4G00430 IR Chr4 186602 186671-186873 186897 -	0.999210421	0.999380181	0.999655223	0.99959522
AT2G31141 IR Chr2 13271726 13271958-13272104 13272145 -	0.999957012	0.999948751	0.999650585	0.999895714

Table 3.4. The most significant (Top 50) differential AS events identified by SUPPA in each contrast group. In column 1, AS events are presented as follows: gene_id, event-type, seqname, coordinates-of-the-event, strand: either '+' or '-'.

<p align="center"><u>Contrast group 1: AzadC 22°C vs Ctrl 22°C</u></p>

AS event	ΔPSI value	P value ≤ 0.05
AT1G25175 A3'SS Chr1 8827623-8828298 8827623-8829112 +	0.382074756	0
AT1G72930 A5'SS Chr1 27439973-27440072 27439930-27440072 +	0.190010647	0
AT2G24600 IR Chr2 10452421 10453359-10453463 10453902 -	0.57320443	0
AT2G32690 A3'SS Chr2 13863989-13864311 13863971-13864311 -	0.333586629	0
AT2G32690 A3'SS Chr2 13864010-13864311 13863971-13864311 -	0.35173249	0
AT2G32690 A5'SS Chr2 13864010-13864311 13864010-13864323 -	0.196687471	0
AT2G39730 A5'SS Chr2 16571066-16571180 16571066-16571191 -	-0.01818078	0

AT2G41100 A3'SS Chr2 17137411-17137521 17137411-17137524 +	0.28920784	0
AT2G41100 IR Chr2 17138107 17138206-17138307 17138573 +	0.328498275	0
AT2G47110 IR Chr2 19344635 19345236-19345289 19345345 +	-0.131454388	0.000499501
AT4G30660 A5'SS Chr4 14956063-14956161 14956000-14956161 +	0.070163446	0.000999001
AT5G24735 IR Chr5 8468331 8468672-8469020 8469262 -	0.12039773	0.000999001
AT1G25175 A5'SS Chr1 8828544-8829112 8827623-8829112 +	0.237118268	0.001498502
AT1G25175 IR Chr1 8828298 8828544-8829112 8829320 +	-0.282929242	0.001498502
AT4G08991 IR Chr4 5770930 5771090-5771165 5771335 -	-0.42165303	0.001498502
AT5G45340 IR Chr5 18368532 18369070-18369164 18369389 -	0.405884618	0.001498502
AT1G32940 AF Chr1 11937183 11937402-11937800 11937499 11937717-11937800 +	-0.37756312	0.001998002
AT3G45970 IR Chr3 16896166 16896358-16896435 16896729 +	0.121828846	0.001998002
AT4G33100 A5'SS Chr4 15970955-15971105 15970912-15971105 +	-0.326597479	0.001998002
AT5G03120 IR Chr5 734224 734379-734679 735381 +	0.121241561	0.001998002
AT2G32690 A3'SS Chr2 13863971-13864311 13863962-13864311 -	-0.3475672	0.002497503
AT2G32690 A3'SS Chr2 13864010-13864311 13863962-13864311 -	-0.246065163	0.002497503
AT2G32690 A3'SS Chr2 13864025-13864311 13863962-13864311 -	-0.429987544	0.002497503
AT2G32690 A3'SS Chr2 13864025-13864311 13863971-13864311 -	0.349454718	0.002497503
AT2G32690 A3'SS Chr2 13864028-13864311 13863971-13864311 -	0.23218047	0.002497503
AT4G32060 A5'SS Chr4 15502446-15502512 15502301-15502512 +	-0.176350576	0.002497503
AT4G32060 IR Chr4 15502218 15502301-15502512 15502725 +	-0.107408162	0.002497503
AT5G44572 A3'SS Chr5 17969433-17970071 17969433-17970079 +	0.399251569	0.002497503
AT1G25175 A3'SS Chr1 8827623-8828374 8827623-8829112 +	0.449673727	0.002697303
AT1G25175 IR Chr1 8827570 8827623-8828298 8828544 +	-0.327968017	0.002697303
AT1G25175 IR Chr1 8827286 8827335-8827428 8827462 +	-0.189674144	0.002997003
AT2G35050 A3'SS Chr2 14768872-14769524 14768872-14772270 +	-0.087949731	0.002997003
AT2G43530 IR Chr2 18070056 18070170-18070333 18070718 +	-0.078539688	0.002997003
AT4G08991 A3'SS Chr4 5771090-5771165 5771056-5771165 -	0.491189659	0.002997003
AT1G25175 ES Chr1 8827623-8828298 8828544-8829112 +	0.617515438	0.003211075
AT1G61340 IR Chr1 22628516 22628792-22629360 22629434 +	0.313187861	0.003496504
AT2G41110 A5'SS Chr2 17140490-17140819 17140454-17140819 +	0.089305479	0.003496504
AT1G02080 A3'SS Chr1 374922-375122 374922-375125 +	0.124596579	0.003996004
AT1G13650 IR Chr1 4681883 4681946-4682066 4682382 -	0.095481225	0.003996004
AT2G29340 A5'SS Chr2 12598311-12598869 12598306-12598869 +	0.23879983	0.003996004
AT2G29340 IR Chr2 12597910 12598016-12598115 12598306 +	0.26763308	0.003996004
AT3G61610 IR Chr3 22799027 22799040-22799499 22799590 +	0.276227232	0.003996004
AT4G32060 IR Chr4 15502218 15502446-15502512 15502725 +	0.092575885	0.004162504
AT1G19720 IR Chr1 6819564 6819576-6819716 6819818 -	0.121108612	0.004495505
AT4G13495 A5'SS Chr4 7843302-7843376 7843248-7843376 +	-0.035762281	0.004495505

AS event	ΔPSI value	P value
AT4G26530 A3'SS Chr4 13391394-13391486 13391394-13391492 +	-0.05765541	0.004495505
AT5G19250 A3'SS Chr5 6472064-6472300 6472064-6472324 +	-0.023856829	0.004495505
AT5G22380 IR Chr5 7409277 7409563-7409879 7410102 -	0.242691731	0.004495505
AT4G33150 A3'SS Chr4 15991093-15991446 15991069-15991446 -	0.137005968	0.004995005

Contrast group 2: AzadC 4°C vs Ctrl 4°C

AS event	ΔPSI value	P value ≤ 0.05
AT1G20696 A5'SS Chr1 7181073-7181160 7181069-7181160 +	0.313474862	0.04995005
AT1G32860 A3'SS Chr1 11907837-11907944 11907727-11907944 -	-0.290585125	0.04995005
AT1G34418 A3'SS Chr1 12582384-12582469 12582384-12582507 +	-0.130111321	0.04995005
AT1G77260 IR Chr1 29024427 29024549-29024627 29024884 -	0.08338321	0.04995005
AT1G78070 ES Chr1 29355825-29356300 29356466-29356665 +	0.035397032	0.04995005
AT2G02470 IR Chr2 654136 654262-654337 654985 +	0.088039968	0.04995005
AT2G02470 IR Chr2 654136 654262-654364 654985 +	0.132841868	0.04995005
AT3G11820 IR Chr3 3729305 3730006-3730295 3730848 -	-0.298424297	0.04995005
AT3G23830 A3'SS Chr3 8607699-8607982 8607687-8607982 -	0.060807676	0.04995005
AT3G26920 IR Chr3 9920975 9921840-9921917 9922066 +	0.269555946	0.04995005
AT4G10170 IR Chr4 6344301 6344452-6344587 6345610 +	-0.104456168	0.04995005
AT4G33467 A3'SS Chr4 16101840-16101926 16101837-16101926 -	-0.013019426	0.04995005
AT5G11010 IR Chr5 3483632 3483863-3484154 3484342 +	0.234102026	0.04995005
AT5G54930 IR Chr5 22305733 22305935-22306027 22306885 -	0.397656265	0.04995005
AT1G75420 EI Chr1 28306039 28306149-28306247 28306310 +	0.084262912	0.04945055
AT1G78070 A5'SS Chr1 29356708-29357085 29355825-29357085 +	0.015642561	0.04945055
AT3G19820 AF Chr3 6881628-6882096 6882121 6881628-6882212 6882315 -	0.130114433	0.04945055
AT4G21865 A3'SS Chr4 11602292-11602566 11602292-11602646 +	-0.300781553	0.04945055
AT5G38470 A5'SS Chr5 15407105-15407195 15407101-15407195 +	0.0277002	0.04945055
AT3G02470 IR Chr3 509305 509399-509797 509952 +	-0.213227878	0.048951049
AT3G62800 A3'SS Chr3 23227174-23227447 23227169-23227447 -	-0.111347431	0.048951049
AT4G37180 A3'SS Chr4 17505532-17505643 17505532-17505647 +	0.098315659	0.048951049
AT3G52920 A3'SS Chr3 19625738-19625859 19625738-19625868 +	0.024744596	0.048451549
AT1G02090 IR Chr1 387584 387672-388268 388406 -	0.073004562	0.047952048
AT1G14820 A5'SS Chr1 5105924-5106163 5105924-5106722 -	0.093514913	0.047952048

AT1G55450 A5'SS Chr1 20706234-20706676 20706234-20706678 -	0.072638671	0.047952048
AT1G62430 IR Chr1 23108264 23108415-23108517 23108770 -	0.071568572	0.047952048
AT2G21660 EI Chr2 9265249 9265484-9265575 9265622 -	-0.322218662	0.047952048
AT4G00040 IR Chr4 14627 14833-14921 16079 +	0.061151068	0.047952048
AT4G25640 IR Chr4 13077109 13077195-13077270 13077388 -	0.008151806	0.047952048
AT5G19855 AL Chr5 6711943 6712279-6712390 6712282 6712291-6712390 -	0.166057244	0.047952048
AT5G19855 IR Chr5 6712390 6712455-6712795 6712852 -	0.03503437	0.047952048
AT2G22710 IR Chr2 9651719 9652735-9653132 9653207 +	-0.258762508	0.047452548
AT2G45990 IR Chr2 18920943 18921132-18921262 18921511 +	-0.045087565	0.047452548
AT3G05600 IR Chr3 1623244 1623752-1623820 1624070 -	0.272761739	0.047452548
AT3G61750 IR Chr3 22858571 22858970-22859048 22859669 -	0.090735527	0.047452548
AT4G01590 A3'SS Chr4 689096-689185 689093-689185 -	0.049624233	0.047452548
AT4G28150 IR Chr4 13978102 13978187-13978272 13978452 -	0.044453879	0.047452548
AT4G36730 A3'SS Chr4 17310679-17310758 17310673-17310758 -	-0.021407625	0.047452548
AT4G36730 IR Chr4 17311693 17311790-17312207 17312479 -	-0.182272957	0.047452548
AT5G14440 A3'SS Chr5 4656176-4656498 4656173-4656498 -	-0.056149109	0.047452548
AT5G53540 IR Chr5 21749904 21750199-21750283 21750441 -	0.047300054	0.047452548
AT1G54170 IR Chr1 20222648 20222775-20222849 20223070 -	0.056455122	0.047202797
AT1G54170 ES Chr1 20223345-20223832 20223904-20224022 -	0.056455122	0.047202797
AT1G28600 A3'SS Chr1 10052167-10052254 10052160-10052254 -	-0.023702516	0.046953047
AT1G71340 IR Chr1 26886535 26886615-26886706 26886770 -	0.068050356	0.046953047
AT4G27960 IR Chr4 13916573 13917240-13917343 13917420 -	-0.267304188	0.046953047
AT4G27960 IR Chr4 13917134 13917240-13917343 13917500 -	-0.023901515	0.046953047
AT5G24530 IR Chr5 8381929 8382253-8382900 8383401 +	0.017622864	0.046953047

Contrast group 3: AzadC 4°C vs AzadC 4°C

AS event	ΔPSI value	P value ≤ 0.05
AT1G01060 IR Chr1 37373 37398-37569 37780 -	-0.554345146	0
AT1G01910 A3'SS Chr1 313452-313567 313418-313567 -	0.292029554	0
AT1G01910 IR Chr1 313145 313418-313567 313759 -	0.339304898	0
AT1G04080 IR Chr1 1053807 1054200-1054663 1054806 +	0.255924813	0
AT1G09920 ES Chr1 3225441-3225994 3226123-3226817 -	0.670724769	0
AT1G10890 A3'SS Chr1 3630064-3630255 3630064-3630279 +	-0.221183716	0
AT1G10890 IR Chr1 3627997 3628229-3629098 3629304 +	0.338016404	0
AT1G10890 IR Chr1 3627997 3628972-3629098 3629304 +	0.253949108	0
AT1G10910 IR Chr1 3643454 3643582-3643732 3644187 +	0.263035085	0
AT1G12650 A3'SS Chr1 4306028-4306149 4306028-4306169 +	-0.204053575	0
AT1G12750 A3'SS Chr1 4347790-4348241 4347537-4348241 -	0.196380229	0
AT1G13350 A5'SS Chr1 4575835-4575962 4575835-4576239 -	0.458879717	0
AT1G14170 EI Chr1 4844654 4845150-4845226 4845283 -	-0.24216955	0
AT1G14820 ES Chr1 5105924-5106163 5106421-5106722 -	0.236915101	0
AT1G15200 A3'SS Chr1 5229231-5229346 5229051-5229346 -	0.318067697	0
AT1G18660 IR Chr1 6420975 6421303-6421398 6421501 +	-0.315929324	0
AT1G19400 IR Chr1 6712981 6713787-6714169 6714483 -	-0.31143253	0
AT1G22140 IR Chr1 7814663 7814937-7815130 7815292 -	-0.285076466	0
AT1G22750 A3'SS Chr1 8052480-8052606 8052480-8052663 +	-0.239704866	0
AT1G22750 A3'SS Chr1 8052480-8052615 8052480-8052663 +	-0.310921025	0
AT1G24825 A3'SS Chr1 8776339-8777014 8776339-8777828 +	0.099583237	0
AT1G24825 ES Chr1 8776339-8777014 8777260-8777828 +	0.366097572	0
AT1G25098 IR Chr1 8813192 8813245-8813920 8814450 +	-0.233358318	0
AT1G25175 A3'SS Chr1 8827623-8828298 8827623-8829112 +	-0.445494798	0
AT1G28060 EI Chr1 9779757 9779825-9779905 9780038 +	-0.238272341	0
AT1G28330 ES Chr1 9934288-9934451 9934566-9934794 -	0.395130525	0
AT1G34340 A5'SS Chr1 12532696-12532780 12532661-12532780 +	-0.36735443	0
AT1G44750 ES Chr1 16892730-16893565 16893831-16894072 +	0.433075271	0
AT1G48030 A3'SS Chr1 17719236-17719377 17719173-17719377 -	0.211743037	0
AT1G48410 A3'SS Chr1 17890649-17890735 17890643-17890735 -	0.196898971	0
AT1G49500 A3'SS Chr1 18320676-18321415 18320671-18321415 -	0.244262569	0
AT1G50440 IR Chr1 18685777 18685966-18686144 18686312 +	-0.510810636	0
AT1G53040 IR Chr1 19764830 19765029-19765120 19765307 -	-0.369205721	0
AT1G53510 IR Chr1 19972658 19972807-19973144 19973203 -	0.19865425	0
AT1G53510 ES Chr1 19972807-19972922 19972964-19973144 -	0.295322339	0
AT1G54380 A3'SS Chr1 20299689-20300122 20299244-20300122 -	0.413631817	0
AT1G55340 A3'SS Chr1 20652822-20652902 20652822-20652932 +	0.242949711	0
AT1G56220 IR Chr1 21043704 21043889-21044357 21044998 +	0.15785152	0
AT1G56660 IR Chr1 21237888 21237949-21238803 21240558 +	-0.471737547	0
AT1G62710 IR Chr1 23225071 23225270-23225363 23225448 -	-0.167248079	0
AT1G65270 A3'SS Chr1 24244536-24244646 24244536-24244795 +	0.226103437	0

AS event		
AT1G65270 IR Chr1 24244376 24244536-24244795 24245074 +	0.262655019	0
AT1G65280 IR Chr1 24245122 24245407-24245489 24245893 +	-0.236643072	0
AT1G66260 IR Chr1 24696767 24696916-24697662 24697910 -	0.286486394	0
AT1G67300 A3'SS Chr1 25194778-25194893 25194740-25194893 -	-0.332509779	0
AT1G68660 IR Chr1 25778407 25778560-25779020 25779161 -	-0.289477583	0
AT1G69610 A3'SS Chr1 26187108-26187193 26187108-26187204 +	0.253179348	0
AT1G72640 A3'SS Chr1 27347498-27347585 27347489-27347585 -	0.198080879	0
AT1G73480 ES Chr1 27630935-27631319 27631365-27631572 +	0.220737636	0
AT1G78070 IR Chr1 29355739 29356466-29356665 29356708 +	-0.326187254	0

Contrast group 4: Ctrl 4°C vs Ctrl 22°C

AS event	ΔPSI value	P value ≤ 0.05
AT1G18180 A3'SS Chr1 6257593-6257695 6257593-6257701 +	0.211797	0.04995
AT1G26630 A3'SS Chr1 9206891-9206972 9206891-9206994 +	-0.00328	0.04995
AT1G31020 A3'SS Chr1 11057901-11057986 11057901-11057990 +	-0.09286	0.04995
AT1G31355 IR Chr1 11228988 11229209-11229523 11229990 +	-0.19564	0.04995
AT1G48030 A3'SS Chr1 17719236-17719377 17719173-17719377 -	0.071197	0.04995
AT1G48030 IR Chr1 17717008 17717320-17717421 17718682 -	-0.07109	0.04995
AT1G51690 A5'SS Chr1 19167568-19167961 19167565-19167961 +	0.12079	0.04995
AT1G53390 A3'SS Chr1 19919398-19919637 19919398-19919641 +	0.118147	0.04995
AT1G53390 A5'SS Chr1 19920296-19920453 19920292-19920453 +	0.133781	0.04995
AT1G53390 ES Chr1 19918998-19919322 19919398-19919637 +	0.144474	0.04995
AT1G58180 IR Chr1 21539005 21539053-21539169 21539300 -	0.091149	0.04995
AT1G58180 IR Chr1 21539169 21539300-21539394 21539458 -	-0.09049	0.04995
AT1G75180 A5'SS Chr1 28216380-28216776 28216380-28217195 -	0.035795	0.04995
AT2G02910 A5'SS Chr2 848044-848125 848044-848130 -	0.190993	0.04995
AT2G20900 IR Chr2 8991429 8991511-8991722 8991770 -	-0.15326	0.04995
AT2G21960 A5'SS Chr2 9355369-9355482 9355364-9355482 +	0.041684	0.04995
AT2G21960 A5'SS Chr2 9355371-9355482 9355364-9355482 +	0.145577	0.04995
AT3G07580 IR Chr3 2420838 2421275-2421381 2421433 -	-0.17242	0.04995
AT3G20630 IR Chr3 7205539 7205593-7205863 7205940 -	-0.07701	0.04995
AT3G27380 IR Chr3 10129610 10130509-10130612 10131374 -	0.0337	0.04995
AT3G60240 A3'SS Chr3 22262256-22262372 22262256-22262378 +	-0.07721	0.04995
AT3G60240 A3'SS Chr3 22262512-22262632 22262512-22262638 +	0.082844	0.04995
AT4G19840 AF Chr4 10774273 10774581-10774969 10774652 10774657-10774969 +	0.016937	0.04995
AT4G29170 A3'SS Chr4 14382987-14383457 14382987-14383552 +	0.275812	0.04995
AT5G20950 A3'SS Chr5 7110819-7111196 7110797-7111196 -	-0.05634	0.04995
AT5G24155 IR Chr5 8179703 8180231-8180549 8180660 -	-0.36104	0.04995

AS event	ΔPSI value	P value
AT5G45190 IR Chr5 18278879 18279030-18279120 18279295 -	-0.11863	0.04995
AT1G03380 IR Chr1 837471 837908-838069 838314 +	-0.11064	0.049451
AT1G06550 A3'SS Chr1 2006116-2006196 2006109-2006196 -	0.103206	0.049451
AT1G07728 IR Chr1 2395451 2396434-2396529 2397345 +	0.083457	0.049451
AT1G08680 IR Chr1 2764593 2764689-2765056 2765451 +	-0.1973	0.049451
AT1G47240 A3'SS Chr1 17310366-17310737 17310307-17310737 -	-0.07823	0.049451
AT1G48090 A3'SS Chr1 17755777-17755939 17755769-17755939 -	-0.05488	0.049451
AT2G23950 IR Chr2 10186891 10187569-10187643 10188022 -	-0.11943	0.049451
AT3G13340 IR Chr3 4332878 4332964-4333125 4333259 +	0.105306	0.049451
AT3G16240 IR Chr3 5505764 5506014-5506414 5507050 +	0.017512	0.049451
AT4G12460 A3'SS Chr4 7391726-7391877 7391726-7392688 +	-0.35825	0.049451
AT4G12460 A3'SS Chr4 7391726-7391877 7391726-7392691 +	-0.32388	0.049451
AT4G30160 AF Chr4 14753314 14753371-14753726 14753432 14753502-14753726 +	0.305525	0.049451
AT1G78070 A3'SS Chr1 29357219-29357303 29357219-29357306 +	0.018268	0.049151
AT1G07640 A5'SS Chr1 2355699-2355796 2355699-2355986 -	0.217224	0.048951
AT1G13700 IR Chr1 4694372 4694972-4695209 4695344 -	0.161491	0.048951
AT1G13700 IR Chr1 4695428 4695638-4695979 4696023 -	0.499389	0.048951
AT1G21400 ES Chr1 7495266-7495366 7495477-7495572 +	0.084544	0.048951
AT1G21450 IR Chr1 7508969 7509175-7509717 7511773 +	-0.16778	0.048951
AT1G31870 A3'SS Chr1 11438496-11438674 11438496-11438774 +	0.073549	0.048951
AT1G59840 IR Chr1 22028085 22028251-22028487 22029379 +	-0.13091	0.048951
AT1G64140 A3'SS Chr1 23806302-23806924 23806299-23806924 -	-0.04023	0.048951
AT2G34680 A5'SS Chr2 14628317-14628647 14628317-14628860 -	-0.10922	0.048951
AT2G43200 IR Chr2 17958230 17958898-17959015 17959225 +	-0.28553	0.048951

Contrast group 5: AzadC 22°C vs Ctrl 4°C

AS event	ΔPSI value	P value ≤ 0.05
AT1G31175 IR Chr1 11141926 11141961-11142056 11142716 +	0.192001048	0.04995005
AT1G34418 A3'SS Chr1 12582384-12582469 12582384-12582523 +	-0.079709275	0.04995005
AT1G60990 IR Chr1 22462026 22462390-22462720 22462956 -	0.103784612	0.04995005
AT1G60990 IR Chr1 22462026 22462393-22462720 22462956 -	0.175290073	0.04995005
AT1G75670 IR Chr1 28415486 28415590-28415674 28415759 -	-0.062734525	0.04995005
AT2G04039 IR Chr2 1333262 1333565-1333640 1333703 +	-0.092782253	0.04995005
AT2G32690 A3'SS Chr2 13864025-13864311 13863971-13864311 -	0.089304237	0.04995005
AT2G36670 ES Chr2 15367302-15367400 15367414-15367724 -	-0.191285013	0.04995005
AT3G26100 A5'SS Chr3 9537827-9537905 9537402-9537905 +	0.12192759	0.04995005
AT3G27430 ES Chr3 10152616-10152892 10153053-10153551 +	-0.042235725	0.04995005

AT3G43540 IR Chr3 15430660 15431017-15431095 15431298 +	-0.056176294	0.04995005
AT4G08460 A3'SS Chr4 5378205-5380531 5378202-5380531 -	0.123556307	0.04995005
AT4G08460 ES Chr4 5378202-5378698 5378832-5380531 -	-0.157725512	0.04995005
AT4G14880 A3'SS Chr4 8520066-8520332 8520063-8520332 -	0.050740875	0.04995005
AT5G27390 IR Chr5 9674391 9674726-9674820 9674892 -	-0.053774539	0.04995005
AT5G55896 A3'SS Chr5 22631438-22632426 22627853-22632426 -	0.37809816	0.04995005
AT5G55896 ES Chr5 22632511-22632678 22632769-22633201 -	-0.136512358	0.04995005
AT3G26890 A5'SS Chr3 9910792-9910873 9910792-9911368 -	0.375296201	0.049825175
AT3G26890 A5'SS Chr3 9910792-9910873 9910792-9911415 -	0.287953841	0.049825175
AT1G67900 A3'SS Chr1 25464987-25467032 25464987-25467052 +	-0.185303527	0.04978355
AT1G67900 ES Chr1 25467159-25467376 25467421-25467619 +	0.186780334	0.04978355
AT2G21660 A3'SS Chr2 9265541-9265759 9265514-9265759 -	-0.23119514	0.049586777
AT2G21660 A5'SS Chr2 9265541-9265722 9265541-9265759 -	0.383183329	0.049586777
AT1G55520 A3'SS Chr1 20726058-20726144 20726055-20726144 -	0.095437773	0.04945055
AT1G73480 A3'SS Chr1 27630771-27630865 27630771-27630871 +	-0.014145104	0.04945055
AT1G73650 A3'SS Chr1 27688461-27688561 27688457-27688561 -	0.031130775	0.04945055
AT1G80245 A5'SS Chr1 30174627-30174710 30174349-30174710 +	-0.507517071	0.04945055
AT2G39240 IR Chr2 16387741 16387859-16388038 16388502 -	0.184251634	0.04945055
AT2G48070 IR Chr2 19663747 19663838-19664218 19664465 +	-0.038359333	0.04945055
AT3G21710 IR Chr3 7648989 7649060-7649348 7649894 +	0.383497695	0.04945055
AT4G00970 IR Chr4 419602 419735-420477 420687 +	0.148888776	0.04945055
AT4G00970 IR Chr4 420778 421021-421113 421263 +	0.143782852	0.04945055
AT4G34640 IR Chr4 16539652 16539721-16539996 16540145 +	0.069636515	0.04945055
AT4G35920 A3'SS Chr4 17012904-17013006 17012875-17013006 -	0.18394715	0.04945055
AT4G38970 A5'SS Chr4 18163992-18164127 18163992-18164131 -	0.003036497	0.04945055
AT5G01910 IR Chr5 357675 358084-358164 358229 -	0.599188497	0.04945055
AT5G04280 A3'SS Chr5 1192571-1192966 1192571-1194495 +	-0.085914261	0.04945055
AT1G73650 A3'SS Chr1 27688457-27688561 27688453-27688561 -	0.037628444	0.049284049
AT4G02430 IR Chr4 1068974 1069081-1069175 1069272 +	0.449327975	0.049236478
AT4G32660 A5'SS Chr4 15758025-15758155 15758004-15758155 +	-0.094483971	0.049034299
AT4G32660 IR Chr4 15757934 15758004-15758155 15758302 +	0.146979732	0.049034299
AT1G15350 A3'SS Chr1 5279194-5279482 5279189-5279482 -	-0.092079994	0.048951049
AT3G27310 IR Chr3 10087891 10087995-10088075 10088204 -	0.112026878	0.048951049
AT3G55170 IR Chr3 20452910 20453210-20453279 20453435 -	0.03764485	0.048951049
AT3G59600 ES Chr3 22016707-22016934 22017091-22017605 +	-0.055985924	0.048951049
AT4G00180 A3'SS Chr4 73710-74268 73707-74268 -	0.071041796	0.048951049
AT4G18975 A3'SS Chr4 10393695-10393768 10393673-10393768 -	0.164413816	0.048951049
AT5G15190 A3'SS Chr5 4933043-4933118 4933034-4933118 -	0.103685847	0.048951049
AT4G22570 IR Chr4 11882647 11882796-11882864 11882956 -	0.039794683	0.048701299

AS event	ΔPSI value	P value ≤ 0.05
AT5G06980 A3'SS Chr5 2168109-2168219 2168109-2168232 +	0.105215918	0.048701299

AS event	ΔPSI value	P value ≤ 0.05
AT1G27630 A5'SS Chr1 9612429-9612590 9611876-9612590 +	-0.15901	0.04995
AT1G33720 IR Chr1 12222407 12223202-12223587 12224108 -	0.193478	0.04995
AT1G54610 A3'SS Chr1 20394264-20394585 20394261-20394585 -	0.080235	0.04995
AT1G62200 IR Chr1 22982036 22982945-22983033 22983598 -	0.09666	0.04995
AT1G73760 IR Chr1 27739591 27739728-27739830 27739928 -	0.092062	0.04995
AT2G16920 A5'SS Chr2 7338827-7339134 7338827-7339145 -	0.101758	0.04995
AT2G22720 A3'SS Chr2 9658994-9659075 9658994-9659095 +	-0.10563	0.04995
AT2G47250 A3'SS Chr2 19400925-19401240 19400816-19401240 -	-0.10177	0.04995
AT2G47960 A5'SS Chr2 19626856-19627234 19626828-19627234 +	0.102798	0.04995
AT3G09405 IR Chr3 2896019 2896079-2896230 2896354 -	-0.28755	0.04995
AT3G61010 IR Chr3 22574073 22574183-22574455 22574491 -	-0.15904	0.04995
AT4G02450 A3'SS Chr4 1074353-1074627 1074344-1074627 -	-0.15428	0.04995
AT4G02450 A3'SS Chr4 1074371-1074627 1074344-1074627 -	-0.22872	0.04995
AT4G02450 A3'SS Chr4 1074389-1074627 1074344-1074627 -	-0.26051	0.04995
AT4G10120 IR Chr4 6314787 6314925-6315017 6315290 +	0.083352	0.04995
AT4G23300 ES Chr4 12182854-12182934 12183061-12183172 +	-0.32894	0.04995
AT4G38225 IR Chr4 17927565 17928196-17928285 17928485 +	0.045228	0.04995
AT5G08450 A3'SS Chr5 2732642-2732741 2732605-2732741 -	0.139782	0.04995
AT5G08450 IR Chr5 2730906 2732602-2732741 2732893 -	0.258892	0.04995
AT5G27380 IR Chr5 9669630 9669752-9669835 9669965 -	-0.0819	0.04995
AT5G51300 IR Chr5 20848794 20849263-20849845 20852384 -	-0.11351	0.04995
AT5G60580 IR Chr5 24353562 24354185-24354296 24354369 +	0.179765	0.04995
AT1G36390 A3'SS Chr1 13701639-13701799 13701603-13701799 -	-0.18954	0.049451
AT2G31350 A3'SS Chr2 13368504-13368774 13368504-13368777 +	0.044667	0.049451
AT2G31350 A5'SS Chr2 13369278-13369375 13369265-13369375 +	0.049035	0.049451
AT2G31350 IR Chr2 13369375 13369498-13370094 13370194 +	-0.04519	0.049451
AT3G27990 IR Chr3 10397606 10397683-10397772 10398175 -	0.145414	0.049451
AT4G19660 A5'SS Chr4 10696813-10696895 10696813-10696918 -	-0.13312	0.049451
AT4G19660 EI Chr4 10696266 10696525-10696607 10696813 -	0.139352	0.049451
AT5G37370 A5'SS Chr5 14814848-14815165 14814848-14815698 -	-0.13788	0.049451
AT2G32690 A3'SS Chr2 13864028-13864311 13863962-13864311 -	0.415872	0.049042
AT2G23980 IR Chr2 10202598 10202911-10202981 10203193 -	0.099881	0.048951
AT2G23980 IR Chr2 10203941 10204031-10204388 10204467 -	0.132417	0.048951

AT2G44420 ES Chr2 18330636-18330846 18331028-18331133 +	-0.14443	0.048951
AT2G45380 ES Chr2 18700897-18701098 18701377-18701537 -	-0.45022	0.048951
AT3G12280 A3'SS Chr3 3913845-3913928 3913842-3913928 -	0.058247	0.048951
AT4G20380 IR Chr4 11004835 11004896-11005025 11005150 +	0.369041	0.048951
AT4G22890 A3'SS Chr4 12007355-12007788 12007355-12007791 +	-0.15724	0.048951
AT4G22890 A3'SS Chr4 12009185-12009258 12009185-12009262 +	-0.03608	0.048951
AT4G27700 A3'SS Chr4 13827400-13827502 13827393-13827502 -	0.027289	0.048951
AT5G12210 A3'SS Chr5 3947558-3947671 3947558-3947674 +	0.058099	0.048951
AT5G37480 A5'SS Chr5 14885378-14885497 14885378-14885807 -	-0.30079	0.048951
AT5G37480 ES Chr5 14885378-14885497 14885551-14885807 -	-0.04017	0.048951
AT5G50280 A3'SS Chr5 20460100-20460178 20460100-20460196 +	0.149751	0.048951
AT5G53050 ES Chr5 21511966-21512348 21512443-21512644 -	-0.06934	0.048951
AT5G61410 A3'SS Chr5 24683858-24684045 24683852-24684045 -	-0.01353	0.048951
AT3G44630 AF Chr3 16195860 16196078-16196264 16196111 16196147-16196264 +	-0.23152	0.048701
AT5G26850 A3'SS Chr5 9445530-9445936 9445530-9445944 +	-0.1694	0.048701
AT5G26850 A5'SS Chr5 9445646-9445936 9445530-9445936 +	0.181545	0.048701
AT2G45070 AF Chr2 18587493-18588054 18588141 18587493-18588192 18588452 -	0.090043	0.048618

To further, investigate to which extent this remains true for the most common local AS events (IR, A3'SS, A5'SS, and ES) in Arabidopsis, total significant differential AS events has been split into different AS categories (IR-excluding exitrons, A3'SS, A5'SS, SE, EI, AF, AL, and MX. As the number of differential AS exitrons represent only 8.5% of the total IR events detected by SUPPA and AF, AL, MX events are less represented in our AS events, we have considered IR without exitrons and EIs as another type of IR splicing events and excluded AF, AL, MX from our study (Table 3.5).

Table 3.5. Number of differential AS events in different contrast group. C1 : Contrast group 1 : AzadC 22°C vs Ctrl 22°C, C2 : Contrast group 2 : AzadC 4°C vs Ctrl 4°C, C3 : Contrast group 3 : AzadC 4°C vs AzadC 4°C, C4 : Contrast group 4 : Ctrl 4°C vs Ctrl 22°C , C5 : Contrast group 5 : AzadC 22°C vs Ctrl 4°C, C6 : Contrast group 6 : Ctrl 22°C vs AzadC 4°C.

	IR	EI	A3'SS	A5'SS	ES	AF	AL	MX	Total number of events
C1	118	7	78	48	8	7	1	0	267
C2	298	8	99	65	21	7	3	1	502
C3	814	45	651	358	206	52	11	1	2138
C4	1108	54	791	415	217	61	12	2	2660
C5	1138	56	792	426	229	70	12	3	2726
C6	779	46	674	359	192	41	12	1	2122

Then, the distribution of mean PSI was plotted against along with the expression of different AS events in different contrast groups (Figure 3.9).
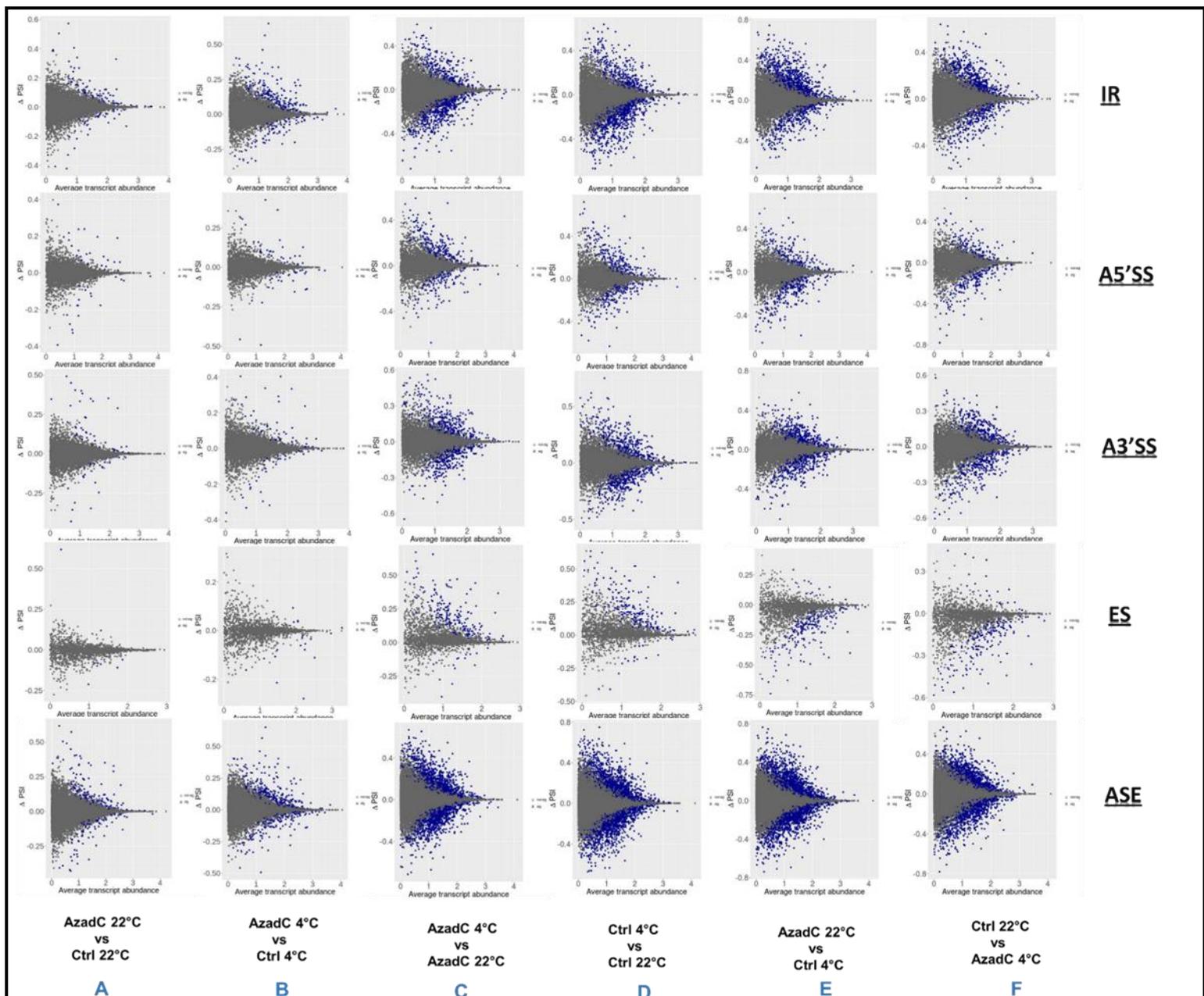
Figure 3.9. Plot representing the distribution of mean PSI (ΔPSI) detected by SUPPA along with the expression of different AS events in different contrast groups, and the P value of this difference. The x axis is the DeltaPSI (ΔPSI); the y axis is the average transcript abundance. Blue and grey dots represent significant (p-value < 0.05) and non-significant events (p-value > 0.05), respectively. IR: Intron retention, A3'SS: alternative 3' splice site, A5'SS: alternative 5'SS, ES: exon skipping, ASE: all splicing events. For ASE events, under the same temperature conditions, AzadC and Ctrl plants display the lowest number of significant DAS events (267 and 502, respectively) in contrast group A and B. This number increases to 2138 and 2660 in AzadC and Ctrl plants, respectively upon shift from normal temperature (22^0 C) to cold stress (4^0 C). The number of ASE reaches a maximum of 2726 DAS events in contrast group E in which changes in gene expression are regulated by DNA methylation changes. The distribution of mean PSI of individual AS events shows that an IR (First panel from top) event is the major type changing for all contrast groups. Similar to overall ASE changes detected, 125, 306, 859, 1162, 1194 and 843 differential IR events were detected in contrast groups A, B, C, D, E, and F, respectively. The second major AS event type detected is A3'SS for which, similar to RI, 78, 99, 651, 791, 792, and 674 differential A3'SS are detected in groups A, B, C, D, E, and F, respectively. A5'SS splicing event takes place as the third major differential AS event

detected, where 48, 65, 358, 415, 426, 359 are detected in A, B, C, D, E, and F, respectively. Finally, the least represented AS event, ES, shows 8, 21, 206, 217, 229, and 129 differential ES are detected in groups A, B, C, D, E, and F, respectively. This data clearly show that regardless of the AS event, DNA methylation is likely to regulate differential AS upon temperature shift.

Interestingly, IR events are the most prevalent AS event influenced methylation changes and/or cold stress followed by A5'SS and A3'SS, whereas ES was least affected upon DNA methylation or temperature changes. This is similar to the overall frequency of alternative splicing events observed in Arabidopsis (Marquez et al., 2015). Thus, changes in DNA methylation seem to affect all types of splicing events similarly and doesn't affect the frequency of local AS events under normal growth conditions and upon cold stress. Additionally, this data show that DNA methylation regulates the expression of different types of AS events, potentially through different organisation of chromatin structure around splice junction which can subsequently affect the reorganisation of splice sites by the splicing machinery.

3.5 Discussion

In Arabidopsis and many other plant species, transcriptome changes play a crucial role in plant growth and adaptation to stressful environmental signals. AS is a co-transcriptional mechanism that contributes to transcriptome plasticity and dynamics through diversifying the transcript pool to determine plant gene expression profiles (Calixto et al., 2018; S. A. Filichkin et al., 2015, 2018), However, to date, the role of epigenetic landscape in modulating gene expression and AS profiles upon cold stress in Arabidopsis is not fully understood. To understand this, in this chapter, epigenetically different Arabidopsis plants (AzadC and Ctrl) with the same genetic background has been used to examine if they exhibit any similarities/changes in gene expression and AS profiles upon cold stress treatment as well as normal growth conditions.

Deep RNA-seq is a specific and sensitive technique that allows reliable detection and quantification of individual transcripts, and helps towards the detection of novel transcripts and genes (Nagalakshmi, Waern, & Snyder, 2010). Furthermore, the sensitivity of RNA-seq, which is due to the high read coverage, allows the detection and accurate quantification of low expressed or rare transcripts hence; helping towards more precise differential gene expression and AS analysis.

To obtain authentic and accurate gene expression and AS analysis from RNA-seq data, RNA-Seq reads needs to be first quantified using a high quality and complete reference transcriptome (Brown, Calixto, & Zhang, 2017). For this reason, AtRTD2 has been used as reference transcript database to quantify RNA-seq reads using Salmon; a newly developed accurate, fast, and lightweight algorithms (Patro, Duggal, Love, Irizarry, & Kingsford, 2017; R. Zhang et al., 2017). Compared to previous Arabidopsis transcript datasets (TAIR10, AtRTD1, and Araport11), AtRTD2 contained around 82k non-redundant transcript isoforms which are a result of stringent filtering of mis-assembled and false transcripts. The authenticity of AtRTD2 has been demonstrated by high resolution reverse transcriptase polymerase chain reaction (HR RT-PCR), which showed high correlation with transcript quantification levels. Quantification of RNA-seq reads has been measured for isoform quantification using TPM (Rather than counts, TMM, FPKM) because is normalized for gene length first, and then for sequencing depth. The use of high-quality Arabidopsis reference transcript database and accurate lightweight program for RNA-seq reads quantification provided an adequate gene expression and AS analysis, which showed that epigenetic differences in genetically identical Arabidopsis plants are sufficient to modulate transcriptome under normal conditions and cold stress.

The major finding displayed in this chapter, is that AzadC plants with hypomethylation levels display significant changes in gene expression and AS profiles under normal conditions, which are different from the ones observed for Ctrl. This implies the role of DNA methylation in modulating the chromatin structure to subsequently affect RNAPll processivity and accumulation and affect gene expression and AS profiles. Additionally, the results presented in this chapter also emphasise the importance of DNA in plants cold stress response; where plants with epigenetic differences respond differently to cold stress at the gene and transcript levels. These results suggest that DNA methylation can perceive environmental signals in plants to help their environmental fitness, as differences in DNA methylation between Ctrl and AzadC result in different responses to cold stress. The second finding presented here is that changes in DNA methylation affect the overall transcript expression rather than affecting the frequency of AS events. This was detected by SUPPA, which showed that plants with differences in DNA methylation don't display any changes in the frequency of local AS events where IR events remains the highest type of AS event in Ctrl and AzadC.

Interestingly, GO analysis show that differences in DNA methylation between AzadC and Ctrl result in different gene expression upon cold stress of genes involved in plants development, growth, hormone signalling, cold response, and circadian rhythm. The observed differences in gene expression between AzadC and Ctrl were detected in multiple cellular components such as cell wall, cell membrane, and vacuole. At the transcript level, differences in DNA methylation between AzadC and Ctrl plants also result in changing AS profiles of genes involved in mRNA splicing in the nucleus.

Overall, the results presented in this chapter are the first evidence in Arabidopsis showing that changes in epigenetic features are sufficient to change gene expression and AS profiles of genes which are involved in multiple physiological processes and cold acclimation. Despite the evidence (the major three finding explained in this section) presented here, future studies can benefit from these findings to validate the changes in gene expression and AS profiles through using HR-RT PCR and/or mutants. Alternatively, future studies can focus on using CRISPR/deadCas9 systems coupled with demethylation enzymes to engineer important traits and modulate splicing variation (See the final chapter).

References

Alexander, R. D. *et al.* (2010) 'Splicing-Dependent RNA polymerase pausing in yeast', *Molecular Cell*, 40(4), pp. 582–593. doi: 10.1016/j.molcel.2010.11.005.

Amado, L. *et al.* (1997) 'Development-dependent inheritance of 5-azacytidine-induced epimutations in triticale: Analysis of rDNA expression patterns', *Chromosome Research*, 5, pp. 445–450. doi: 10.1023/A:1018460828720.

Barrero-Gil, J. and Salinas, J. (2013) 'Post-translational regulation of cold acclimation response', *Plant Science*, 205–206, pp. 48–54. doi: 10.1016/j.plantsci.2013.01.008.

Brown, J. W. S., Calixto, C. P. G. and Zhang, R. (2017) 'High-quality reference transcript datasets hold the key to transcript-specific RNA-sequencing analysis in plants', *New Phytologist*, 213(2), pp. 525–530. doi: 10.1111/nph.14208.

Burn, J. E. *et al.* (1993) 'DNA methylation, vernalization, and the initiation of flowering', *Proc. Natl. Acad. Sci.*, 90(1), pp. 287–291. doi: 10.1073/PNAS.90.1.287.

Calixto, C. P. G. *et al.* (2018) 'Rapid and dynamic alternative splicing impacts the Arabidopsis cold response transcriptome.', *The Plant cell*, 30(7), pp. 1424–1444. doi: 10.1105/tpc.18.00177.

Chamala, S. *et al.* (2015) 'Genome-Wide Identification of Evolutionarily Conserved Alternative Splicing Events in Flowering Plants', *Frontiers in Bioengineering and Biotechnology*, 3, p. 33. doi: 10.3389/fbioe.2015.00033.

Chaudhary, S. *et al.* (2019) 'Perspective on Alternative Splicing and Proteome Complexity in Plants', *Trends in Plant Science*, 24(6), pp. 496–506. doi: 10.1016/j.tplants.2019.02.006.

Chen, W., Luo, L. and Zhang, L. (2010) 'The organization of nucleosomes around splice sites', *Nucleic Acids Research*, 38(9), pp. 2788–2798. doi: 10.1093/nar/gkq007.

Chodavarapu, R. K. *et al.* (2010) 'Relationship between nucleosome positioning and DNA methylation', *Nature*, 466(7304), pp. 388–392. doi: nature09147 [pii]\n10.1038/nature09147.

Christman, J. (2002) '5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy.', *Oncogene*, 21(35), pp. 5483–5495. doi: 10.1038/sj.onc.1205699.

Cokus, S. J. *et al.* (2008) 'Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning', *Nature*, 452(7184), pp. 215–219. doi: 10.1038/nature06745.

Ding, Y. *et al.* (2014) 'In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features', *Nature*, 505(7485), pp. 696–700. doi: 10.1038/nature12756.

Dowen, R. H. *et al.* (2012) 'Widespread dynamic DNA methylation in response to biotic stress', *Proceedings of the National Academy of Sciences*, 109(32), pp. E2183-91. doi: 10.1073/pnas.1209329109.

Ehrlich, M. *et al.* (1982) 'Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells', *Nucleic Acids Research*, 10(8), pp. 2709–2721. doi: 10.1093/nar/10.8.2709.

Fieldes, M. A. (1994) 'Heritable effects of 5-azacytidine treatments on the growth and development of flax (Linum usitatissimum) genotrophs and genotypes', *Genome*, 37(1), pp.

1–11. doi: 10.1139/g94-001.

Filichkin, S. A. *et al.* (2015) 'Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis', *Molecular Plant*, 8(2), pp. 207–227. doi: 10.1016/j.molp.2014.10.011.

Filichkin, S. A. *et al.* (2018) 'Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching', *Frontiers in Plant Science*, 9, p. 5. doi: 10.3389/fpls.2018.00005.

Garg, R. *et al.* (2015) 'Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response', *Scientific Reports*, 5, p. 14922. doi: 10.1038/srep14922.

Gasch, A. *et al.* (2006) 'The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains', *Journal of Biological Chemistry*, 281(1), pp. 356–364. doi: 10.1074/jbc.M508047200.

Gelfman, S. *et al.* (2013) 'DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure', *Genome Research*, 23(5), pp. 789–799. doi: 10.1101/gr.143503.112.

Goodrich, J. A. and Tjian, R. (2010) 'Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation', *Nature Reviews Genetics*, 11(8), pp. 549–58. doi: 10.1038/nrg2847.

Hajheidari, M., Koncz, C. and Eick, D. (2013) 'Emerging roles for RNA polymerase II CTD in Arabidopsis', *Trends in Plant Science*, 18(11), pp. 633–643. doi: 10.1016/j.tplants.2013.07.001.

Hirose, Y. and Manley, J. L. (2000) 'RNA polymerase II and the integration of nuclear events', *Genes and Development*, 14(12), pp. 1415–1429.

Huff, J. T. and Zilberman, D. (2014) 'Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes', *Cell*, 156(6), pp. 1286–97. doi: 10.1016/j.cell.2014.01.029.

Jabre, I. *et al.* (2019) 'Does co-transcriptional regulation of alternative splicing mediate plant stress responses?', *Nucleic Acids Research*. Oxford University Press, 47(6), pp. 2716–2726. doi: 10.1093/nar/gkz121.

James, A. B. *et al.* (2012) 'Alternative Splicing Mediates Responses of the Arabidopsis Circadian Clock to Temperature Changes', *The Plant Cell*, 24(3), pp. 961–981. doi: 10.1105/tpc.111.093948.

Janoušek, B., Široký, J. and Vyskot, B. (1996) 'Epigenetic control of sexual phenotype in a dioecious plant, Melandrium album', *Molecular and General Genetics*, 250(4), p. 483–490. doi: 10.1007/s004380050101.

King, G. J. (1995) 'Morphological development in Brassica oleracea is modulated by in vivo treatment with 5-azacytidine', *Journal of Horticultural Science*, 70, pp. 333–342. doi: 10.1080/14620316.1995.11515304.

Knight, M. R. and Knight, H. (2012) 'Low-temperature perception leading to gene expression and cold tolerance in higher plants', *New Phytologist*, 195(4), pp. 737–751. doi: 10.1111/j.1469-8137.2012.04239.x.

Kumar, S. V. and Wigge, P. A. (2010) 'H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis', *Cell*, 140(1), pp. 136–147. doi: 10.1016/j.cell.2009.11.006.

Kumpatla, S. P. *et al.* (1997) 'Epigenetic Transcriptional Silencing and 5-Azacytidine-Mediated Reactivation of a Complex Transgene in Rice', *Plant Physiology*, 115(2), pp. 361–373. doi: 10.1104/pp.115.2.361.

Lenasi, T. and Barboric, M. (2010) 'P-TEFb stimulates transcription elongation and pre-mRNA splicing through multilateral mechanisms', *RNA Biology*, 7(2), pp. 145–150. doi: 10.4161/rna.7.2.11057.

Ling, Y. *et al.* (2018) 'Thermopriming triggers splicing memory in Arabidopsis', *Journal of Experimental Botany*, 69(10), pp. 2659–2675. doi: 10.1093/jxb/ery062.

Lister, R. *et al.* (2008) 'Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis', *Cell*, 133(3), pp. 523–536. doi: 10.1016/j.cell.2008.03.029.

Liu, J. *et al.* (2015) 'Genetic and epigenetic control of plant heat responses', *Frontiers in Plant Science*, 6, p. 267. doi: 10.3389/fpls.2015.00267.

Liu, M.-J. *et al.* (2015) 'Determinants of nucleosome positioning and their influence on plant gene expression', *Genome Research*, 25(8), pp. 1182–1195. doi: 10.1101/gr.188680.114.

Lu, X. *et al.* (2017) 'Single-base resolution methylomes of upland cotton (Gossypium hirsutum L.) reveal epigenome modifications in response to drought stress', *BMC Genomics*, 18(1), p. 297. doi: 10.1186/s12864-017-3681-y.

Luna, E. *et al.* (2012) 'Next generation systemic acquired resistance', *Plant physiology*, 158(2), pp. 844–853. doi: 10.1104/pp.111.187468.

Marfil, C. F., Asurmendi, S. and Masuelli, R. W. (2012) 'Changes in micro RNA expression in a wild tuber-bearing Solanum species induced by 5-Azacytidine treatment', *Plant Cell Reports*, 31, pp. 1449–1461. doi: 10.1007/s00299-012-1260-x.

Marquez, Y. *et al.* (2012) 'Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis', *Genome Research*, 22(6), pp. 1184–1195. doi: 10.1101/gr.134106.111.

Mastrangelo, A. M. *et al.* (2012) 'Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity', *Plant Science*, 185–186, pp. 40–49. doi: 10.1016/j.plantsci.2011.09.006.

Mavrich, T. N. *et al.* (2008) 'Nucleosome organization in the Drosophila genome', *Nature*, 453(7193), pp. 358–362. doi: 10.1038/nature06929.

McClung, C. R. and Davis, S. J. (2010) 'Ambient thermometers in plants: From physiological outputs towards mechanisms of thermal sensing', *Current Biology*, p. 20(24):1086-1092. doi: 10.1016/j.cub.2010.10.035.

Nagalakshmi, U., Waern, K. and Snyder, M. (2010) 'RNA-seq: A method for comprehensive transcriptome analysis', in *Current Protocols in Molecular Biology*, pp. 1–13. doi: 10.1002/0471142727.mb0411s89.

Nahkuri, S., Taft, R. J. and Mattick, J. S. (2009) 'Nucleosomes are preferentially positioned at exons in somatic and sperm cells', *Cell Cycle*, 8(20), pp. 3420–3424. doi:

10.4161/cc.8.20.9916.

Palusa, S. G., Ali, G. S. and Reddy, A. S. N. (2007) 'Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: Regulation by hormones and stresses', *Plant Journal*, 49(6), pp. 1091–1107. doi: 10.1111/j.1365-313X.2006.03020.x.

Pan, Y. *et al.* (2011) 'DNA methylation alterations of rice in response to cold stress', *Plant OMICS*, 4(7), pp. 364–369.

Patro, R. *et al.* (2017) 'Salmon provides fast and bias-aware quantification of transcript expression', *Nature Methods*, 14(4), pp. 417–419. doi: 10.1038/nmeth.4197.

Rasmann, S. *et al.* (2012) 'Herbivory in the Previous Generation Primes Plants for Enhanced Insect Resistance', *Plant Physiology*, 158(2), pp. 854–863. doi: 10.1104/pp.111.187831.

Reddy, A. S. N. *et al.* (2013) 'Complexity of the alternative splicing landscape in plants.', *The Plant cell*, 25(10), pp. 3657–83. doi: 10.1105/tpc.113.117523.

Roy, D. *et al.* (2014) 'Differential acetylation of histone H3 at the regulatory region of OsDREB1b promoter facilitates chromatin remodelling and transcription activation during cold stress', *PLoS One*, 9(6), p. e100343. doi: 10.1371/journal.pone.0100343.

Sano, H. *et al.* (1990) 'A single treatment of rice seedlings with 5-azacytidine induces heritable dwarfism and undermethylation of genomic DNA', *MGG Molecular & General Genetics*, 220, pp. 441–447. doi: 10.1007/BF00391751.

Schwartz, S., Meshorer, E. and Ast, G. (2009) 'Chromatin organization marks exon-intron structure.', *Nature Structural & Molecular Biology*, 16(9), pp. 990–995. doi: 10.1038/nsmb.1659.

Secco, D. *et al.* (2015) 'Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements', *eLife*, 4, p. e.09343. doi: 10.7554/eLife.09343 [doi].

Shen, H., Kan, J. L. C. and Green, M. R. (2004) 'Arginine-Serine-Rich Domains Bound at Splicing Enhancers Contact the Branchpoint to Promote Prespliceosome Assembly', *Molecular Cell*, 13(3), pp. 367–376. doi: 10.1016/S1097-2765(04)00025-5.

Steward, N. *et al.* (2002) 'Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress', *Journal of Biological Chemistry*, 277(40), pp. 37741–37746. doi: 10.1074/jbc.M204050200.

Syed, N. H. *et al.* (2012) 'Alternative splicing in plants - coming of age', *Trends in Plant Science*, 17(10), pp. 616–623. doi: 10.1016/j.tplants.2012.06.001.

Thomashow, M. F. (2010) 'Molecular Basis of Plant Cold Acclimation: Insights Gained from Studying the CBF Cold Response Pathway', *Plant Physiology*, 154(2), pp. 571–577. doi: 10.1104/pp.110.161794.

Tilgner, H. *et al.* (2009) 'Nucleosome positioning as a determinant of exon recognition', *Nature Structural and Molecular Biology*, 16(9), pp. 996–1001. doi: 10.1038/nsmb.1658.

Trincado, J. *et al.* (2018) 'SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions', *Genome Biology*, 19, pp. 40. doi: 10.1186/s13059-018-1417-1

Ullah, F. *et al.* (2018) 'Exploring the relationship between intron retention and chromatin accessibility in plants', *BMC Genomics*, 19(1), p. 21. doi: 10.1186/s12864-017-4393-z.

Vyskot, B. *et al.* (1995) 'Meiotic transmission of a hypomethylated repetitive DNA family in tobacco', *Theoretical and Applied Genetics*, 91(4), pp. 659–664. doi: 10.1007/BF00223294.

Xu, J. *et al.* (2016) 'Quantitative trait variation is revealed in a novel hypomethylated population of woodland strawberry (Fragaria vesca)', *BMC Plant Biology*, 16(1), p. 240. doi: 10.1186/s12870-016-0936-8.

Yu, H. *et al.* (2016) 'Transcriptome Survey of the Contribution of Alternative Splicing to Proteome Diversity in Arabidopsis thaliana', *Molecular Plant*, 9(5), pp. 749–752. doi: 10.1016/j.molp.2015.12.018.

Zhang, G. *et al.* (2010) 'Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome', *Genome Research*, 20(5), pp. 646–654. doi: 10.1101/gr.100677.109.

Zhang, R. *et al.* (2017) 'A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing', *Nucleic Acids Research*, 45(9), pp. 5061–5073. doi: 10.1093/nar/gkx267.

Zhu, J. *et al.* (2018) 'RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis', *Nature Plants*. Springer US, 4(12), pp. 1112–1123. doi: 10.1038/s41477-018-0280-0.

Chapter 4. Nucleosome and DNA methylation profiles in AzadC and Ctrl plants modulate gene expression and AS patterns

4.1 Introduction

In eukaryotes, DNA methylation occurs in symmetric CG and CHG (where H = A, T or C) and the asymmetric CHH contexts (Ehrlich et al., 1982). In plants, DNA methylation is largely dependent on the CpG context representing 24%, whereas CHG and CHH is only 6.7% and 1.7%, respectively of the Arabidopsis methylated genome (Cokus et al., 2008; Lister et al., 2008). Interestingly, nucleosome DNA is more highly methylated and exons, rather than the introns, are marked at the DNA level by high occupancy of nucleosomes and are preferentially positioned at intron-exon and exon-intron boundaries in both mammals and Arabidopsis (Chodavarapu et al., 2010; M.-J. Liu et al., 2015; Mavrich et al., 2008; S. Schwartz et al., 2009). Since exons are usually GC-rich, transcription through nucleosome-rich regions with compact chromatin tends to be slower (Chodavarapu et al., 2010; Churchman & Weissman, 2011; M.-J. Liu et al., 2015; Singh & Padgett, 2009). Interestingly, nucleosome occupancy is also lower in alternatively spliced exons compared to constitutively spliced exons (Wei Chen et al., 2010; Gelfman et al., 2013; S. Schwartz et al., 2009; Tilgner et al., 2009). These findings indicate that nucleosome positioning influences DNA methylation patterning throughout the genome and that DNA methyltransferases preferentially target nucleosome-bound DNA, suggesting a role for DNA methylation in exon definition. Furthermore, similarities between Arabidopsis and human nucleosomal DNA indicate that the relationships between nucleosomes and DNA methyltransferases are conserved. RNA polymerase II (RNAPII) is also enriched on exons relative to introns, consistent with the hypothesis that nucleosome positioning regulates RNAPII speed (Berget, 1995; Chodavarapu et al., 2010; Kornblihtt, 2015; Nahkuri et al., 2009; S. Schwartz et al., 2009; Shayevitch et al., 2018)

DNA methylation patterns in fungi, plants and animals indicate that gene body methylation in eukaryotes is highly conserved and may influence AS (Maria Kalyna, Lopato, Voronin, & Barta, 2006; Mei, Boatwright, Feng, Schnable, & Brad Barbazuk, 2017; Rauch et al., 2014; C. Zhang, Yang, & Yang, 2015). Indeed, DNA methylation affects exon recognition and is influenced by the GC architecture of exons and flanking introns in human (Gelfman et al., 2013). DNA methylation in the honey bee is almost exclusively found in exons. Interestingly, a strong correlation was also found between methylation patterns on alternative exons and splicing patterns of these exons in workers and queens. Intriguingly, reduction in methylation

in the honey bee via RNAi of the dnmt3 (methyl transferase) gene resulted in widespread changes of AS in fat tissues (Li-Byarlay et al., 2013). Recently, a mechanistic link between DNA methylation and AS splicing was demonstrated. A DNA-binding protein, CCCTC-binding factor (CTCF) promoted inclusion of weak upstream exons in the CD45 gene by mediating local RNAPII pausing. Methylation of exon 5 abolished CTCF binding and resulted in complete loss of exon 5 from CD45 transcripts (Shukla et al., 2011). Excitingly, a direct link was very recently provided between DNA methylation and AS by perturbing DNA methylation patterns of alternatively spliced exons using CRISPR-dCas9 proteins (for details, see engineering splicing in Chapter 5) and methylating/demethylating enzyme fusions (Shayevitch et al., 2018). Interestingly, this work demonstrated that changes in the methylation pattern of alternatively spliced exons mediated inclusion levels but had no effect on introns or constitutively spliced exons (Shayevitch *et al.,* 2019).

Plants exhibit extensive DNA methylation variation under different developmental and stress conditions. For example, methylation profiling of leaves revealed that 2.48% of the genome is hyper-methylated under drought stress and influenced the expression of dozens of stress-responsive (hormone related) genes, however DNA methylation patterns were almost completely reversed when plants were re-watered (Lu et al., 2017). Similarly, data from Arabidopsis, Cork oak (*Quercus suber L.*), *Brassica napus* and cotton (*Gossypium hirsutum*) revealed the differential effect of heat stress on global methylation patterns (Junzhong Liu et al., 2015). Interestingly, global DNA methylation levels increase in leaves upon water deficiency in barley, however root tissues display reduced DNA methylation and affect gene expression levels (Chwialkowska et al., 2016). A recent study in Arabidopsis also demonstrated that *Swedish* accessions show that higher levels of gene body methylation are critical for plant adaptation to cooler regions (Dubin et al., 2015).Taken together, these data suggest that stress-induced dynamic changes in DNA methylation regulate plant transcription in an organ specific manner and also regulate plant transcription to acclimatize and adapt plants to different stresses. Since DNA methylation patterns are highly conserved between plants and animals, influence nucleosome occupancy and define exons and RNAPII processivity, it is likely that differential DNA methylation and associated chromatin structure may influence co-transcriptional AS mechanism in plants and animals in a similar manner. For example, PRMT5 methyltransferase (also known as SKB1) increase H4R3sme2 levels in Arabidopsis and suppress the transcription of FLC and a number of stress-responsive genes (Z. Zhang et al., 2011). Upon salt stress, SKB1 disassociate from the chromatin resulting in a reduced

H4R3sme2 level inducing the expression of FLC and stress-responsive genes via increasing the methylation of the small nuclear ribonucleoprotein Sm-like4 (LSM4) (Z. Zhang et al., 2011). In addition, skb1 mutants display pre-mRNA splicing defects caused by reduced Arg symmetric dimethylation of LSM4 (Z. Zhang et al., 2011). This data shows that SKB1 alters the methylation status of H4R3sme2 and LSM4 to link transcription to pre-mRNA splicing during stress responses. Collectively, this data suggests that stress-induced changes in DNA methylation may provide the context through which stress-responsive genes regulate their transcription and co-transcriptional AS patterns and is further supported by similarities of nucleosome positioning and DNA methylation between mammals and plants (S. Schwartz et al., 2009)(Chodavarapu et al., 2010). However, further work needs to illuminate the relationship between dynamic DNA methylation and nucleosome occupancy patterns in regulating plants gene expression and AS profiles in response to environmental stresses.

The results presented in chapter 3 show that plants with epigenetic differences (AzadC plants with hypomethylation DNA levels compared to wild type Ctrl plants) display different gene expression and AS profiles during normal growth conditions or upon cold stress treatment. This data show that DNA methylation potentially modulate the chromatin landscape (i.e: nucleosome occupancy) to influence RNAPII processivity, and subsequently gene expression and splice site selection. To investigate if differences in gene expression and AS profiles detected between Ctrl and AzadC plants are associated with differences in genome-wide distribution of nucleosome occupancy, micrococcal nuclease (MNase) digestion combined with high-throughput sequencing (MNase-seq) has been performed for Ctrl and AzadC at 22°C and treated with cold stress (4°C). Additionally, whole genome bisulphite sequencing (WGBS) has been performed for AzadC plants grown at 22°C and treated with cold stress (4°C) to obtain genome-wide single-base resolution quantification of DNA methylation around genomic features (Exons, introns, and splice sites) and illuminate how DNA methylation patterns change in response to cold stress.

4.2 Materials and Method

4.2.1 Micrococcal nuclease

Cell nuclei are treated with MNase nuclease to cleave double-strand internucleosomal regions of the chromatin, resulting in mononucleosomal fragments associated with single-strand DNA. The genomic DNA is then purified from mononucleosomal fragments for downstream analysis. Previously, MNase is used to determine whether a DNA fragment of interest is within a nucleosome or to detect nucleosome positioning (Carey M, 2005). In the first case, the purified genomic DNA is separated by agarose gel electrophoresis to obtain a ladder of bands corresponding to nucleosome core and the linker visualized by ethidium bromide staining. Then, if a probe corresponding to the DNA fragment of interest is hybridized to the ladder of nucleosomal bands (Southern blot analysis); the DNA fragment in question is within the nucleosome. In the second case, a treatment with a restriction enzyme is essential before agarose gel electrophoresis and Southern blot analysis (Carey M, 2005). MNase technique combined with gel electrophoresis and Southern blot limit the detection of nucleosome positioning to specific region of the genome (Carey M, 2005). However, MNase digestion combined with high-throughput sequencing method is now considered as strong technique to determine genome-wide nucleosome positioning. For that, in this study, MNase-Seq technology has been used to give insights into the location of nucleosomes at various levels of the chromatin and assess changes in the chromatin structure among different samples. Therefore, in this section, MNase-seq is described in detail.

4.2.1.1 Nuclei isolation and Micrococcal nuclease digestion

2 grams (g) of Arabidopsis rosettes were harvested, flash frozen in liquid nitrogen and ground to fine powder with a pestle and mortar. Next, the homogenized plant material was added to 10 millilitres (ml) of nuclei extraction buffer A (Table 4.1) in a 50 ml falcon tube and mixed well by vortex. Then, the obtained plant homogenate was filtered in 50 ml centrifuge tubes using a 70 micrometres (µm) nylon mesh placed in a funnel

Table 4.1. Nuclei extraction buffer A.

Reagent	Stock solution concentration	Final concentration
Sucrose	2 M	0.25 M
KCl	1 M	60 mM
MgCl2	1 M	15 mM
CaCl2	1 M	1 mM
PIPES	1 M	15 mM

Then, tubes were centrifuged at 10,000 x g for 20 minutes at 4°C. Afterwards, the supernatant was discarded by decantation and pellets were resuspended in 300 microliters (µl) of nuclei extraction buffer B (Table 4.2).

Table 4.2. Nuclei extraction buffer B.

Table 12. Reagent	Stock solution concentration	Final concentration
Sucrose	1 M	0. 25 M
Tris-HCl pH=8	1 M	10 mM
MgCl2	1 M	10 mM
Triton X-100	100%	1% V/V
B-Mercaptoethanol	1 M	5 mM
PMSF	1 M	1 mM

Next 300 µl of nuclei extraction buffer C (Table 4.3) were placed into an empty 2 ml eppendorf tube and layered by the pellet resuspended in nuclei extraction buffer B. Then, samples were centrifuged at 12,000 x g for 1 hour at 4°C.

Table 4.3. Nuclei extraction buffer C.

Reagent	Stock solution concentration	Final concentration
Sucrose	2 M	1.7 M
Tris-HCl pH=8	1 M	10 mM
MgCl2	1M	10 mM
Triton X-100	100%	0.50%
B-mercaptoethanol	1M	5 mM
PMSF	1 M	1 mM

Afterwards, supernatant were discarded by pipetting and pellets were mixed in 250 µl of MNase buffer (Table 4.4).

Table 4.4. MNase buffer.

Reagent	Stock solution concentration	Final concentration
Sucrose	2 M	0.3 M
Tris-HCl pH=7.5	1 M	20 mM
CaCl2	1M	3 mM

Then, DNA concentrations were measured using Qubit® DNA quantification kit (Table 4.5).

Table 4.5. Nuclei concentration.

Sample	Concentration (ug/ml)
Con 22°C Rep 1	13.2
Con 22°C Rep 2	10.4
Con 4°C Rep 1	7.8
Con 4°C Rep 2	5.8
Aza-dc 22°C Rep 1	13.3
Aza-dc 22°C Rep 2	8.6
Aza-dc 4°C Rep 1	15.9
Aza-dc 4°C Rep 2	7.10

Afterwards, around 350 nanograms (ng) of nuclei suspensions were incubated with 0.02 U/ul of MNase at 37 °C for 3 minutes followed by adding 40 ul of 2x stop buffer (Table 4.6) to stop the enzymatic reaction.

Table 4.6. 2X Stop buffer.

Reagent	Stock solution concentration	Final concentration
EDTA	0.5 M	0.05 mM
SDS	5%	1%

Then, 1X Proteinase K buffer (Table 4.7) and 1 μl of Proteinase K (stock 10 mg/ml) were mixed with the samples and incubated overnight at 37°C.

Table 4.7. 10X Proteinase buffer.

Reagent	Stock solution concentration	Final concentration
EDTA	0.5 M	50 mM
SDS	5%	5%
Tris-HCl pH=7.8	1 M	100 mM

4.2.1.2 Library preparation

To prepare nucleosome purified samples for next-generation sequencing on the Illumina platform, NEBNext® Ultra™ II DNA library kit Prep for Illumina® and ChIP-Seq was used for library preparation. The kit starting material requires around 500 picograms (pg)–1 micrograms (μg) into 50 μl of fragmented DNA which was obtained from step 4.2.1.1.

4.2.1.2.1 End Prep

First, the following components were added to a sterile nuclease-free tube: NEBNext Ultra II End Prep Enzyme Mix (3 μl), NEBNext Ultra II End Prep Reaction Buffer (7 μl) and 50 μl of fragmented DNA. Then, the mixture was pipetted up and down 10 times to mix thoroughly. Afterwards, a quick spin was performed to collect all liquid from the sides of the tube. Samples were then placed in a thermocycler, with the heated lid set to ≥ 75°C, and the following program was run: 30 minutes at 20°C-30 minutes at 65°C-Hold at 4°C for ∞.

4.2.1.2.2 Adaptor ligation

Since the DNA input samples were between 5-100 ng, adaptors were diluted in Tris/NaCl, pH 8.0 in a ratio of 1:10. Afterwards, the following mixture was prepared: End Prep Reaction Mixture (From step 4.2.1.2.1): 60 μl, NEBNext Ultra II Ligation Master Mix (30 μl), NEBNext Ligation Enhancer (1μl), NEBNext Adaptor for Illumina (2.5 μl). Then, a 200 μl pipette was set to 80 μl and the entire volume was pipetted up and down at least 10 times to mix thoroughly. Then, a quick spin was performed to collect all liquid from the sides of the tube. Afterwards, samples were incubated at 20°C for 15 minutes in a thermocycler with the heated lid off. Next, 3 μl of USER™ Enzyme was added to the ligation mixture. Finally, samples were mixed well and incubated at 37°C for 15 minutes with the heated lid set to 50°C.

4.2.1.2.3 Size selection of Adaptor-ligated DNA

Since the DNA starting material is greater than 50 ng, the adapter-ligated DNA was selected using SPRIselect beads. Beads were first resuspended by vortex, then 50 μl (~ 0.5x) of beads was added to the 96.5 μl ligation reaction obtained from step 4.2.1.2.2. Samples were then mixed well by pipetting up and down 10 times and incubated on bench top for 5 minutes at room temperature. Afterwards, tubes were placed on an appropriate magnetic stand to separate the beads from the supernatant. Once the solution is clear, the supernatant containing the DNA was carefully transferred to a new tube and beads that contain the unwanted large fragments were discarded. Next, 25 μl (0.25x) of resuspended SPRIselect beads was mixed with the supernatant 10 times and incubated on the bench top for 7 minutes at room temperature. Tubes were then placed on an appropriate magnetic stand to separate the beads from the supernatant. Once the solution is clear, beads containing the DNA were carefully transferred to a new tube and supernatants that contain the unwanted large fragments were discarded. Next, 200 μl of

80% freshly prepared ethanol were added to the tubes placed on the magnetic stand and incubated at room temperature for 30 seconds. Then, the supernatant was carefully removed and discarded without disturbing the beads that contain DNA targets. Beads were then air-dried for 5 minutes while the tubes are on the magnetic stand with the lid open. Afterwards, tubes were removed from the magnetic stand and the DNA target was eluted from the beads into 17 μl of 0.1X TE. After mixing, tubes were placed back on the magnetic stand. Finally, once the solution is clear, 15 μl of the clear solution containing DNA target was transferred to a new PCR tube.

4.2.1.2.4 PCR enrichment of Adaptor-ligated DNA

To a sterile strip tube, the following components were added: Adaptor Ligated DNA Fragments (From step 4.2.1.2.3): 15 μl, NEBNext Ultra II Q5 Master Mix: 25 μl, Index Primer/i7 Primer: 5 μl, Universal PCR Primer/i5 Primer: 5 μl. Afterwards, a 100 μl pipette was set to 40 μl and then samples were pipetted up and down 10 times to mix thoroughly. Then, a quick spin was performed to collect all liquid from the sides of the tube. Finally, the tubes were set on a thermocycler and a PCR amplification was performed using the following cycling conditions: 1 cycle of initial denaturation (98°C for 30 seconds), 12 cycles of denaturation Annealing/Extension (98°C for 10 seconds/ 65°C for 75 seconds), and final extension (65°C for 5 minutes), and final hold at 4°C for ∞.

4.2.1.2.5 Cleanup of PCR Reaction

First, 45 μl (0.9x) of resuspended beads were mixed to the PCR reaction by pipetting up and down at least 10 times. Then, samples were incubated on bench top for at least 5 minutes at room temperature. Next, tubes were placed on an appropriate magnetic stand to separate the beads from the supernatant. After the solution becomes clear, the supernatant was carefully removed and discarded, and 200 μl of 80% freshly prepared ethanol was added to the tubes while in the magnetic stand. Afterwards, samples were incubated at room temperature for 30 seconds and supernatants were carefully removed and discarded, and samples were placed back on the magnet to remove traces of ethanol with a p10 pipette tip. Next, the beads were air-dried for 5 minutes, while the tubes are on the magnetic stand with the lid open. Samples were then eluted by adding 33 μl of 0.1X TE and mixing it well by pipetting up and down 10 times. Then, tubes were incubated for at least 2 minutes at room temperature. Finally, tubes were placed on the magnetic stand and when the solution is clear, 30 μl of eluted DNA was transferred to a

new PCR tube and stored at –20°C. Finally, the size distribution of the library and its quantification were determined using the Agilent High Sensitivity DNA Assay at the Earlham institute, United Kingdom.

4.2.1.3 MNase sequencing procedure

The library pool was quantified using a KAPA Library Quant Kit (Roche Diagnostics Limited) before being diluted to 3 nM and spiked with 5% PhiX Control V3 (Illumina FC-110-3001). Then the library was denatured with NaOH and neutralised with Tris before addition of Illumina's ExAmp mix and loading onto the Illumina cBot at a final loading concentration of 300 pM. The flow cell was clustered using a HiSeq 4000 PE Cluster Kit (Illumina, PE-410-1001), utilising the Illumina HiSeq_3000_4000_HD_Exclusion_Amp_v1.0 method on the Illumina cBot. Following clustering, the patterned flow cell was loaded onto the Illumina HiSeq 4000 instrument following the manufacturer's instructions. Each paired sequencing read was 76bp long. The sequencing chemistry used was HiSeq 4000 SBS Kit (Illumina, FC-410-1003) with HiSeq Control Software 3.3.52 and RTA 2.7.3. Reads in bcl format were converted to FASTQ format by bcl2fastq2 (Illumina). Below a table summarising the number of reads generated from each sample (Table 4.8)

Table 4.8. Mnase sequencing reads information generated from all samples.

SampleName	Number of Reads	Mean Q30 to base Read 1	Mean Q30 to base Read 2
AzadC_22°C_R1	34,776,948	75	75
AzadC_22°C_R2	43,411,620	75	75
AzadC_4°C_R1	37,630,623	75	75
AzadC_4°C_R2	3,312,008	75	75
Ctrl_22°C_R1	84,714,794	75	75
Ctrl_22°C_R2	56,717,599	75	75
Ctrl_4°C_R1	55,197,020	75	75
Ctrl_4°C_R2	64,459,223	75	75

4.2.1.4 Bioinformatics analysis of MNase-Sequencing data

MNase-seq data analysis has been performed using a combination of Linux and python command lines. To begin with, FASTAQ files obtained from Earlham sequencing facility have been quality controlled (QC) using Fastqc version 0.11.8 and trimmed using trimmomatic version 0.32 with default parameters. Then, Bowtie version 1 has been used to map MNase-seq reads to Arabidopsis thaliana TAIR.10 reference genome, with the option "-m 1" on to obtain uniquely mapped reads only (Reads with mapping quality score equal to 255) in Sequence Alignment Map (SAM) file format. Afterwards, Samtools version has been used to convert mapping file in SAM format into Binary Alignment Map (BAM) file format. Since the obtained BAM files contains alignments in random order with respect to their position in the reference genome, samtools "view" has been used to order the alignments based upon the order of chromosomal coordinates. Then, "bamToBed" module in Bedtools version 2.29.0 has been used to convert the sorted BAM file to BEDPE format. Finally, the chromosome, starting coordinate, ending coordinate for each sequencing read in BEDPE file format has been extracted to build a 3-column BED file for detection of nucleosome positioning. Given that each biological replicate was sequenced independently and has different data quality (different fastqc evaluation, different mapping ratio, different unique mapping ratio), the data quality of each biological replicate has been assessed at the FASTQ file stage, then the BED files of biological replicates were merged before nucleosome positioning detection.

4.2.1.4.1 Detection of genome-wide nucleosome positioning using improved nucleosome positioning algorithm

Zhang et al. (Y. Zhang, Shin, Song, Lei, & Shirley, 2008) have developed NPS algorithm to detect nucleosomes from sequencing of MNase digested DNA fragments. However, the accuracy of NPS has been criticized by Chen et al., who argued that the algorithm couldn't detect nucleosome which are very visible even after adjusting or eliminating the thresholds for all the filtering steps (Weizhong Chen et al., 2014). To solve NPS defects, Chen et al. identified the technical problems of the algorithm resulting in mis-detection of nucleosomes and created an improved nucleosome positioning algorithm (iNPS), which combines the theoretical core algorithm of NPS and the resolution of its technical problems (Weizhong Chen et al., 2014). iNPS algorithm believes that accurate nucleosome positioning is based on generating a nucleosome sequencing profile that is able to automatically display nucleosome distribution in

a wave-form, which helps determining the peaks of the wave-form profile. Compared to NPS and to current nucleosome detection algorithms, iNPS has proven its reliability and high performance in detecting higher quality, a lower false positive nucleosomes, which are strongly associated with diverse biological terms (Y. Zhang et al., 2008). Due to its high performance and accuracy, iNPS has been used in this study to detect nucleosome positioning (Supplementary table 8-12). Briefly, iNPS generates a wave-form nucleosome signal profile from tag coordinate bed data (nucleosome scoring), which are further smoothen using discrete Gaussian convolution. Three Gaussian derivatives are then performed to detect important sites on the smoothed wave-form profile (maximun/minimum-extremum points, inflection points, and most winding positions). A pair of inflection points form a "main" nucleosome if a maximun-extremum point falls between them, otherwise it would be classified as a "shoulder". Then, shoulder candidates are classified as independent nucleosomes or the dynamic shifting of the "main" neighbouring nucleosome. To avoid the effect of big nucleosomes on detecting the borders (inflection points) of small nucleosomes, inflection border adjustment is performed using inflection points on the mildly smoothed profile. Nucleosome with accurate inflection borders are then merged to form a "doublet" if they are extremely close with similar height. Low quality nucleosomes are filtered out based on six criteria implemented by the iNPS algorithm (Chen et al., 2014). The confidence level of detected nucleosome is calculated using both upper- and lower-tailed Poisson test; in which, the first test identifies tag enrichment within the peak region and the second identifies the tag depletion within the adjacent "valley" regions flanking the corresponding nucleosome. This results in two respective scores '–log10(P-value_of_peak)' and '–log10(P-value_of_valley)' for each detected nucleosomes.

4.2.1.4.2 Illustration of nucleosome profiles around genomic features

For each chromosome, iNPS outputs two results files: *.like_bed and *.like_wig which are used to illustrate the results of nucleosome positioning. The *.like_bed file records the position information for each detected nucleosome. In which, column 1-4 (representing the chromosome, start, end, and index number of each nucleosome) are similar to UCSC BED file, and can be used to illustrate the position of each detected nucleosome you could extract the first 3 or 4 columns to build a 3-column or 4-column BED file, then illustrate the BED file by using IGV (a genome browser software). The *.like_wig contains 7 columns, where column 1 is the coordinate and column 2-7 are signal profile columns. This file can be used with Microsoft Excel or with IGV. In the first case, the nucleosome profile of the genomic region

of interest can be illustrated by extracting the genomic coordinate information in column 1 and any another signal profile column from column 2-7, which is then illustrated using "Scatter with Smooth Lines" from Excel. In the second case, column 1 (coordinates) can be extracted as well as another signal profile column from column 2-7 to build a UCSC wiggle file to illustrate the profile by using IGV. For example, by using column 1 and 2 to build wiggle file, the raw profiles of nucleosome signal can be illustrated, and by using column 1 and 3, the smoothed profiles of nucleosome signal can be illustrated, whereas, using column 1 and 7 can help illustrate the profiles of detected nucleosome peaks.

Based on this, for each sample, a wiggle file containing column 1 and 7 was generated using Linux command lines. Then, around transcription start and end sites (TSSs and TESs; respectively) coordinates information of transcription sites were extracted from Arabidopsis TAIR.10 annotation file to build a 4-Column BED file containing (chromosome, TSS, TES, and strand information). The resulting 4-Column BED file and the wiggle files were then used by deepTools3 to calculate two matrices for each sample (One for TSSs and other TESs) using "computeMatrix" module; containing the average nucleosome profile peaks within +/- 1000 base pairs (bp) around TSSs and TESs. Then, "PlotProfile" module from deepTools3 was used to illustrate the profiles of nucleosome peaks from the calculated matrices.

In a similar manner, to plot nucleosome profiles around 3' and 5' splice sites (SS), nucleosome profiles within -500/+500 bp around 3'SS and 5'SS respectively were collected using "computeMatrix" module of deepTools3, then nucleosome profiles around 3'SS and 5'SS were illustrated using the "PlotProfile" module from the same tool. The beginning/ending coordinates of exons (except TSS/TTS) were considered as the splice sites coordinates, taking into consideration the DNA strand of each gene. If a gene is located at positive strand, the beginning/ending coordinates of exons are 3'/5' SS correspondingly, while if a gene is located at negative strand, the beginning/ending coordinates of exons are 5'/3' splicing sites correspondingly. An additional matrix was computed using "computeMatrix" module of deepTools3 for each sample to plot nucleosome profiles within -500/+500 bp around exons; for which the coordinates were the same as the ones used to determine splice sites coordinates. Then, nucleosome profiles within -500/+500 bp around exons were illustrated using the "PlotProfile" module from deepTools3.

Furthermore, to plot nucleosome profiles around each group of genes based on their expression, from the RNA-Seq data, genes were grouped based on their TPM values generated by Salmon

(five groups). Then, the average nucleosome profiles around TSSs and TESs for each group of gene (for each sample) were collected using "computeMatrix" module of deepTools3. Then nucleosome profiles around TSSs and TEs for each group of gene were illustrated using the "PlotProfile" module from the same tool.

Moreover, to plot nucleosome profiles for different AS event, coordinate start and end for each exon (For A3'SS, A5'SS, and ES events) and intron (For IR events) of each sample were collected from SUPPA, then nucleosome profiles within -200/+200 bp around exons and introns involved in each AS events were collected using "computeMatrix" module of deepTools3. Then, nucleosome profiles within -200/+200 bp around exons involved in A3'SS, A5'SS, and ES AS events and introns for IR events were illustrated using the "PlotProfile" module from the same tool. Similarly, nucleosome profiles were aligned to the 3'SS of exons and introns involved in different AS events grouped according to their Percent Spliced in (PSI) index, and flanking sequences. First, AS events obtained from SUPPA in each sample were grouped into four groups based on their PSI values (PSI ≤ 20%, 20% <PSI≤50%, 50%<PSI ≤80%, PSI ≥80%). Then for each group in each AS event, coordinate start and end of each exon (For A3'SS, A5'SS, and ES events) and intron (For IR events) of each sample were collected. Then nucleosome profiles within -200/+200 bp around 3'SS of exons involved in A3'SS, A5'SS, and ES AS events and introns involved in IR were collected using "computeMatrix" module of deepTools3. Then, nucleosome profiles within -200/+200 bp around exons involved in A3'SS, A5'SS, and ES AS events and introns for involved in IR events were illustrated using the "PlotProfile" module from the same tool.

Finally, to plot nucleosome profiles across specific AS events, the coordinates involved in the AS events were extract from SUPPA output. Then, the nucleosome signal level corresponding to those coordinates were extracted from the 7th column of *.like_wig file. Afterwards, the nucleosome profiles for each event were presented using "Scatter with Smooth Lines" option from Excel.

4.2.1.4.3 Differential nucleosome positioning analysis

Since iNPS algorithm is specific to determine nucleosome positioning rather than detecting differentially positioning nucleosomes (DPNs), DANPOS version 2.1.2 was compulsory algorithmic module to integrate in the analysis presented in this chapter to obtain DPNs. Due to the accuracy and specificity of iNPS in determining nucleosome positioning compared to

other algorithm, including DANPOS, only tags contributing to the iNPS-detected nucleosome were selected for DNPs analysis. Therefore, from the 3-column BED file obtained in section 4.2.1.4, the following formula has been applied to obtain the mid-point of each tag: (Chromosome start + chromosome end)/2. Then, after referring to the *like_bed file for detected nucleosome peaks obtained in section 4.2.1.4 (In which the column 2 and 3 are the start and end coordinates of each nucleosome peak), only tags having their mid-point locating with any nucleosome peak, the corresponding tag (chromosome, start, end) were selected for DNPs analysis. Afterwards, DANPOS was run with the parameters '-q,--height'=1 (the intensity cutoff for nucleosome calling), '-z,--smooth_width'=100 (the smooth width before peak calling), '-e,--edge'=1 (detect edges for peaks), '-k,--keep'=1 (saving mid-stage files), '-x,--pcfer'=0 (no nucleosome calling), '-n,--nor'=N (no normalization), '--frsz'=150 (setting the average size of DNA fragment to 150 bp) and '--clonalcut'=0 (don't adjust clonal signal). DANPOS scores the difference of nucleosome signal between two samples of each contrast group using pvalues and false positive rates (FDRs); hence, significant differentially positioned nucleosomes were selected only if 'point_diff_FDR'≤0.01 and 'smt_diff_FDR'≤0.05. Then, a 2,000-bp sliding window was moved across the genome with a 500 bp step size to select the windows enriched with differentially positioned nucleosomes (~ top 1% windows that have ≥2 differentially positioned nucleosomes are selected). Plant Biomart was then used to identify genes associated with DPNs in selected windows.

4.2.2 Genomic DNA extraction and bisulphite treatment

To determine DNA methylation status, selected samples were treated with bisulphite followed by standard Illumina 100 bp paired-end reads sequencing. After the discovery of CpG methylation sites, a class of restriction enzymes dependent on methylation was developed to cleave the methylated sites in the genome and their cleavage activity can be blocked if a specific base is modified (li *et al.* 2011). Although this method is easy to use, the data obtained from these techniques is limited by the presence of CpG sites in the studied sequence and requires large amounts of genomic DNA (li *et al.* 2011). Therefore, bisulfite DNA sequencing discovered by Frommer et al. was regarded as the most accurate and sensitive technology for DNA methylation analysis that generates quantitative accuracy for a wide spectrum of samples handling (Formmer *et al.* 1992). This qualitative and quantitative method is capable of identifying 5-methylcytosine (5mC) at a single base resolution. Frommer et al. developed this technique based on their findings that 5mC and unmethylated cytosines respond differently to sodium bisulphite (SB) treatment. Once the single-stranded DNA is treated with SB, the

unmethylated cytosine residues will be converted to uracil which is detected as thymine after PCR amplification followed by sequencing (Formmer *et al.* 1992). In this technique, PCR amplification is essential to determine the methylation patterns within a certain locus using specific methylation primers. Otherwise, genome-wide methylation patterns can be identified by sequencing the direct PCR product, sub-cloning sequencing or DNA Seq. Since our aim was to detect genome-wide changes of methylation patterns, WGBS was the most suitable technique to follow. Therefore, DNA was extracted using the DNeasy® Plant Mini Kit and treated as follow. First, leaf tissues (≤100 mg wet weight) were frozen with liquid nitrogen and disrupted using a mortar and pestle then, 400 µl Buffer AP1 and 4 µl RNase A was added to each sample, well vortex, and incubated for 10 minutes at 65°C. Next, 130 µl Buffer P3 was added to the samples and incubated for 5 minutes on ice after mixing. Samples were then centrifuged for 5 minutes at 20,000 x g (14,000 rpm). Afterwards, the lysate was pipetted into a QIAshredder spin column placed in a 2 ml collection tube and centrifuged for 2 minutes at 20,000 x g. Next, the flow-through was transferred into a new tube and 1.5 volumes of Buffer AW1 was added and mixed well by pipetting. Then, 650 µl of the mixture was transferred into a DNeasy Mini spin column placed in a 2 ml collection tube. Afterwads, samples were centrifuged for 1 min at ≥6000 x g (≥8000 rpm), the flowthrough was discarded, and the spin column was then placed into a new 2 ml collection tube. Next, 500 µl Buffer AW2 was then added to the samples and tubes were centrifuged for 1 min at ≥6000 x g. After centrifugation, the flowthrough was discarded and another 500 µl Buffer AW2 was added followed by tubes centrifugation for 2 minutes at 20,000 x g. Finally, the spin column was transferred to a new 1.5 ml microcentrifuge tube and 30 µl Buffer AE was added for elution. Samples were then incubated for 5 minutes at room temperature (~25°C) and tubes were centrifuged for 1 minute at ≥6000 x g. Finally, purified DNA was sent to bisulfite treatment, library preparation, and sequencing at the Earlham Institute.

4.2.2.1 Library preparation

For WGBS, libraries were prepared for 6 samples using a KAPA high throughout Library Prep Kit -with amendments (Part No: KK8234). This kit has been optimised for 1µg of input DNA with a size selection using Beckman Coulter XP beads (Part No: A63880). The DNA was QC'd with a High Sensitivity Qubit assay (part No: Q32854) and 1 µg of each sample was taken forward for processing. The DNA was sheared using the Covaris LE220 sonicator (Covaris) to an average size of 350 bp. Methylated barcoded adapters (NEXTFlex bisulfite barcodes (BiooScientific _ 511913)) weredded to the treated ends of sheared DNA. Bisulfite Conversion

122

of DNA library was performed using EZ DNA Methylation-Gold™ Kit. Treating DNA with bisulfite chemically modifies non-methylated cytosines into uracil, methylated cytosines remain unchanged. After PCR enrichment, the insert size of the libraries was verified by running an aliquot of the DNA library on a PerkinElmer GX using the High Sensitivity DNA chip (Part No: 5067-4626) and the concentration determined by using a High Sensitivity Qubit assay.Then, 12-plex equimolar pool was prepared ready for q-PCR and sequencing on the HiSeq 4000 using v1 chemistry and 150 bp paired-end reads over 2 lanes.

4.2.2.1 Whole genome bisulphite sequencing procedure

The constructed WGBS libraries were normalised and equimolar pooled, the final pool was quantified using a KAPA Library Quant Kit (Roche Diagnostics Limited) and found to be 10.11 nM. The library pool was diluted to 3 nanomolars (nM) and spiked with 10% PhiX Control V3 (Illumina FC-110-3001). Then, DNA was denatured with NaOH and neutralised with Trisbuffer before addition of Illumina's ExAmp mix and loading onto the Illumina cBot, to give a final loading concentration of 300 pM. The flow cell was clustered using a HiSeq 4000 PE Cluster Kit (Illumina, PE-410-1001), utilising the Illumina HiSeq_3000_4000_HD_Exclusion_Amp_v1.0 method on the Illumina cBot. Following clustering, the patterned flow cell was loaded onto the Illumina HiSeq4000 instrument following the manufacturer's instructions. Each paired sequencing read was 151bp long. The sequencing chemistry used was HiSeq 4000 SBS Kit (Illumina, FC-410-1003) with HiSeq Control Software 3.3.52 and RTA 2.7.3. Reads in bcl format were converted to FASTQ format by bcl2fastq2 (Illumina). The total number of raw reads generated in the WGBS-seq data was ~ 23 M per biological replicate (Table 4.9).

Table 4.9. WGSB sequencing reads information generated from AzadC samples grown at 22°C and 4°C.

Sample name	Number of reads	Mean Q30 to base Read 1	Mean Q30 to base Read 2
AzadC_22°C_R1	30,719,706	151	134
AzadC_22°C_R2	23,899,768	151	134
AzadC_22°C_R3	22,016,740	151	134

AzadC_4°C_R1	19,396,984	151	129
AzadC_4°C_R2	22,634,063	151	134
AzadC_4°C_R3	24,493,759	151	139

4.2.2.2 Bioinformatics analysis of whole genome bisulphite data

WGBS data analysis has been performed using a combination of Linux, Python, Perl, and R command lines. To begin with, FASTAQ files obtained from Earlham sequencing facility have been QC'd using Fastqc version 0.11.8 and trimmed using trim glore version 0.5. Trim galore has been chosen due to its specificity in filtering low quality reads (Prehed score less than 20) and trimming adapter sequences from WGBS data. Trim glore offers the option of trimming the first 13 bp of Illumina standard adapters ('AGATCGGAAGAGC') by default, in addition to trimming the first bp from the 3' end of all reads to avoid problems with invalid alignments of completely overlapping long reads and hence incorrect methylation calls. Additionally, trim glore has the option to trim a fixed amount of bases from the 5' end of reads, which might be helpful in the case bisulfite-Seq paired-end library where the end repair procedure introduces unmethylated cytosines.

4.2.2.3.1 Alignment of sequencing reads to Arabidopsis reference genome

After ensuring the removal of unwanted sequencing reads with low prehed score quality and adapter contamination, the next step is to perform mapping to Arabidopsis TAIR.10 reference genome. For this step, Bismark version v0.15.0 was the best tool of choice due to its capability of supporting alignments of bisulphite-treated reads. First, Bismark performs fully bisulfite conversion of sequence reads, where each C in the forward sequence read is transformed to T and each G in the reverse forward read into A, before they are aligned to similarly converted versions of the genome. To infer the methylation of each cytosine in the genome, Bismark search for the best alignment (out the four alignments running against the genome in parallel) to compare them afterwards to the normal genomic sequence. Upon alignments completion, Bismark generates a BAM or SAM file which can be processed for further analysis, in addition to a run report containing a summary of alignments parameters in addition to percentage of methylation cytosine in each context CpG, CHG or CHH context (where H can be either A, T or C). In which, the percentage is calculated individually for each context following the

equation: % methylation (context) = 100 * methylated Cs (context) / (methylated Cs (context) + unmethylated Cs (context)). Bismark has also been the best tool of choice given its capability of offering directional or non-directional mode alignments mode, depending on bisulphite treated library preparation. In a directional library, complementary bisulphite converted strands to the original top (CTOT) and original bottom (CTOB) are generated during BS-PCR step, in addition to the original top (OT) and original bottom (OB) strands. Yet, during sequencing of this library type, CTOT and CTOB are not taken into consideration as they are ligated to the wrong kind of adapter at their 5' end. Therefore during Bismark alignment step, if the option "--directional" is specified, Bismark only take into consideration the OT and OB strands given that CTOT and CTOB should not be present theoretically in the BS-Seq library. Alternatively, in non-directional library, all four strands (OT, OB, CTOT, and CTOB) are constructed in a way where they can serve as valid reads for alignments. Subsequently, specifying the option "—non-directional" instructs Bismark to use all four strands during the alignment stage. Given these information, in this study, Bismark alignment for the WGBS data has been performed as follows. First, Arabidopsis reference genome has been indexed using the module "—bowtie2", Then, as described previously in this section, the reference genome needs to be converted (C->T and G->A versions) using the "bismark_genome_preparation" module form Bismark. Finally, and since the library used in this study is directional and paired-ended, Bismark alignment has been run with the "--directional" and "—paired-ended" options on. Although Bismark offers the option to perform extracting methylation information in the three methylation contexts (CpG, CHG or CHH context; where H can be either A, T or C), in this analysis, the option "--bismark_methylation_extractor" has been turned off and the methylation the methylation call for every single C analysed in the three contexts has been performed using MethylKit v1.11.0 as described below.

4.2.2.3.2 Extracting methylation calls from Bismark alignments

After obtaining SAM alignments from bismark, SAM files were sorted by chromosome and read position columns, using the "sort" module from Samtools. Sorted SAM files were then processed as follows MethylKit v1.11.0 (Akalin et al., 2012). First, methylation calls has been read using "methylRaw" option from Methylkit in the three methylation context and the methylation call files have been saved separately for each. During this step, two groups were defined by treatment vector; where AzadC grown at 22°C and treated by cold stress were taken as cold and treatment vectors, respectively. Then "getMethylationStats" has been used to obtain the histogram for percent methylation distribution. Typically, percent methylation histogram

should present two peaks on both ends. In any given cell, bases are either methylated or not. Therefore, investigating the methylation status from complex tissue should yield a methylation pattern where the genome presents a combination of methylated and unmethylated regions (Figure 4.1-4.3). Additionally, coverage per base information can be plotted using "getCoverageStats" from Methylkit. The histogram presents bars alongside numbers; representing the percentage of locations contained in a certain bin. The histograms presented here, show that the experiments doesn't suffer from PCR duplication bias, where no secondary peak has been detected towards the right hand side of the histogram (Figure 4.1-4.3). Furthermore, bases that have coverage below 10X and more than 99.9th percentile were discarded using "filterByCoverage" function from Methylkit also offers the option to obtain bisulphite conversion statistics for different biological replicates as indicated in table 4.10 (Akalin et al., 2012).

A
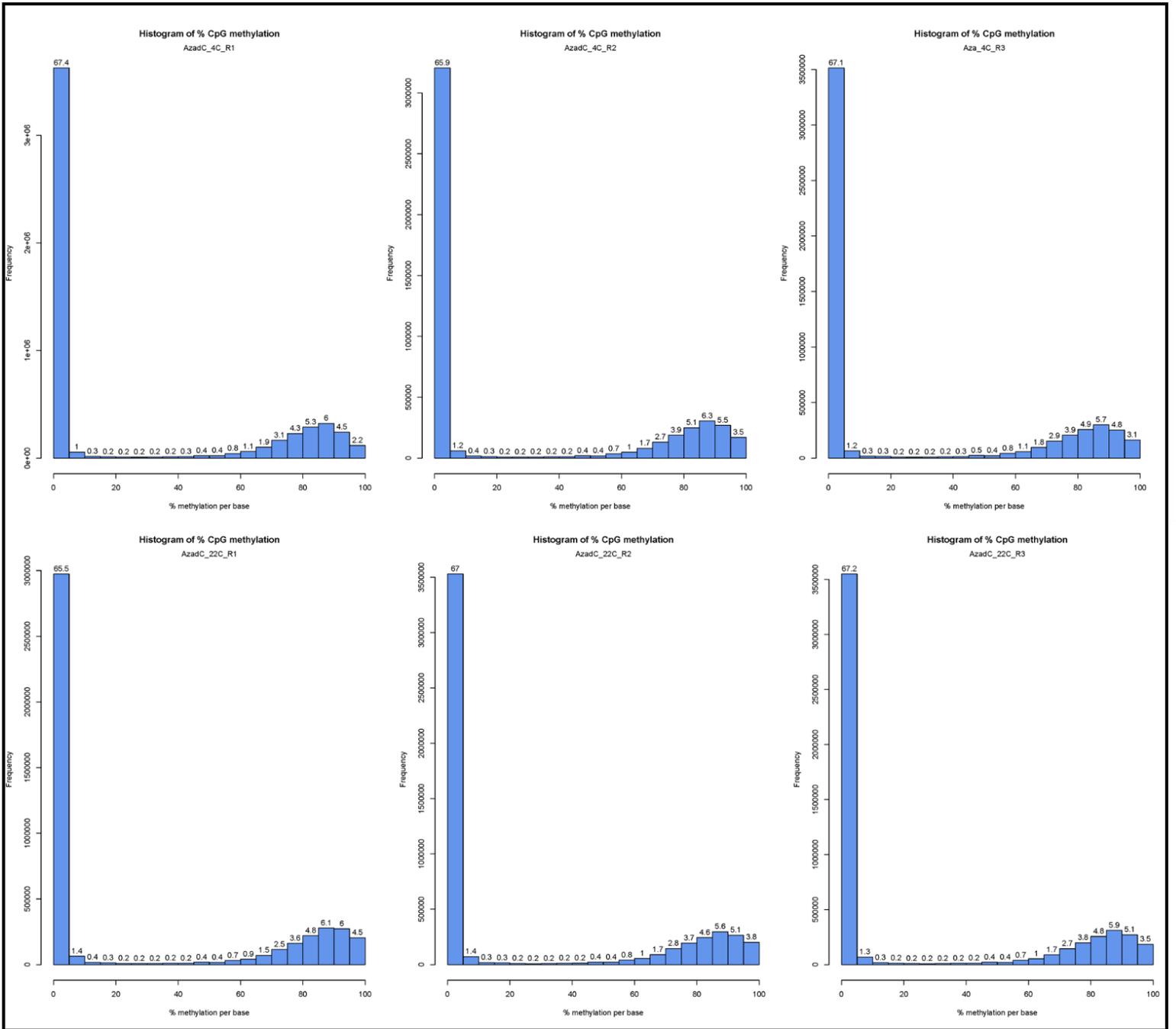
B
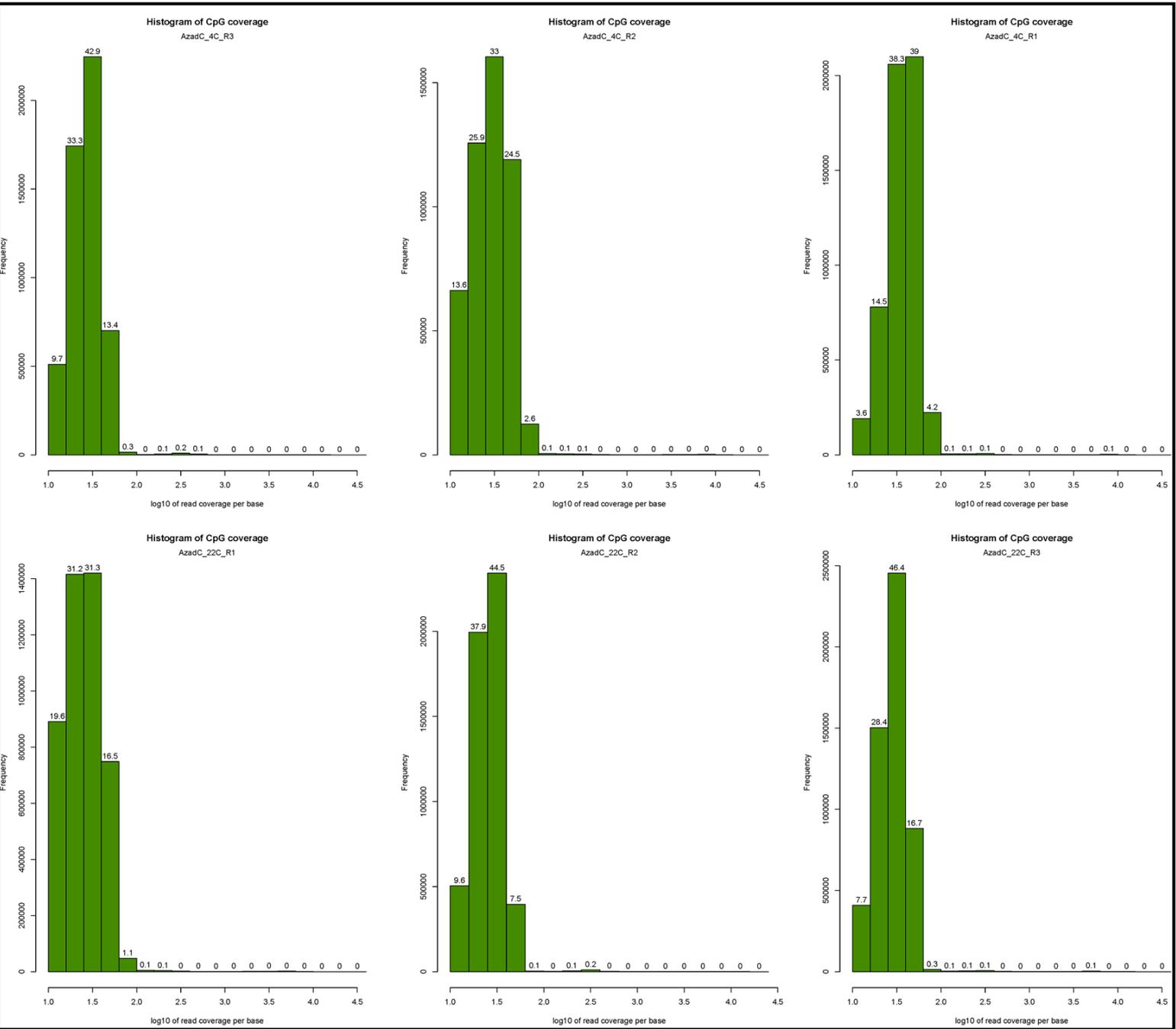
Figure 4.1. Histogram of CpG methylation percentage (A) and coverage (B) of different samples. (A) The x-axis represents methylation percentage per base whereas, the y-axis represents the percentage of location contained in the corresponding bin. **(B)** The x-axis represents the log10 read coverage per base, whereas the y-axis represents the read coverage location contained in the corresponding bin.

A

Histogram of % CHG methylation

B
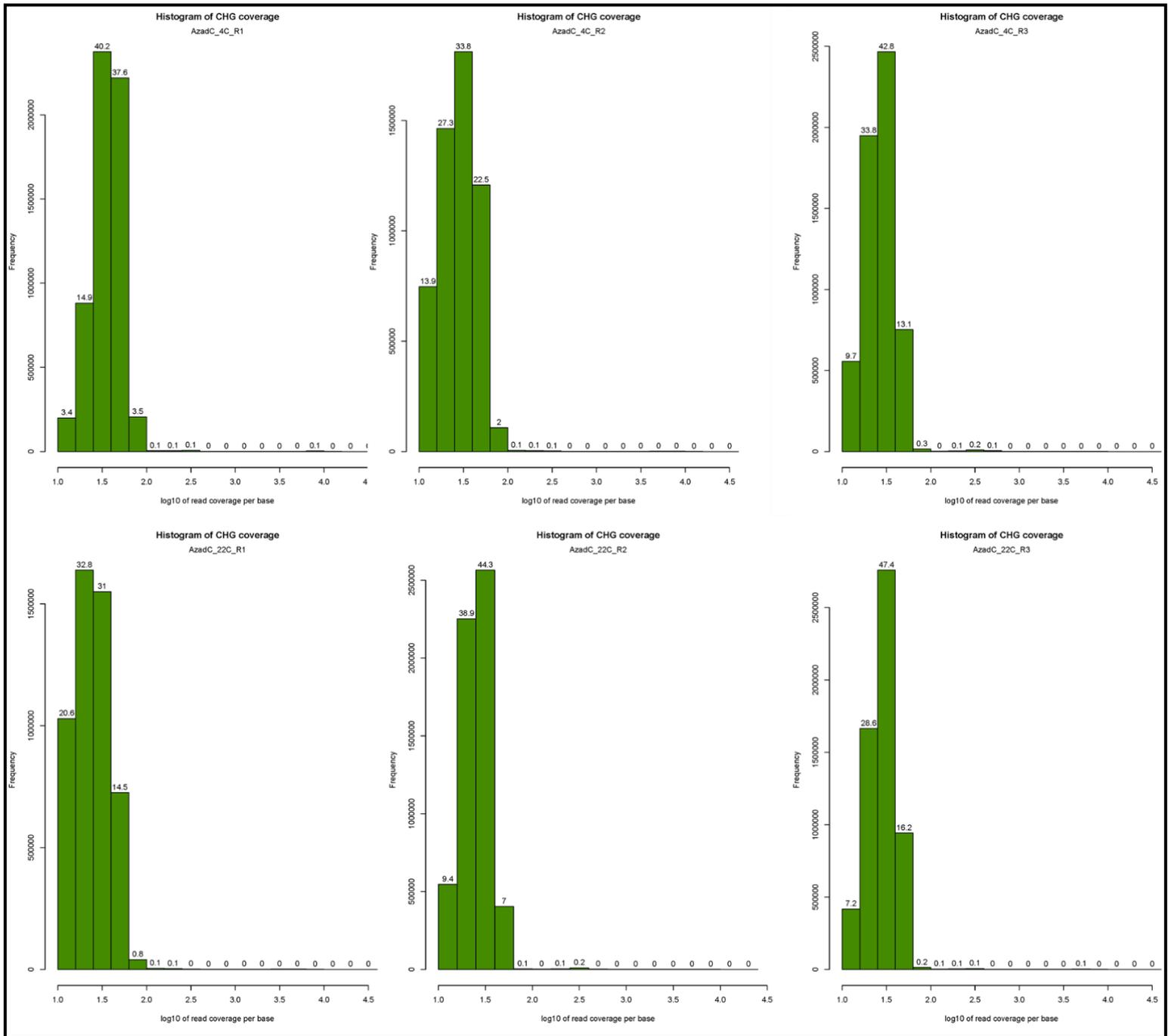
Figure 4.2. Histogram of CHG methylation percentage (A) and coverage (B) of different samples. (A) The x-axis represents methylation percentage per base whereas, the y-axis represents the percentage of location contained in the corresponding bin. **(B)** The x-axis represents the log10 read coverage per base, whereas the y-axis represents the read coverage location contained in the corresponding bin.

A

Histogram of % CHH methylation
AzadC_4C_R1

Histogram of % CHH methylation
AzadC_4C_R2

Histogram of % CHH methylation
AzadC_4C_R3

Histogram of % CHH methylation
AzadC_22C_R1

Histogram of % CHH methylation
AzadC_22C_R2

Histogram of % CHH methylation
AzadC_22C_R3
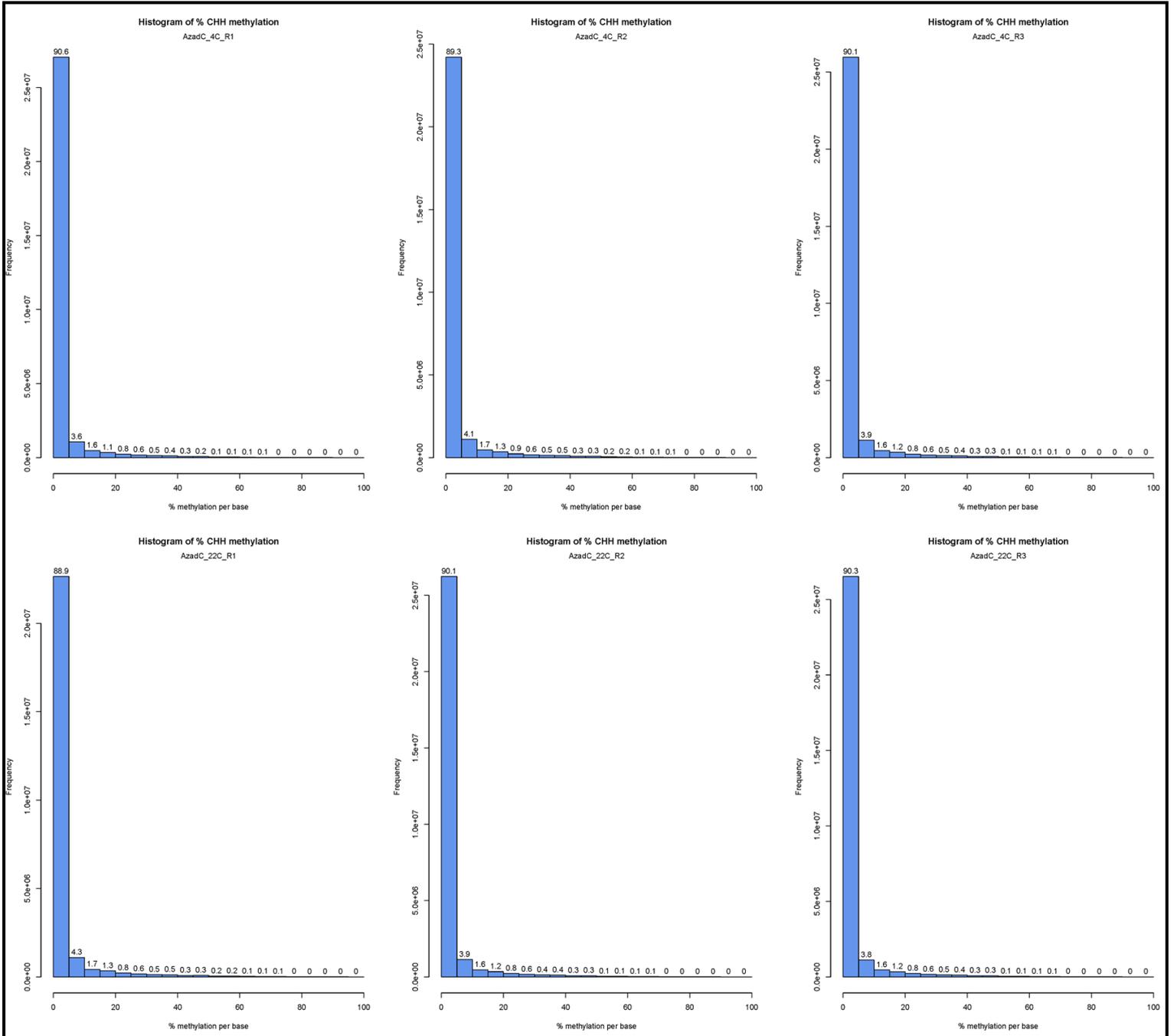
B
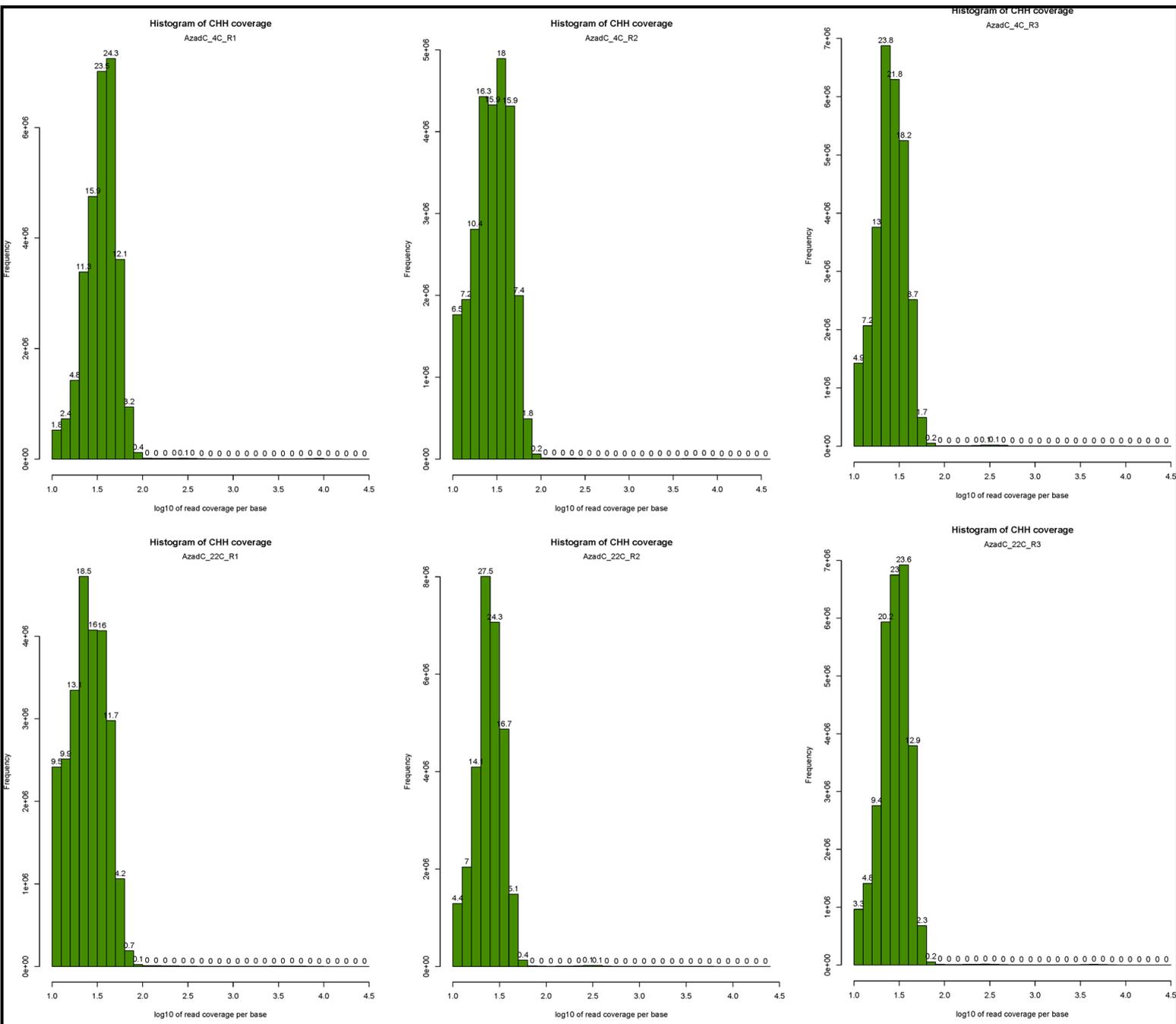
Figure 4.3. Histogram of CHG methylation percentage (A) and coverage (B) of different samples. (A) The x-axis represents methylation percentage per base whereas, the y-axis represents the percentage of location contained in the corresponding bin. (B) The x-axis represents the log10 read coverage per base, whereas the y-axis represents the read coverage location contained in the corresponding bin.

Table 4.10 Bisulphite conversion statistics reported by MethylKit v1.10.0 for AzadC 22°C and AzadC 4°C biological replicates.

	AzadC 22°C- Replicate1	AzadC 22°C- Replicate2	AzadC 22°C- Replicate3	AzadC 22°C- Replicate1	AzadC 22°C- Replicate2	AzadC 22°C- Replicate3
Total otherC considered (>95% C+T)	30195363	34895655	35194670	35768103	32484132	34583763
Average conversion rate	96.478933 47	96.940571 02	96.939308 97	96.956865 53	96.568533 07	96.893922 67
Total otherC considered (Forward) (>95% C+T)	15107657	17444680	17598330	17886008	16246344	17291960
Average conversion rate (Forward)	96.472534 32	96.940557 15	96.936821 13	96.955747 23	96.566399 6	96.893436 53
Total otherC considered (Reverse) (>95% C+T)	15087706	17450975	17596340	17882095	16237788	17291803
Average conversion rate (Reverse)	96.485341 07	96.940584 89	96.941797 1	96.957984 07	96.570667 67	96.894408 81

4.2.2.3.3 Identification and annotation of differentially methylated regions

After obtaining good quality descriptive statistics for each sample and filtering out reads that don't display appropriate read coverage, the next step is to identify differentially methylated regions (DMRs). To begin with, for each methylation context, the bases coverage for all samples were merged into one object using "unite" function from MethylKit, which is an essential step to start a comparative analysis between the biological samples. Hence, the resulting object contains methylation information for regions that are covered in all samples. Since each sample was sequenced in three biological replicate, only methylation contexts covered with at least 2 samples per group will be returned using (min.per.group) parameter available from the "unite" function of MethylKit. Given the purpose of obtaining DMRs in tilling windows rather than single bas-pair resolution, the function "min.per.group" from methylKit has been used to summarize methylation information over 1000 bp windows with a step size of 1000 bp. The tilling function sums up C and T counts from each covered cytosine and returns a total C and T count for each tile. Afterwards, DMRs were calculated using the "calculateDiffMeth" function from MethylKit. Since the dataset contains biological replicates, calculation of DMRs is automatically performed by Methylkit using logistic regression to

calculate P-values. Based on the formula below, the logistic regression model uses πi as a methylation proportion to model the log odds ratios. This is performed through the treatment vector which denotes the sample group membership for the methylation context in the model. The "Treatment" variable is used to predict the log-odds ratio of methylation proportions (Akalin et al., 2012).

Logistic regression formula: $\log\left(\dfrac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 Treatment_i$

P-values were then adjusted to q-values using the SLIM method (Akalin et al., 2012). Afterwards, for the three methylation contexts, only DMRs with q-value less than 0.05 and a methylation difference higher than 5% were selected using "getMethylDiff" function form MethylKit. Hyper-methylated and Hypo-methylated regions were obtained using (type) parameter of "getMethylDiff" function form MethylKit (Akalin et al., 2012).

To annotate DMRs, Genomation package version 1.16.0 was used in parallel with MethylKit. Given that Genomation requires GRanges objects, DMRs object obtained by Methylkit were converted to Granges objects. Then, Arabidopsis genome annotation was read from a BED file containing annotation information using "readTranscriptFeatures" function from Genomation. This function will return a Granges list containing introns, exons, TSS, and promoter coordinates. Next, annotation of DMRs was performed using the "annotateWithGeneParts" function from Genomation. The following annotation will return a Granges object containing the percentage of target features overlapping with annotation, which can be displayed in a histogram form as shown in here in the results section

4.2.2.3.4 Illustrating methylation percentage across genomic features

For each sample, the methylation call files generated by MethylKit were used to output a wiggle file for each sample containing methylation percentage for each base pair calculated as C/(C+T) from the first nucleotide of both strands. For each sample, three wiggle files corresponding to each methylation context (CpG, CHH, CHG) were generated.
To plot methylation percentage around 3'SS and 5'SS, all SJs were stacked (100 bp exon + 100 bp intron for the donor, 100 bp intron + 100 bp exon for the acceptor) and methylation percentage in the three methylation contexts around these regions were collected using "computeMatrix" module of deepTools3, then methylation profiles around 3'SS and 5'SS were

illustrated using the "PlotProfile" module from the same tool. The beginning/ending coordinates of exons (except TSS/TTS) were considered as the splice sites coordinates, taking into consideration the DNA strand of each gene. If a gene is located at positive strand, the beginning/ending coordinates of exons are 3'/5' SS correspondingly, while if a gene is located at negative strand, the beginning/ending coordinates of exons are 5'/3' splicing sites correspondingly. An additional matrix was computed using "computeMatrix" module of deepTools3 for each sample to DNA methylation percentage within -500/+500 bp around exons; for which the coordinates were the same as the ones used to determine splice sites coordinates. Then, nucleosome profiles within -500/+500 bp around exons were illustrated using the "PlotProfile" module from deepTools3.

Furthermore, to plot CpG methylation percentage around each group of genes based on their expression, from the RNA-Seq data, genes were grouped based on their TPM values generated by Salmon (five groups). Then, the average methylation percentage around TSSs and TESs for each group of gene (for each sample) were collected using "computeMatrix" module of deepTools3. Then nucleosome profiles around TSSs and TEs for each group of gene were illustrated using the "PlotProfile" module from the same tool.

Finally, to plot CpG methylation percentage for different AS event, coordinate start and end for each exon (For A3'SS, A5'SS, and ES events) and intron (For IR events) of each sample were collected from SUPPA, then methylation percentage within -200/+200 bp around exons and introns involved in each AS events were collected using "computeMatrix" module of deepTools3. Next, DNA CpG methylation percentage within -200/+200 bp around exons involved in A3'SS, A5'SS, and ES AS events and introns for IR events were illustrated using the "PlotProfile" module from the same tool. Similarly, methylation percentage were aligned to the 3'SS of exons and introns involved in different AS events grouped according to their PSI index, and flanking sequences. First, AS events obtained from SUPPA in each sample were grouped into four groups based on their PSI values (PSI ≤ 20%, 20% <PSI≤50%, 50%<PSI ≤80%, PSI ≥80%). Then for each group in each AS event, coordinate start and end of each exon (For A3'SS, A5'SS, and ES events) and intron (For IR events) of each sample were collected. Then methylation percentage within -200/+200 bp around 3'SS of exons involved in A3'SS, A5'SS, and ES AS events and introns involved in IR were collected using "computeMatrix" module of deepTools3. Then, methylation percentage within -200/+200 bp around exons involved in A3'SS, A5'SS, and ES AS events and introns for involved in IR events were illustrated using the "PlotProfile" module from the same tool.

4.3 Results

4.3.1 DNA methylation and nucleosome occupancy define intron and exon boundaries

In Arabidopsis, DNA methylation and nucleosome positioning have been found to differentially mark promoter regions, gene bodies as well as exons and introns, indicating a potential link of chromatin architecture to gene expression and splicing regulation (Chodavarapu et al., 2010). Since the RNA-seq data show genome-wide changes in gene expression and AS due to methylation differences, the next step was to investigate whether nucleosome occupancy and DNA methylation levels differentially mark promoter regions, exons, introns (in AzadC and Ctrl plants under normal and cold conditions). Towards this goal, MNase-seq of AzadC and Ctrl Arabidopsis plants grown at 22°C and subjected to cold stress has been used. Then, the distribution of nucleosome density in -2000/+2000 bp regions flanking the TSS and TTS has been analysed to show that nucleosome occupancy levels are significantly lower around the TSS and their flanking regions among all samples for five Arabidopsis chromosomes (Figure 4.4).
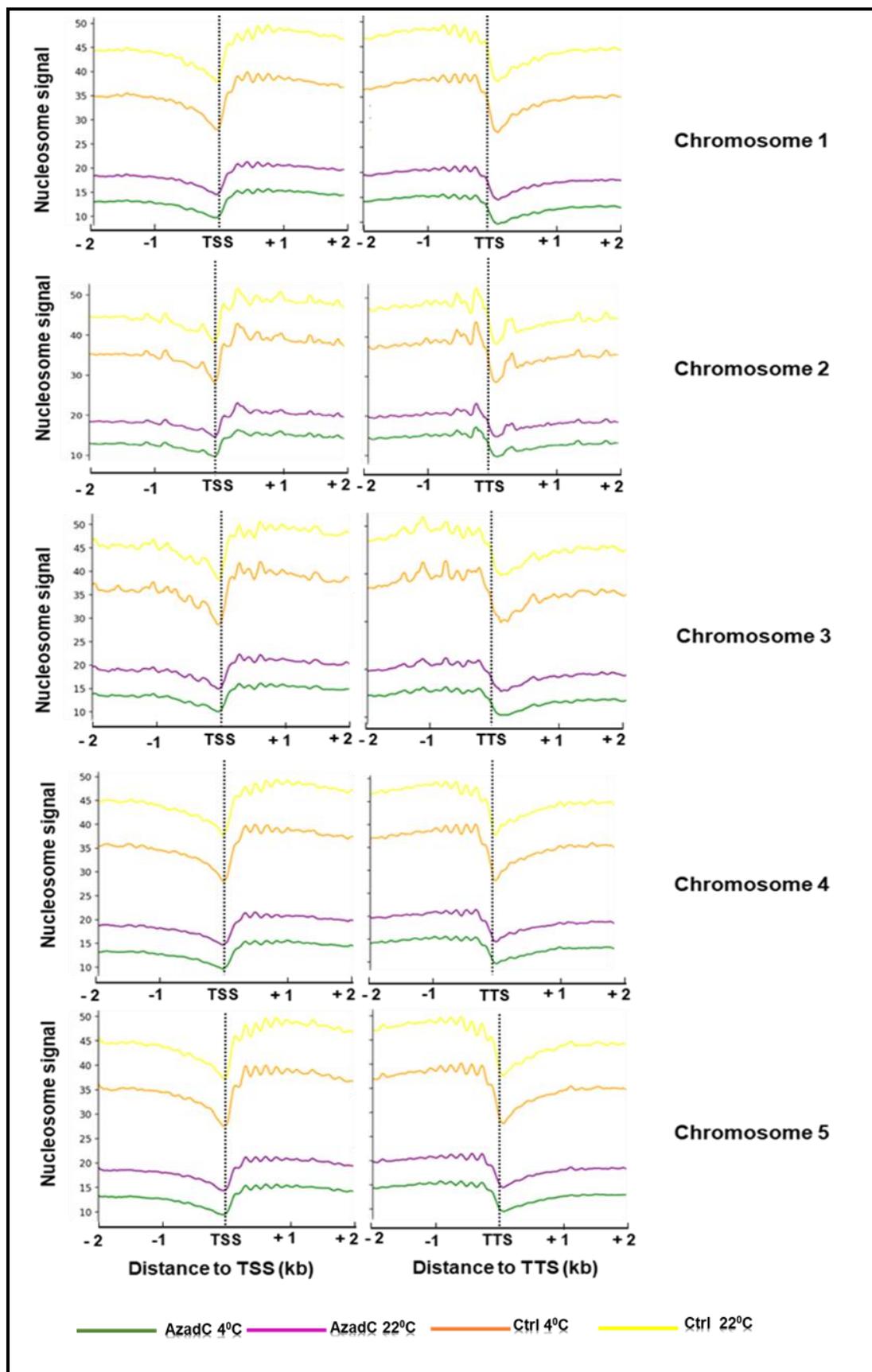
Figure 4.4. Nucleosome occupancy profiles in -2000/+2000 bp regions flanking the transcription start site (TSS, left) and transcription termination site (TTS, right) for Arabidopsis chromosomes. The x- axis represents the distance to TSS (kb); the y- axis represents the nucleosome signal level. Nucleosome occupancy profiles for all

samples and across all chromosomes, show a distinctive nucleosome occupancy patterns near the TSS and TTS that drops significantly in the surrounding regions. Cold stress induces lower nucleosome occupancy levels regardless of the treatment (orange line compared to yellow for Ctrl and green line compared to purple for AzadC). For both temperatures, AzadC nucleosomes levels around the TSS, TTS, and flanking regions remains lower than Ctrl plants (green and purple lines compared to orange and yellow).

Yet, it was clear that nucleosome occupancy levels are reduced in plants subjected to cold stress relative to plants grown at $22^{\circ}C$ and in AzadC compared to Ctrl. Then, global patterns of nucleosome occupancy over all internal exons and flanking regions has been profiled. A sharp peak of nucleosome occupancy is detected on exons, surrounded by regions of lower density in the flanking introns (Figure 4.5). Despite the similarity of nucleosome occupancy profile between different conditions, the level of nucleosome signal is affected by cold stress and upon DNA demethylation. Indeed, a significant decrease in nucleosome occupancy levels was detected in AzadC vs Ctrl plants with a further reduction in nucleosome signal strength among cold-treated plants compared with those growing at $22^{\circ}C$ (Figure 4.5 A). Remarkably, regardless of nucleosome occupancy levels among different groups, exons always showed higher nucleosome occupancy and can be differentiated from their flanking regions (introns).

To illuminate the relationship between DNA methylation, nucleosome occupancy and exon definition, the association between nucleosome patterns and DNA methylation patterns has been investigated in AzadC plants. For that, the methylation percentage (exons only) in three contexts (CG, CHH, and CHG), including 500 bp upstream and downstream of flanking introns has been illustrated (Figure 4.5 B-D).
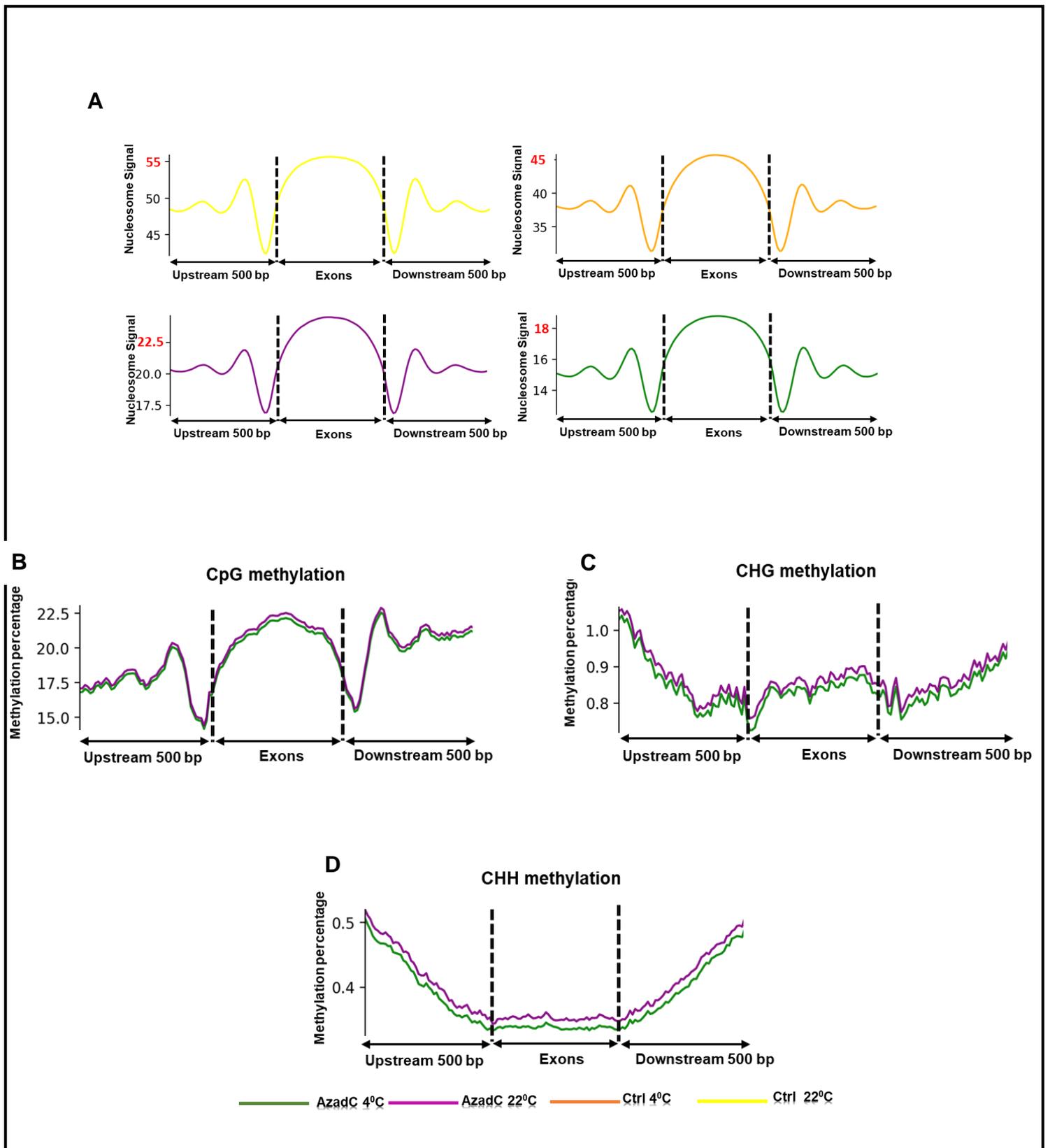
Figure 4.5. Nucleosome occupancy (A) and DNA methylation levels in CpG (B), CHG (C), and CHH context in exons and flanking regions. For A, the x axis represents exons scaled to 500 bp and their upstream and downstream flanking regions (500 bp); the y axis represents the nucleosome signal level. For all samples, exons are well defined by nucleosome occupancy that drops around ~ 25 bp upstream and downstream exons to increase again across flanking regions, but remains lower than exons level. For both temperatures, AzadC nucleosome levels around exons and flanking introns remains less than in Ctrl plants (green and purple lines compared to orange and yellow). Red numbers indicate the highest level of nucleosome occupancy detected in each sample. For B, C, and D: The x axis represents exons scaled to 500 bp and their upstream and downstream flanking regions (500 bp); the y axis represents the DNA methylation percentage in each sequence context.

Similar to nucleosome profiles, DNA methylation define exons, especially in CpG DNA methylation context (B), while this is less pronounced for CHG (C) and CHH (D). For all methylation contexts, AzadC plants subjected to cold stress display slightly lower DNA methylation percentage compared to AzadC grown at normal temperature (green and purple lines, respectively).

Interestingly, the percentage of methylated CpG dinucleotides (mCpG) accumulates at high levels along exons relative to flanking DNA, while mCHG methylation is suppressed in gene bodies compared to surrounding DNA. Like nucleosome profiles, AzadC plants subjected to cold stress displayed a significant decrease in DNA methylation in all contexts compared to AzadC plants grown at normal conditions (Figure 4.6 B-D). Nucleosome occupancy and DNA methylation sequences profiles were then aligned around global 5'SS and 3'SS to analyse the distribution of both epigenetic features at intron-exon junctions. A sharp drop in nucleosome occupancy at ~25 bp upstream of the acceptor site (3'SS), corresponding to the location of the branch point in the RNA transcript was detected (Figure 4.6 A-B). In Arabidopsis, branch points are located −11 to −60 bp upstream of the acceptor site, and the polypyrimidine stretch downstream of the branch point is A and T rich (Tolstrup, Rouzé, & Brunak, 1997). The AT rich sequences inhibit nucleosome formation in the DNA sequence (Peckham et al., 2007), which may be helpful to promote the binding of SFs to their corresponding cis-elements. Cytosine nucleotides at the splice sites show similar patterns of nucleosome occupancy and CG methylation on the sense and antisense strands for AzadC treated plants (Figure 4.6 C). Similarly, a weak correlation between nucleosome occupancy and mCHH and mCHG methylation is also observed (Figure 4.6 D-E).
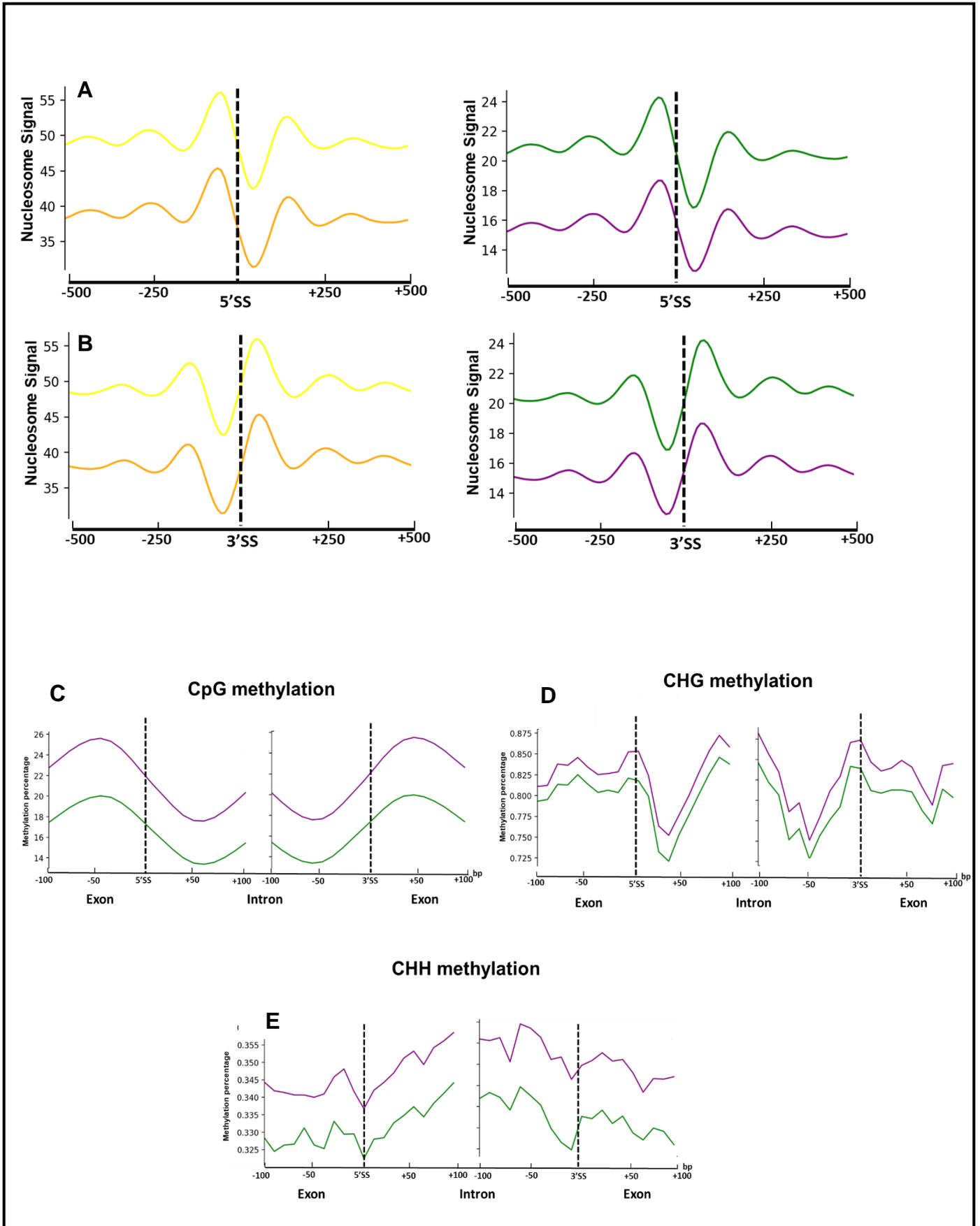
Figure 4.6. Splice site nucleosome occupancy (A and B) and DNA methylation for CpG, CHG, and CHH contexts (C-E). For **A and B,** the x axis represents 5'SS and 3'SS alongside 500 bp upstream and downstream the splice sites; the y axis represents the nucleosome signal level. For all samples, a drop of nucleosome occupancy is detected upstream and downstream 5'SS and 3'SS respectively. For both temperatures, AzadC nucleosome

levels around the splice sites and flanking regions is less than Ctrl plants (green and purple lines compared to orange and yellow). Red numbers indicate the highest level of nucleosome occupancy detected in each sample. **For C, D, and E:** the x axis is the position relative to the acceptor site (left) and donor site (right); the y axis represents the DNA methylation percentage in each sequence context. Similar to nucleosome profiles, DNA methylation levels drop upstream and downstream 5'SS and 3'SS, respectively in the CpG DNA methylation context (B), whereas it is less pronounced for the case of CHG (C) and CHH (D). For all methylation contexts, AzadC plants subject to cold stress display slightly lower DNA methylation percentage compared to AzadC grown at normal temperature (green and purple lines respectively).

4.3.2 DNA methylation and nucleosome occupancy can modulate expression and alternative splicing patterns

To reveal the relationship between DNA methylation, nucleosome occupancy and their influence on gene expression and splicing patterns, genes were ranked into five equal bins (in terms of gene number) based on their TPM values using RNA-seq data generated from each sample. Nucleosomes are then aligned to 1000 bp upstream and 1,000 bp downstream of the TSS of each gene. Compared to genes with lower transcript abundance, genes with higher transcript abundance exhibited lower nucleosome occupancy around TSS and TTS and flanking regions (Figure 4.7 A), which is consistent with previous studies in Arabidopsis, rice, maize, and humans (Fincher et al., 2013; G. Li et al., 2014; M.-J. Liu et al., 2015). Upon cold treatment, lower nucleosome occupancy levels are detected for all gene ranks in AzadC treated and Ctrl plants, while this decrease is more pronounced in AzadC plants, indicating the role of nucleosome re-arrangement in condition-specific gene expression. To further illuminate the potential link between DNA methylation and nucleosome occupancy to modulate chromatin organization, genes are ranked as described above and the methylation profiles for each bin was compared with the expression level of the corresponding genes in each bin. Genes with higher expression levels had lower CpG methylation levels around the TSS and TTS, whereas a loose correlation between gene expression and CpG methylation was observed in the middle of gene bodies (Figure 4.7 B), which is in line with previous studies from maize and Arabidopsis (Hollister, Smith, Guo, & Ott, 2011; Regulski et al., 2013; L. Yang, Takuno, Waters, & Gaut, 2011).
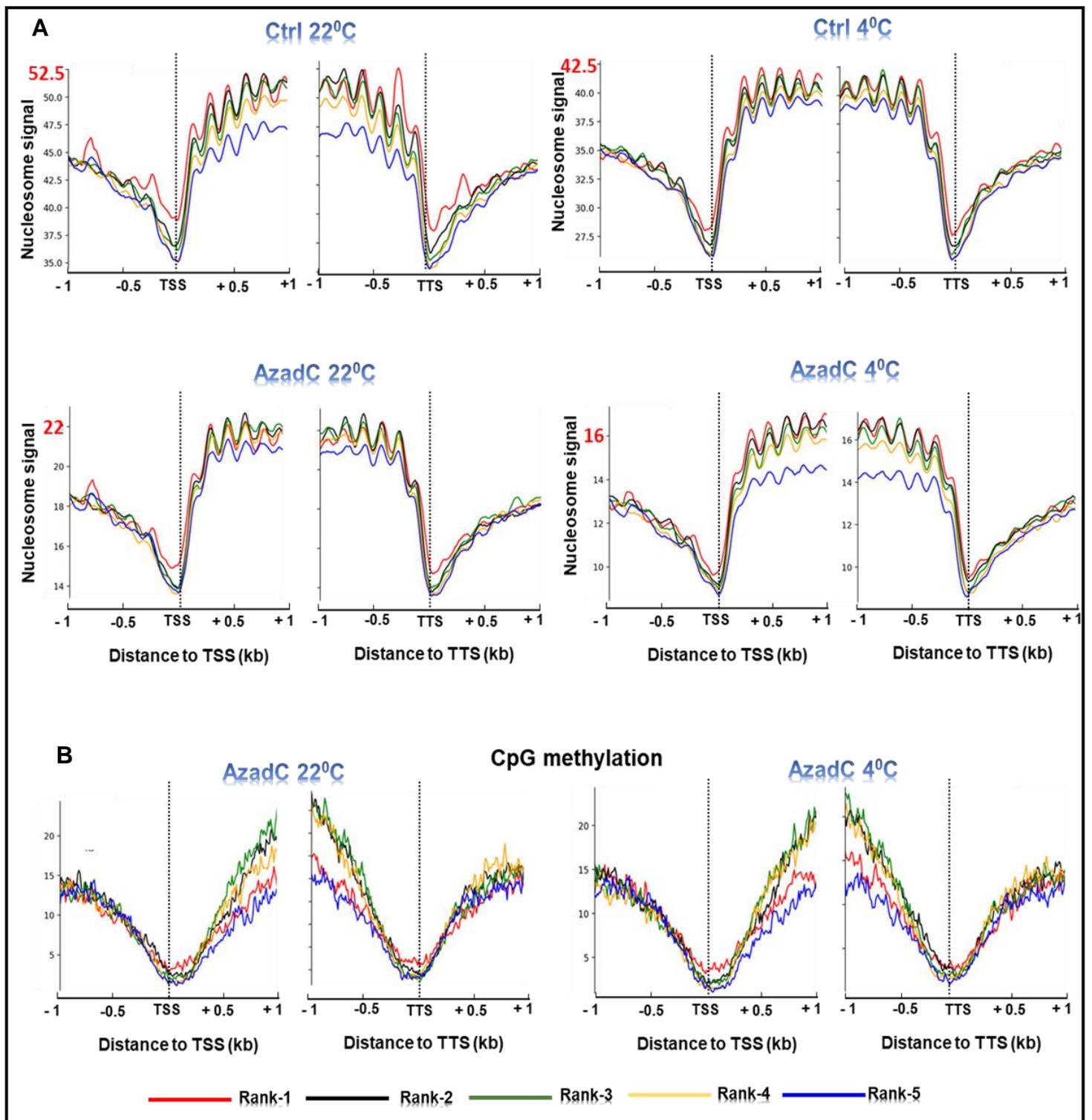
Figure 4.7 Relationships between nucleosome occupancy (A) and CG methylation (B) and gene expression in -1000/+1000 bp regions flanking the transcription start site (TSS, left) and transcription termination site (TTS, right). The x axis represents the distance to TSS (kb); the y axis represents the nucleosome signal level. Genes were divided according to their expression levels (Rank 1: lowest, Rank 5: Highest) into five equal bins then nucleosome occupancy and CG methylation were plotted for each inbred. Red numbers indicate the highest level of nucleosome occupancy detected in each sample.

Additionally, DNPs analysis using DANPOS version 2.1.2 (K. Chen et al., 2013) show genome-wide changes in nucleosome positioning upon cold stress as well as DNPs in cold-responsive genes regulating their expression and AS profiles through DNA methylation changes (Figure 4.8, table 4.11, the rest of table 4.11 data can be found in supplementary table 13). Upon similar temperature conditions, AzadC and Ctrl plants display 11415 and 8052 DNPs, respectively. This number increase to 13961 and 15241 in C and D, respectively where AzadC and Ctrl plants are shifted from 22°C to 4°C. Similarly, 11093 and 9291 DNPs were detected in E and F. For the identified DNPs , 2133, 1238, 2652, 3034, 1796, and 1361 genes in were identified in contrast groups A, B, C, D, E, and F, respectively which overlap with genomic regions having more than 2 DPNs in 1000 bp genomic window (Supplementary table 15 and 16). The overlap between DE, DAS, and genes detected in DNPs regions show that the first two contrast groups display 0% overlap which might be due to the low number of DE and DAS genes identified in these groups. Contrast group C and D clearly show a slight increase (0.8% and 0.3%, respectively) in the overlap between DE, DAS, and genes detected in the DPNs regions. This pattern remains true for contrast groups E and F showing an overlap of 0.7%. For all contrast groups, DE genes showed more overlap with DNPs in all contrast groups (2%, 1.3%, 4.8%, 2.8%, 3.7%, and 4.6% for contrast groups A, B, C, D, E, and F, respectively) compared to DAS genes (0.4%,0.4%, 1.1%, 0.6%, 1.2%, 1.2% for contrast groups A, B, C, D, E, and F respectively (Supplementary table 17).
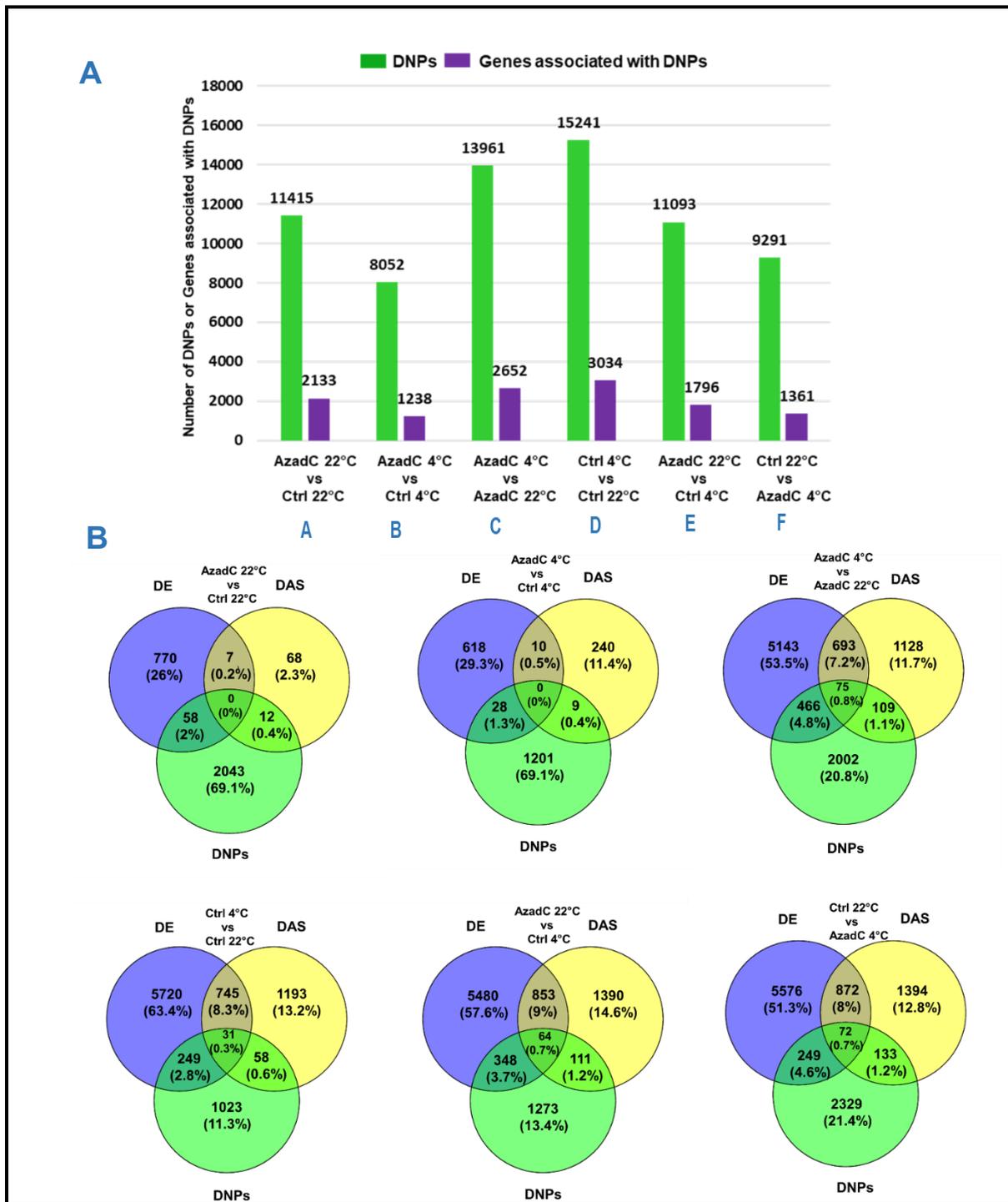
Figure 4.8. Summary of differential nucleosome positioning (DNPs) and their overlap with differentially expressed (DE) and alternatively spliced (DAS) genes. (A) The top panel represents the number of differentially positioned nucleosomes (DNPs) detected in each contrast group and the number of genes overlapping with DPNs. **(B)** The lower panel represents a Venn diagram showing the overlap between DE genes, DAS genes, and genes overlapping with DNPs.

Table 4.11. The most significant (Top 50) DNPs detected in each contrast group. Diff_smt_loc: The point with the biggest difference in occupancy position, smt_diff_FDR:FDR value for difference between treat_smt_val and control_smt_val, Point_diff_FDR: FDR value for treat_point_val and control_point_val.

<div align="center"><u>Contrast group 1: AzadC 22°C vs Ctrl 22°C</u></div>

Chromosome	Diff_smt_location	Smt_diff_FDR	Point_diff_FDR
Chr1	15082321	0.00001	0.00001
Chr2	2601	0	0
Chr2	2891	0.00001	0
Chr2	3491	0	0
Chr2	9991	0	0
Chr2	3619051	0	0
Chr2	3619491	0	0
Chr2	3621181	0	0
Chr2	3626271	0	0
Chr2	3341	0.00001	0.00001
Chr2	3607221	0.00001	0.00001
Chr2	3618121	0.00002	0.00001
Chr2	3619851	0.00001	0.00001
Chr2	3616591	0.00002	0.00002
Chr2	3623391	0.00003	0.00002
Chr2	3605821	0.00003	0.00003
Chr2	3626971	0.00003	0.00003
Chr2	7191021	0.00003	0.00003
Chr3	13590431	0	0
Chr3	13590631	0	0
Chr3	13590781	0	0
Chr3	13591671	0	0
Chr3	13591951	0	0
Chr3	14203281	0	0
Chr3	13587701	0.00001	0.00001
Chr3	13589481	0.00003	0.00001
Chr3	13591161	0.00001	0.00001
Chr3	14195631	0.00001	0.00001
Chr3	14196621	0.00001	0.00001
Chr3	14203801	0.00001	0.00001
Chr3	14196931	0.00002	0.00002
Chr3	13590251	0.00004	0.00003
Chr3	14197331	0.00004	0.00003
Chr3	14201721	0.00003	0.00003
Chr4	3950871	0	0
Chr4	3951561	0	0
Chr4	3952771	0	0
Chr4	3952971	0	0

Chr4	3953551	0	0
Chr4	3950621	0.00001	0.00001
Chr4	3951211	0.00001	0.00001
Chr4	3951391	0.00001	0.00001
Chr4	3952651	0.00004	0.00001
Chr4	3955161	0.00001	0.00001
Chr4	3952221	0.00003	0.00003
Chr4	3966491	0.00004	0.00003
Chr5	11731681	0.00001	0.00001
Chr5	11732281	0.00002	0.00001
Chr5	12810961	0.00001	0.00001

Contrast group 2: AzadC 4°C vs Ctrl 4°C

Chromosome	Diff_smt_loca	Smt_diff_FDR	Point_diff_FDR
Chr1	30108081	0.00005	0.00005
Chr1	18200721	0.00005	0.00005
Chr2	3626271	0	0
Chr2	2981	0.00179	0.00001
Chr2	3351	0.00001	0.00001
Chr2	10291	0.00001	0.00001
Chr2	3619481	0.00004	0.00003
Chr2	3623351	0.01273	0.00004
Chr2	3623601	0.00004	0.00004
Chr2	3619041	0.0001	0.00005
Chr3	14196071	0	0
Chr3	13590451	0.00003	0.00001
Chr3	14203311	0.00001	0.00001
Chr3	13713051	0.00001	0.00001
Chr3	13590631	0.00001	0.00001
Chr3	13590781	0.00001	0.00001
Chr3	13591651	0.00001	0.00001
Chr3	14196951	0.00004	0.00003
Chr3	14197411	0.00003	0.00003
Chr3	14201661	0.00012	0.00005
Chr3	1912681	0.00575	0.00005
Chr3	13589471	0.00108	0.00005
Chr3	13591931	0.00005	0.00005
Chr3	13885311	0.00007	0.00005
Chr4	3952651	0.0096	0
Chr4	3953551	0	0
Chr4	3952981	0.00001	0.00001
Chr4	3955151	0.00005	0.00003
Chr5	3253181	0.00011	0.00003
Chr5	18162591	0.00007	0.00003

Chr5	11731691	0.00003	0.00003
Chr5	665151	0.0002	0.00005
Chr5	7363206	0.00005	0.00005
Chr5	21522371	0.00005	0.00005
Chr5	10069823	0.00005	0.00005
Chr5	17345081	0.00005	0.00005
Chr5	3460691	0.00088	0.00005
Chr5	847821	0.00005	0.00005
Chr5	10023511	0.01597	0.00005
Chr5	9501051	0.00005	0.00005
Chr5	20682971	0.00007	0.00005
Chr5	3491761	0.00023	0.00005
Chr5	7704231	0.00007	0.00005
Chr5	13042351	0.0058	0.00005
Chr5	24525101	0.00044	0.00005
Chr5	14121026	0.00005	0.00005
Chr5	24797831	0.0005	0.00005
Chr5	22181921	0.00012	0.00005
Chr5	6665811	0.00005	0.00005
Chr5	1208021	0.00005	0.00005

Contrast group 3: AzadC 4°C vs AzadC 4°C

Chromosome	Diff_smt_loca	Smt_diff_FDR	Point_diff_FDR
Chr1	15083761	0.00001	0.00001
Chr1	15084111	0.00001	0.00001
Chr1	15085691	0.00001	0.00001
Chr2	9951	0	0
Chr2	3618461	0	0
Chr2	3341	0.00001	0.00001
Chr2	3065661	0.00001	0.00001
Chr2	3616781	0.00001	0.00001
Chr2	3618151	0.00001	0.00001
Chr2	3621211	0.00002	0.00001
Chr2	3621881	0.00001	0.00001
Chr2	3622631	0.00001	0.00001
Chr2	3624651	0.00001	0.00001
Chr2	3627321	0.00001	0.00001
Chr3	13587711	0	0
Chr3	13713041	0	0
Chr3	14195681	0	0
Chr3	13589351	0.00001	0.00001
Chr3	13590281	0.00001	0.00001
Chr3	13590651	0.00001	0.00001
Chr3	13591721	0.00001	0.00001

Chr3	14196081	0.00001	0.00001
Chr3	14197411	0.00001	0.00001
Chr3	14201671	0.00002	0.00001
Chr3	14203301	0.00001	0.00001
Chr4	3950901	0.00001	0.00001
Chr4	3951221	0.00001	0.00001
Chr4	3951941	0.00001	0.00001
Chr4	3952981	0.00001	0.00001
Chr4	3954581	0.00001	0.00001
Chr4	3955831	0.00006	0.00001
Chr4	3966601	0.01993	0.00001
Chr4	3968351	0.00001	0.00001
Chr4	3969221	0.00001	0.00001
Chr4	3978681	0.00001	0.00001
Chr4	3983451	0.00002	0.00001
Chr4	4009021	0.00001	0.00001
Chr4	4009801	0.00002	0.00001
Chr4	4010221	0.00001	0.00001
Chr4	4010771	0.00001	0.00001
Chr4	4011651	0.00001	0.00001
Chr5	3253201	0	0
Chr5	11727841	0	0
Chr5	11734891	0	0
Chr5	11707001	0.00001	0.00001
Chr5	11727121	0.00001	0.00001
Chr5	11727541	0.00001	0.00001
Chr5	11728591	0.00001	0.00001
Chr5	11730501	0.00001	0.00001
Chr5	11731491	0.00001	0.00001

Contrast group 4: Ctrl 4°C vs Ctrl 22°C

Chromosome	Diff_smt_loca	Smt_diff_FDR	Point_diff_FDR
Chr1	1988911	0.04463	0.00997
Chr1	15014461	0.01265	0.00996
Chr1	17119531	0.03483	0.00996
Chr1	21161941	0.03163	0.00996
Chr1	22071681	0.03039	0.00996
Chr1	22579521	0.01074	0.00996
Chr1	27796271	0.01852	0.00996
Chr1	4051751	0.0382	0.00995
Chr1	13169311	0.04668	0.00995
Chr1	22061151	0.01075	0.00995
Chr1	22181221	0.01122	0.00995
Chr2	1974201	0.01074	0.00998

Chr2	11060661	0.01196	0.00998
Chr2	11687341	0.01107	0.00997
Chr2	4325631	0.01684	0.00996
Chr3	648981	0.01593	0.01
Chr3	13612441	0.02769	0.01
Chr3	16226431	0.01	0.01
Chr3	19745251	0.02082	0.01
Chr3	20254751	0.02152	0.01
Chr3	8589501	0.01827	0.00998
Chr3	13021001	0.00998	0.00998
Chr3	19734591	0.03049	0.00998
Chr3	7298471	0.04928	0.00996
Chr3	10443291	0.00996	0.00996
Chr3	15745631	0.01091	0.00996
Chr4	18253461	0.0498	0.01
Chr4	7879641	0.04869	0.00998
Chr4	4326041	0.03784	0.00996
Chr4	13610111	0.03054	0.00996
Chr4	16723731	0.02383	0.00996
Chr4	17799796	0.00996	0.00996
Chr4	18412931	0.04607	0.00996
Chr5	16631611	0.01586	0.01
Chr5	17338431	0.01135	0.00998
Chr5	17628701	0.01446	0.00998
Chr5	19154261	0.01074	0.00998
Chr5	3075831	0.03368	0.00997
Chr5	12241401	0.02409	0.00997
Chr5	816061	0.04028	0.00996
Chr5	1397471	0.02306	0.00996
Chr5	3834541	0.0158	0.00996
Chr5	10106301	0.04058	0.00996
Chr5	11705561	0.00996	0.00996
Chr5	12936641	0.01425	0.00996
Chr5	16118981	0.01686	0.00996
Chr5	16901221	0.01231	0.00996
Chr5	19591121	0.0124	0.00996

Contrast group 5: AzadC 22°C vs Ctrl 4°C

Chromosome	Diff_smt_loca	Smt_diff_FDR	Point_diff_FDR
Chr1	15082321	6.00E-05	6.00E-05
Chr1	7359581	0.00311	7.00E-05
Chr2	10041	1.00E-05	0
Chr2	2611	1.00E-05	1.00E-05
Chr2	3491	1.00E-05	1.00E-05
Chr2	10281	1.00E-05	1.00E-05
Chr2	3619061	1.00E-05	1.00E-05
Chr2	3621181	1.00E-05	1.00E-05
Chr2	3626171	1.00E-05	1.00E-05
Chr2	3619501	2.00E-05	2.00E-05
Chr2	7191021	2.00E-05	2.00E-05
Chr2	3607241	4.00E-05	4.00E-05
Chr2	3619841	5.00E-05	4.00E-05
Chr2	1401	5.00E-05	5.00E-05
Chr2	3618131	5.00E-05	5.00E-05
Chr2	3626971	5.00E-05	5.00E-05
Chr2	3627311	6.00E-05	6.00E-05
Chr2	3616591	7.00E-05	7.00E-05
Chr3	13591671	0	0
Chr3	14203301	0	0
Chr3	13587701	1.00E-05	1.00E-05
Chr3	14195731	1.00E-05	1.00E-05
Chr3	14196071	1.00E-05	1.00E-05
Chr3	14196621	1.00E-05	1.00E-05
Chr3	14196941	1.00E-05	1.00E-05
Chr3	13590371	2.00E-05	2.00E-05
Chr3	13590631	2.00E-05	2.00E-05
Chr3	13591961	3.00E-05	2.00E-05
Chr3	14197401	2.00E-05	2.00E-05
Chr3	14201701	2.00E-05	2.00E-05
Chr3	14203921	5.00E-05	3.00E-05
Chr3	14203801	4.00E-05	4.00E-05
Chr3	13589361	5.00E-05	5.00E-05
Chr3	13591151	5.00E-05	5.00E-05
Chr3	14194811	7.00E-05	6.00E-05
Chr3	14196491	6.00E-05	6.00E-05
Chr3	14201471	0.0001	7.00E-05
Chr4	3950881	1.00E-05	0
Chr4	3952971	0	0
Chr4	3953541	0	0
Chr4	3952771	2.00E-05	2.00E-05
Chr4	3951571	3.00E-05	3.00E-05

Chr4	3951211	4.00E-05	4.00E-05
Chr4	3951391	6.00E-05	6.00E-05
Chr4	3954771	6.00E-05	6.00E-05
Chr5	12810971	2.00E-05	2.00E-05
Chr5	3253201	3.00E-05	3.00E-05
Chr5	11732281	4.00E-05	4.00E-05
Chr5	11185051	6.00E-05	5.00E-05

Contrast group 6: Ctrl 22°C vs AzadC 4°C

Chromosome	Diff_smt_loca	Smt_diff_FDR	Point_diff_FDR
Chr1	15083741	3.00E-05	2.00E-05
Chr2	9981	0	0
Chr2	3621191	0	0
Chr2	2901	2.00E-05	2.00E-05
Chr2	3481	2.00E-05	2.00E-05
Chr2	3607231	2.00E-05	2.00E-05
Chr2	3618131	2.00E-05	2.00E-05
Chr2	3618451	2.00E-05	2.00E-05
Chr2	3619051	2.00E-05	2.00E-05
Chr2	3619511	2.00E-05	2.00E-05
Chr2	3622611	3.00E-05	2.00E-05
Chr2	3627321	2.00E-05	2.00E-05
Chr2	2611	4.00E-05	4.00E-05
Chr2	3619841	5.00E-05	4.00E-05
Chr2	3624641	6.00E-05	6.00E-05
Chr3	13587711	0	0
Chr3	13590641	0	0
Chr3	13591691	0	0
Chr3	14195701	0	0
Chr3	14203291	0	0
Chr3	13589351	2.00E-05	2.00E-05
Chr3	13591971	2.00E-05	2.00E-05
Chr3	13590271	4.00E-05	4.00E-05
Chr3	13590431	4.00E-05	4.00E-05
Chr3	14196071	4.00E-05	4.00E-05
Chr3	14203801	4.00E-05	4.00E-05
Chr3	13591161	5.00E-05	5.00E-05
Chr3	14196921	6.00E-05	5.00E-05
Chr3	14197391	5.00E-05	5.00E-05
Chr3	14201711	5.00E-05	5.00E-05
Chr3	14196631	7.00E-05	6.00E-05
Chr4	3950881	0	0
Chr4	3953541	0	0
Chr4	3952971	1.00E-05	1.00E-05

Chr4	3951211	2.00E-05	2.00E-05
Chr4	3951571	2.00E-05	2.00E-05
Chr4	3952761	2.00E-05	2.00E-05
Chr4	3950611	3.00E-05	3.00E-05
Chr4	3951931	5.00E-05	5.00E-05
Chr4	3954591	6.00E-05	5.00E-05
Chr4	3951391	6.00E-05	6.00E-05
Chr4	3954771	6.00E-05	6.00E-05
Chr4	4009771	6.00E-05	6.00E-05
Chr5	3253211	2.00E-05	2.00E-05
Chr5	11732271	2.00E-05	2.00E-05
Chr5	11734891	2.00E-05	2.00E-05
Chr5	11727821	4.00E-05	4.00E-05
Chr5	11727541	6.00E-05	6.00E-05
Chr5	11730481	6.00E-05	6.00E-05

To understand how epigenetic features are involved in splicing and AS regulation, nucleosome and CG methylation were profiled around the 5'SS and 3'SS sites of alternative and constitutively spliced exons and introns and their flanking regions (Figure 4.9 A). Nucleosome occupancy levels were found to be lower in cassette exons than in constitutively spliced exons, which compares well with previous reports from mammals (S. Schwartz et al., 2009; Tilgner et al., 2009). Similarly, constitutively spliced introns displayed higher nucleosome occupancy level around the donor splice site. By investigating mCG methylation across the same genomic regions, a similar patterns of DNA methylation (mirroring nucleosome occupancy) were found around the donor and acceptor site of alternatively and constitutively spliced exons and introns; however, methylation differences between constitutively and alternatively spliced introns are relatively higher than in exons (Figure 4.9 B).
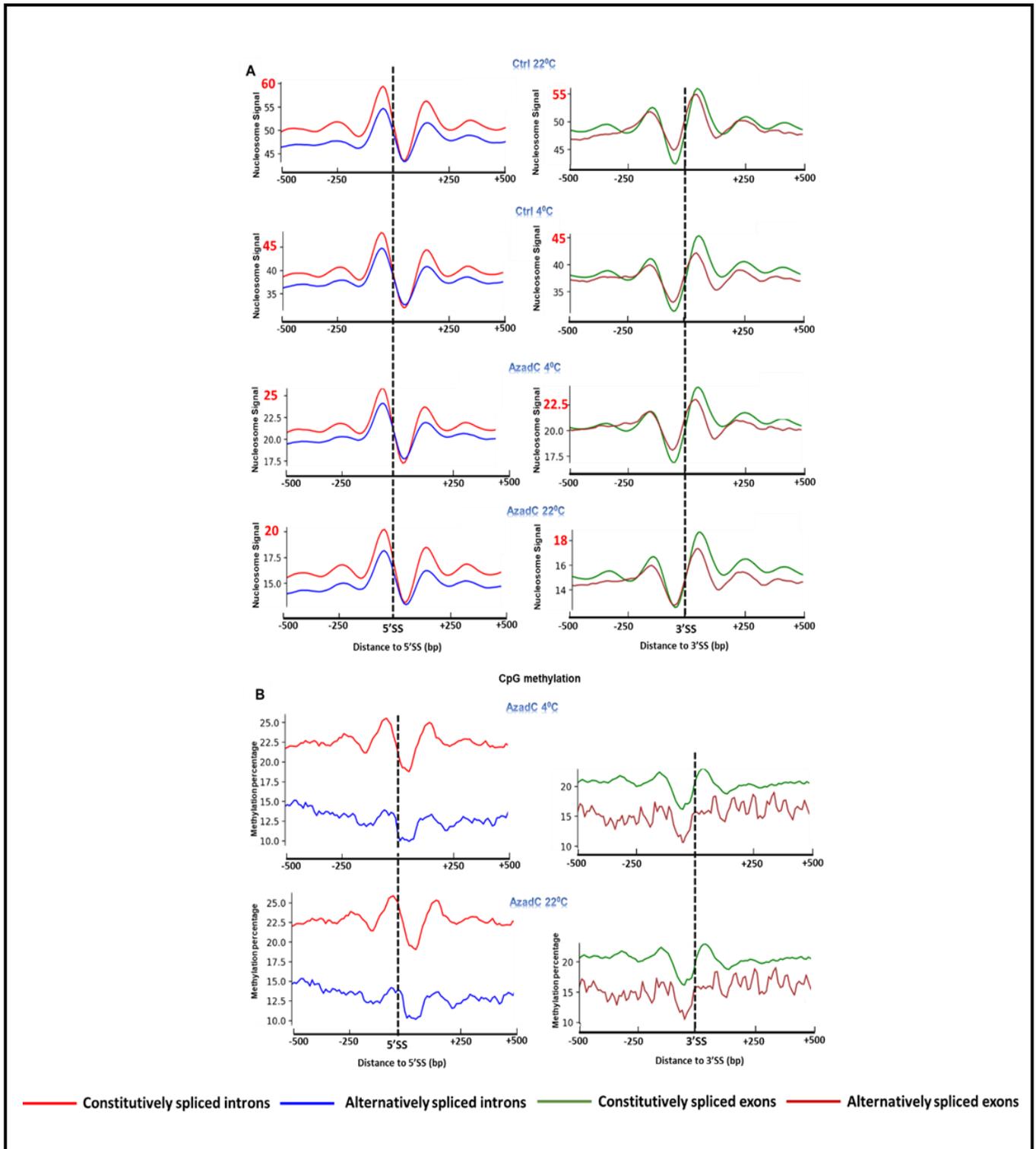
Figure 4.9. Nucleosome occupancy profiles (A) and CG methylation (B) around the donor site of alternatively and constitutively spliced introns (left) the acceptor site of alternatively and constitutively spliced exons (Right). The x-axis is the position relative to acceptor site (left) and donor site (right); the y-axis is the nucleosome signal level for **A** and CG percentage for **B.** Nucleosome occupancy profiles differ between constitutively and alternatively spliced exons and introns. Overall, constitutively spliced exons and introns display higher nucleosome occupancy compared to alternatively spliced ones. This comparison remains true for all for all samples and in the case of DNA methylation level (B) as well.   Red numbers indicate the highest level of nucleosome occupancy detected in each sample.

To test the extent to which DAS genes detected in our contrast groups display different nucleosome profiles to potentially regulate AS patterns, we profiled nucleosome around uniquely DAS genes detected in each contrast group (Supplementary Data Set 4). Interestingly, results show that each contrast group displayed specific nucleosome signal patterns and levels (Figure 4.10).
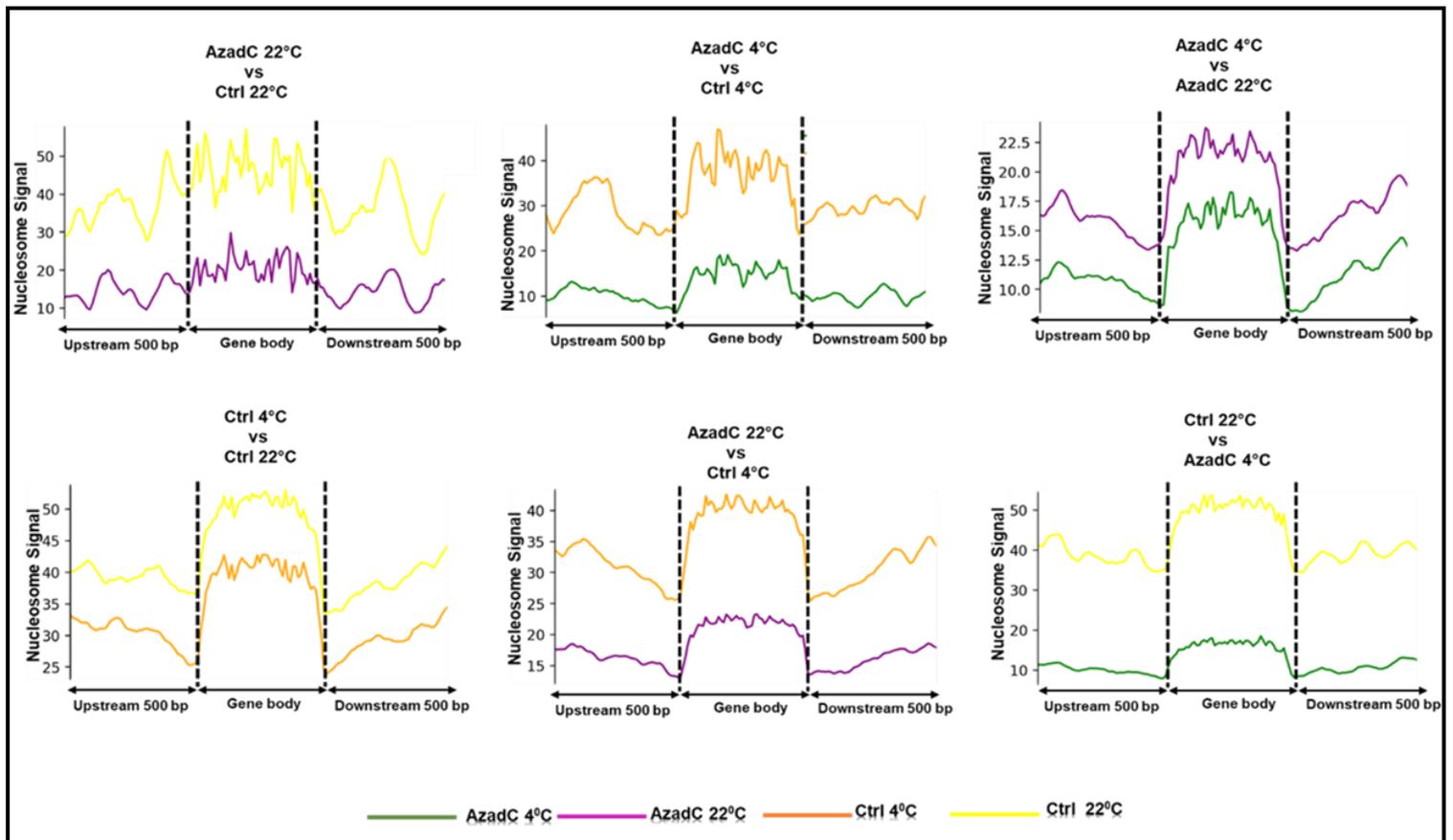
Figure 4.10. Nucleosome occupancy profiles for uniquely alternatively spliced (DAS) genes detected in different contrast groups. Nucleosome profiles are plotted against the gene body of DAS genes and 500 bp upstream and downstream of the gene start and end, respectively. The x axis represents genes scaled to 500 bp and their upstream and downstream flanking regions (500 bp); we grouped 7 (for AzadC 22°C vs Ctrl 22°C), 16 (for AzadC 4°C vs Ctrl 4°C), 117 (for AzadC 4°C vs AzadC 4°C), 199 (for Ctrl 4°C vs Ctrl 22°C), 186 (for AzadC 22°C vs Ctrl 4°C), 121 for Ctrl 22°C vs AzadC 4°C). Average nucleosome profiles in each contrast group are plotted around normalised selected genes. Distinctive nucleosome profiles are observed for DAS genes in each contrast group.

Interestingly, nucleosome levels are significantly lower in the exonic regions associated with A5'SS and ES with ES displaying even lower nucleosome occupancy than that of A5'SS (Figure 4.11 A). Both A5'SS and A3'SS do not have a strong association with nucleosome occupancy which is consistent with previous reports from humans (Zhou, Lu, & Tian, 2012). Interestingly, IR events displayed the lowest nucleosome occupancy compared to other splicing events around exons (Figure 4.5 A). Cold treatment decreased nucleosome

positioning in exons and introns involved in the splicing events among Ctrl and AzadC plants which, in principle, could contribute to reduced exon and intron recognition, henceforth leading to higher exon skipping and intron retention events under cold stress. AzadC plants displayed less nucleosome positioning for all AS compared to Ctrl under the same temperature conditions, indicating the interplay between nucleosome occupancy and DNA methylation in regulating differential AS (Figure 4.11 B).

4.3.3 Characteristic methylation and nucleosome occupancy define exitrons

A subset of retained introns, named 'exitrons', which are internal parts of protein-coding exons, has been identified in Arabidopsis and human (Marquez et al., 2015; Sibley, Blazquez, & Ule, 2016; Dorothee Staiger & Simpson, 2015). Exitrons originate from protein-coding exonic sequences, and their evolution involved intron loss in the exitron-containing exons. Because exitrons are parts of protein-coding exons, they exhibit an absence of stop codons and prevalence of synonymous substitutions. The majority of exitrons have lengths of multiples of three nucleotides, therefore their inclusion or removal do not change the reading frame. Splicing of exitrons affects sequences that encode protein domains, disordered regions and various types of post-translational modifications, hence, affecting protein function and regulatory capacity. At least 6.6% of Arabidopsis and 3.7% of human of protein-coding genes contain exitrons. Intriguingly, exitron regions show higher GC content compared to constitutive and retained introns but lower GC content when compared with different groups of exons. Moreover, exitrons have lower GC content than adjacent sequences of exitron-containing exons (Marquez et al., 2015). Therefore, it is worth to investigate whether differential GC content in exitron sequences has any relation to DNA methylation and nucleosome occupancy in distinguishing exitrons from flanking exonic regions. Towards that goal, nucleosome occupancy and CpG methylation across about 2400 exitrons identified in Arabidopsis (Marquez et al., 2015; R. Zhang et al., 2017) were profiled; 500 bp upstream and downstream from the start (5' SS) and end (3'SS) of exitrons. Interestingly, exitrons display lower nucleosome occupancy when compared to flanking exonic sequences yet higher nucleosome occupancy compared to introns in both AzadC and Ctrl plants (Figure 4.11 C and D). Additionally, nucleosome patterns observed around exitrons are different from those detected around exons (Figure 4.5). Sharp peaks of nucleosome occupancy located before the start and after the end of exitrons were clearly observed, however, slightly lower occupancy in the middle of exitrons was detected. Similarly, DNA methylation levels are higher in exitrons compared to introns but more variable than nucleosome occupancy levels around exons (Figure

4.11 E). Although, nucleosome levels around exitrons are lower in AzadC compared to Ctrl and in plants treated with cold compared to ones grown at 22°C; they exhibited a similar pattern.
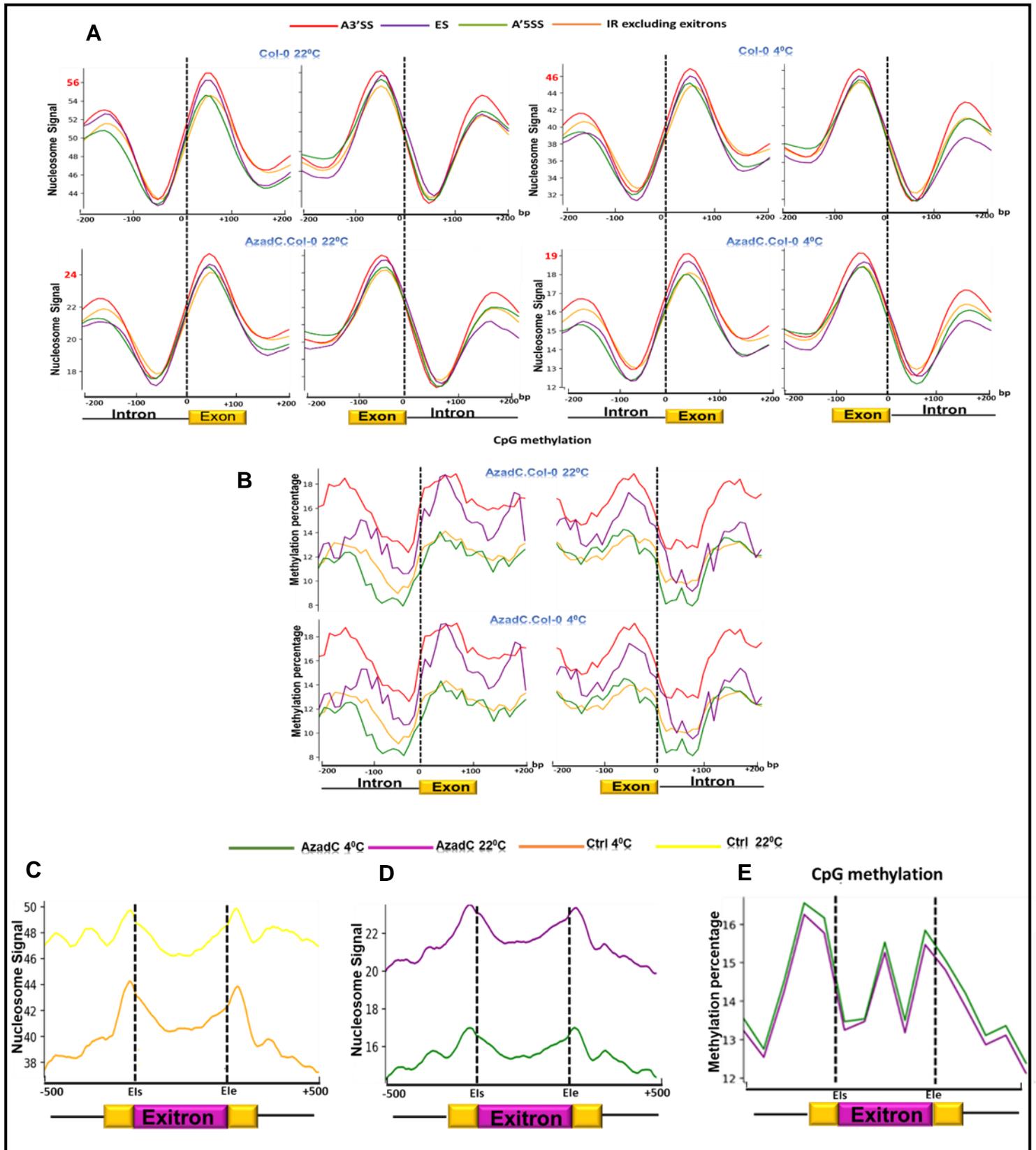
Figure 4.11. The association of nucleosome occupancy (A) and DNA methylation (B) with different AS events. The x axis is the position relative to the acceptor site (left) and donor site (right); the y-axis is the nucleosome signal density for A and CG percentage for **B**. ES: Exon skipping, A3'SS: Alternative 3'SS, A5'SS: Alternative 5'SS,

and IR: Intron retention. Exitron AS events nucleosome profiles **(C and D)** and CpG methylation percentage **(E)** are plotted separately to present the profiles of exitron definition. Each AS event represents a specific nucleosome occupancy level while maintaining the same exon definition. Yet, exitrons, a subset of IR events, display a distinctive nucleosome occupancy and DNA methylation pattern and level compared to other events. EIs and EIe are an exitron's start and end respectively.

Additionally, significantly differentially spliced exitrons were detected and nucleosome profiles were plotted against the coordinate of DAS EIs (Figure 4.12 and table 4.12). Interestingly, DAS EIs nucleosome profiles patterns and levels show differences among different contrast groups. For instance, DAS EIs displayed lower nucleosome occupancy compared to exons yet higher when compared to introns regardless of the contrast group. Additionally, up-regulated and down-regulated exitrons for the same contrast groups display opposite profiles of nucleosome occupancy. This data point towards the importance of nucleosome occupancy in defining a new subset of IR events, the exitrons, and regulating their AS profiles under normal and cold stress conditions.
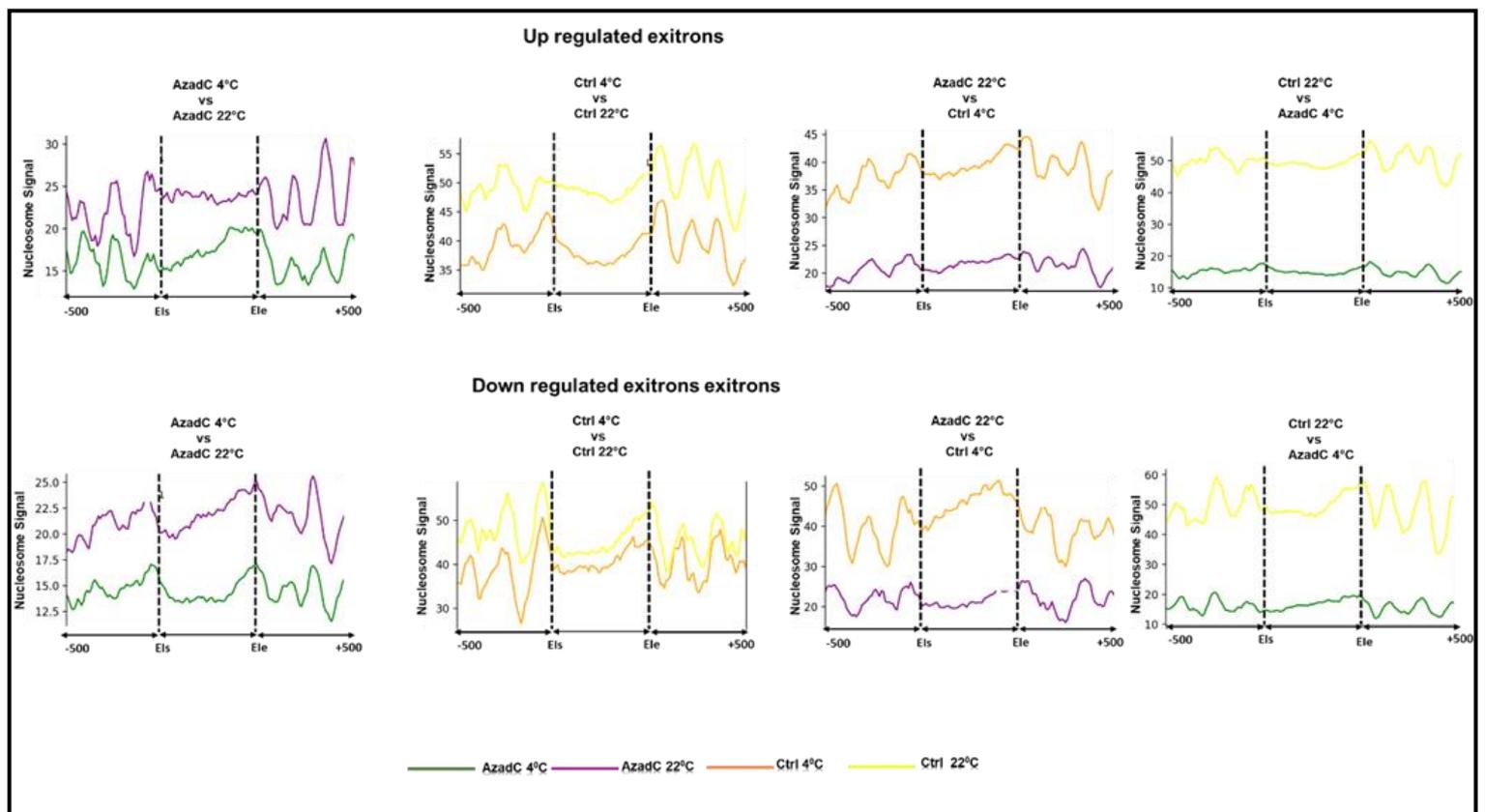
Figure 4.12. Nucleosome profiles around differentially spliced exitrons. The x-axis represents an exitron's start and end alongside the flanking regions; the y-axis represents the nucleosome signal. In each contrast group, upregulated and downregulated exitrons were scaled to 500 bp and nucleosome profiles were plotted across exitrons (EIs and EIe for an exitron's start and end respectively) and 500 bp upstream and downstream exitrons. For the same contrast group, exitrons show opposite nucleosome patterns for up regulated and downregulated exitrons.

Table 4.12. Significant (P < 0.05) differentially alternatively spliced exitrons detected in different contrast groups. EIs: exitron start, EIe: exitron.

Contrast group 1: AzadC 22°C vs Ctrl 22°C

Chromosome	EIs	EIe	Strand	Gene ID	ΔPSI	P value
Chr1	28777603	28777679	+	AT1G76680	0.049713	0.016983
Chr1	30017442	30017518	-	AT1G79790	0.044195	0.030969
Chr2	650395	650501	+	AT2G02450	0.080078	0.021479
Chr4	1118897	1119007	+	AT4G02540	0.079274	0.022977
Chr4	12612953	12613035	+	AT4G24380	0.096768	0.006993
Chr4	18072349	18072410	-	AT4G38680	0.054609	0.015984
Chr5	26123852	26123944	-	AT5G65380	0.078848	0.047952

Contrast group 2: AzadC 4°C vs Ctrl 4°C

Chromosome	EIs	EIe	Strand	Gene ID	ΔPSI	P value
Chr1	9779825	9779905	+	AT1G28060	0.064031	0.043956
Chr1	28306149	28306247	+	AT1G75420	0.084263	0.049451
Chr2	9265484	9265575	-	AT2G21660	-0.32222	0.047952
Chr2	15489532	15489614	-	AT2G36895	0.032213	0.028971
Chr2	15743353	15743429	-	AT2G37510	0.263608	0.022977
Chr4	10814600	10814679	+	AT4G19960	0.067961	0.041958
Chr4	12612953	12613035	+	AT4G24380	-0.04746	0.043457
Chr4	18072349	18072410	-	AT4G38680	-0.10978	0.009491

Contrast group 3: AzadC 4°C vs AzadC.Ctrl 4°C

Chromosome	EIs	EIe	Strand	Gene ID	ΔPSI	P value
Chr1	3153047	3153144	-	AT1G09730	-0.11590455	0.040959041
Chr1	4845150	4845226	-	AT1G14170	-0.24216955	0
Chr1	9779825	9779905	+	AT1G28060	-0.238272341	0
Chr1	17742356	17742438	-	AT1G48090	-0.077881717	0.036713287
Chr1	25000381	25000469	-	AT1G66980	-0.167836091	0.022477523
Chr1	25755241	25755331	+	AT1G68580	0.225480115	0.031968032
Chr1	25778138	25778229	-	AT1G68660	-0.017619117	0.043956044
Chr1	28734608	28734705	+	AT1G76580	-0.10836991	0.034965035
Chr1	29255721	29255803	+	AT1G77800	-0.133119491	0.013486514
Chr1	30082890	30083024	+	AT1G79970	-0.060196677	0.030469531
Chr2	650395	650501	+	AT2G02450	-0.205981772	0.000999001
Chr2	10452749	10452909	-	AT2G24600	-0.27204256	0.026973027
Chr2	13935641	13935702	-	AT2G32850	-0.213502475	0.005994006
Chr2	15489532	15489614	-	AT2G36895	-0.036316943	0.027472528
Chr2	16425150	16425246	+	AT2G39340	0.165338747	0.017982018
Chr2	18028319	18028449	-	AT2G43410	-0.124126699	0.046453547
Chr2	19360450	19360526	-	AT2G47160	-0.106102175	0.04945055
Chr3	392355	392448	+	AT3G02150	-0.078304445	0.015984016

Chr3	4304100	4304206	+	AT3G13300	-0.276386424	0
Chr3	6201686	6201757	+	AT3G18100	-0.318363676	0.020979021
Chr3	6748498	6748601	+	AT3G19460	0.09058721	0.034132534
Chr3	10090592	10090689	+	AT3G27320	-0.17068209	0.034465535
Chr3	20177555	20177621	-	AT3G54500	0.164628052	0.006993007
Chr3	20178021	20178123	-	AT3G54500	0.078754394	0.024864025
Chr3	20178064	20178158	-	AT3G54500	-0.090154899	0.010989011
Chr3	20178064	20178161	-	AT3G54500	-0.145797021	0.005244755
Chr3	20270980	20271074	-	AT3G54760	-0.112944087	0.020979021
Chr3	20748238	20748317	+	AT3G55940	-0.127054913	0.014485515
Chr3	21829086	21829166	-	AT3G59060	0.181385175	0.004995005
Chr3	21922917	21923011	-	AT3G59310	0.18620629	0.011238761
Chr4	1118897	1119007	+	AT4G02540	-0.179098336	0.024975025
Chr4	7125245	7125336	-	AT4G11840	-0.162001906	0.014985015
Chr4	12612953	12613035	+	AT4G24380	-0.05466439	0.024975025
Chr4	13082090	13082160	-	AT4G25650	-0.046967965	0.013486514
Chr4	15130346	15130419	+	AT4G31115	-0.084475544	0.043956044
Chr4	17295297	17295398	-	AT4G36690	-0.062397457	0.018106893
Chr4	18072349	18072410	-	AT4G38680	-0.043029734	0.03996004
Chr5	4557631	4557707	+	AT5G14120	0.063435123	0.030969031
Chr5	6834529	6834645	+	AT5G20250	0.135777511	0.034965035
Chr5	14815843	14815938	-	AT5G37370	-0.32245943	0.008991009
Chr5	18844142	18844225	+	AT5G46470	-0.186451692	0.005994006
Chr5	20967606	20967695	+	AT5G51620	-0.226245364	0.013986014
Chr5	21243164	21243243	+	AT5G52310	-0.234467351	0.006243756
Chr5	23979100	23979178	+	AT5G59470	-0.09919067	0.038961039
Chr5	26123852	26123944	-	AT5G65380	-0.092534622	0.035964036

Contrast group 4: Ctrl 4°C vs Ctrl 22°C

Chromosome	EIs	EIe	Strand	Gene ID	ΔPSI	P value
Chr1	3153047	3153144	-	AT1G09730	-0.18081	0.004995
Chr1	4845150	4845226	-	AT1G14170	-0.27531	0.005994
Chr1	5664288	5664358	+	AT1G16540	-0.26493	0.02997
Chr1	9779825	9779905	+	AT1G28060	-0.27495	0.000999
Chr1	17742356	17742438	-	AT1G48090	-0.17358	0.000749
Chr1	25000381	25000469	-	AT1G66980	-0.1405	0.017982
Chr1	25778138	25778229	-	AT1G68660	-0.02341	0.027972
Chr1	28734608	28734705	+	AT1G76580	-0.11929	0.027972
Chr1	29255721	29255803	+	AT1G77800	-0.10286	0.032218
Chr1	30017442	30017518	-	AT1G79790	0.094254	0.007992
Chr1	30174975	30175061	+	AT1G80245	-0.14717	0.035964
Chr2	650395	650501	+	AT2G02450	-0.14941	0.004496
Chr2	1719127	1719200	+	AT2G04880	-0.06526	0.036464
Chr2	10452749	10452909	-	AT2G24600	-0.2874	0.028472

Chr2	13935641	13935702	-	AT2G32850	-0.21835	0.001499
Chr2	15489532	15489614	-	AT2G36895	-0.06096	0.007992
Chr2	15743353	15743429	-	AT2G37510	-0.39523	0.008991
Chr2	16425150	16425246	+	AT2G39340	0.162416	0.015984
Chr2	18028319	18028449	-	AT2G43410	-0.16247	0.020979
Chr2	19360450	19360526	-	AT2G47160	-0.17602	0.014486
Chr3	392355	392448	+	AT3G02150	-0.07677	0.027972
Chr3	3777224	3777303	+	AT3G11930	-0.13876	0.005994
Chr3	4304100	4304206	+	AT3G13300	-0.24445	0.004496
Chr3	6201686	6201757	+	AT3G18100	-0.23481	0.046454
Chr3	18698338	18698590	-	AT3G50380	-0.13262	0.027972
Chr3	20177555	20177621	-	AT3G54500	0.140485	0.004196
Chr3	20178064	20178158	-	AT3G54500	-0.12406	0.006993
Chr3	20178064	20178161	-	AT3G54500	-0.22072	0
Chr3	20270980	20271074	-	AT3G54760	-0.16684	0.001998
Chr3	20748238	20748317	+	AT3G55940	-0.14032	0.00999
Chr3	21660146	21660216	+	AT3G58570	-0.11779	0.018482
Chr3	21829086	21829166	-	AT3G59060	0.253851	0.00999
Chr3	21922917	21923011	-	AT3G59310	0.210117	0.012737
Chr4	7125245	7125336	-	AT4G11840	-0.14362	0.016983
Chr4	10696525	10696607	-	AT4G19660	-0.19276	0.032967
Chr4	11475849	11475949	+	AT4G21580	-0.0401	0.041958
Chr4	12180318	12180448	-	AT4G23290	-0.10264	0.038462
Chr4	12612953	12613035	+	AT4G24380	0.089566	0.014486
Chr4	12815702	12815812	+	AT4G24900	-0.21231	0.01998
Chr4	13082090	13082160	-	AT4G25650	-0.05532	0.003497
Chr4	15130346	15130419	+	AT4G31115	-0.07841	0.043956
Chr4	15658272	15658397	+	AT4G32440	-0.15107	0.047952
Chr4	17295297	17295398	-	AT4G36690	-0.05483	0.025599
Chr4	18072349	18072410	-	AT4G38680	0.121364	0.002498
Chr4	18416175	18416267	-	AT4G39680	-0.03611	0.04046
Chr5	5081587	5081675	+	AT5G15610	-0.04885	0.042458
Chr5	6024546	6024616	-	AT5G18230	-0.07103	0.031469
Chr5	6062261	6062346	+	AT5G18310	-0.07039	0.032468
Chr5	8454585	8454681	+	AT5G24680	-0.27166	0.030969
Chr5	14815843	14815938	-	AT5G37370	-0.2661	0.017982
Chr5	18844142	18844225	+	AT5G46470	-0.15818	0.008991
Chr5	21243164	21243243	+	AT5G52310	-0.25207	0.007493
Chr5	23713301	23713385	-	AT5G58700	-0.50424	0.047952
Chr5	26772773	26772844	+	AT5G67080	0.445195	0.022478

Chromosome	EIs	EIe	Strand	Gene ID	ΔPSI	P value
Chr1	3153047	3153144	-	AT1G09730	0.172986	0.017982018
Chr1	4845150	4845226	-	AT1G14170	0.255032	0
Chr1	5664288	5664358	+	AT1G16540	0.243862	0.038961039
Chr1	7883125	7883197	-	AT1G22310	0.124077	0.044955045
Chr1	9779825	9779905	+	AT1G28060	0.302304	0
Chr1	17742356	17742438	-	AT1G48090	0.101952	0.01948052
Chr1	25000381	25000469	-	AT1G66980	0.164537	0.017982018
Chr1	25778138	25778229	-	AT1G68660	0.022635	0.021478522
Chr1	26984215	26984282	+	AT1G71720	0.270583	0.040959041
Chr1	28734608	28734705	+	AT1G76580	0.12379	0.026223776
Chr1	29255721	29255803	+	AT1G77800	0.128951	0.015734266
Chr1	30017442	30017518	-	AT1G79790	-0.05006	0.02997003
Chr1	30082890	30083024	+	AT1G79970	0.067477	0.031968032
Chr2	650395	650501	+	AT2G02450	0.229488	0.000499501
Chr2	1719127	1719200	+	AT2G04880	0.059511	0.045954046
Chr2	13935641	13935702	-	AT2G32850	0.210738	0.002997003
Chr2	15489532	15489614	-	AT2G36895	0.068529	0.005994006
Chr2	15743353	15743429	-	AT2G37510	0.298249	0.024975025
Chr2	16425150	16425246	+	AT2G39340	-0.19149	0.007992008
Chr2	18028319	18028449	-	AT2G43410	0.132677	0.035964036
Chr2	19360450	19360526	-	AT2G47160	0.138769	0.02947053
Chr3	392355	392448	+	AT3G02150	0.082864	0.012487513
Chr3	734285	734383	+	AT3G03180	-0.33591	0.032467533
Chr3	1213702	1213774	-	AT3G04500	0.400214	0.013986014
Chr3	3777224	3777303	+	AT3G11930	0.111357	0.021978022
Chr3	4304100	4304206	+	AT3G13300	0.26468	0.001498502
Chr3	6201686	6201757	+	AT3G18100	0.294925	0.010489511
Chr3	17464051	17464137	+	AT3G47390	0.14851	0.046453547
Chr3	20177555	20177621	-	AT3G54500	-0.13724	0.004662005
Chr3	20178021	20178123	-	AT3G54500	-0.06129	0.02952603
Chr3	20178064	20178158	-	AT3G54500	0.119813	0.005994006
Chr3	20178064	20178161	-	AT3G54500	0.21632	0.003496504
Chr3	20270980	20271074	-	AT3G54760	0.163173	0.004995005
Chr3	20748238	20748317	+	AT3G55940	0.132773	0.008991009
Chr3	21829086	21829166	-	AT3G59060	-0.25206	0.004995005
Chr3	21922917	21923011	-	AT3G59310	-0.19521	0.003746254
Chr4	7125245	7125336	-	AT4G11840	0.158754	0.007992008
Chr4	10696525	10696607	-	AT4G19660	0.152288	0.047952048
Chr4	10814600	10814679	+	AT4G19960	0.092361	0.028471529
Chr4	11475849	11475949	+	AT4G21580	0.032759	0.045954046
Chr4	12180318	12180448	-	AT4G23290	0.107767	0.017482518
Chr4	13082090	13082160	-	AT4G25650	0.057473	0.006993007
Chr4	17295297	17295398	-	AT4G36690	0.058877	0.032467533

163

Chr4	18072349	18072410	-	AT4G38680	-0.06676	0.015984016
Chr4	18416175	18416267	-	AT4G39680	0.031309	0.041958042
Chr5	3775519	3775692	+	AT5G11710	-0.04806	0.043456544
Chr5	6024546	6024616	-	AT5G18230	0.063313	0.033216783
Chr5	6834529	6834645	+	AT5G20250	-0.13407	0.040792541
Chr5	8454585	8454681	+	AT5G24680	0.367957	0.007992008
Chr5	14815843	14815938	-	AT5G37370	0.322693	0.026973027
Chr5	18844142	18844225	+	AT5G46470	0.166561	0.003996004
Chr5	19185152	19185240	+	AT5G47240	0.041827	0.034965035
Chr5	20967606	20967695	+	AT5G51620	0.213457	0.00999001
Chr5	20967606	20967699	+	AT5G51620	0.155763	0.031468532
Chr5	21243164	21243243	+	AT5G52310	0.219752	0.004995005
Chr5	26123852	26123944	-	AT5G65380	0.094384	0.020979021

Contrast group 6: Ctrl 22°C vs AzadC 4°C

Chromosome	EIs	EIe	Strand	Gene ID	ΔPSI	P value
Chr1	3153047	3153144	-	AT1G09730	0.123728	0.014486
Chr1	4845150	4845226	-	AT1G14170	0.262449	0.005994
Chr1	9779825	9779905	+	AT1G28060	0.210923	0.002498
Chr1	17742356	17742438	-	AT1G48090	0.149506	0.005245
Chr1	25000381	25000469	-	AT1G66980	0.143797	0.017982
Chr1	25778138	25778229	-	AT1G68660	0.018394	0.036963
Chr1	28734608	28734705	+	AT1G76580	0.103875	0.04021
Chr1	29255721	29255803	+	AT1G77800	0.107023	0.017982
Chr1	30017442	30017518	-	AT1G79790	-0.07497	0.02997
Chr2	650395	650501	+	AT2G02450	0.125904	0.004995
Chr2	9083513	9083583	+	AT2G21195	-0.15162	0.038961
Chr2	10452749	10452909	-	AT2G24600	0.375069	0.04046
Chr2	13935641	13935702	-	AT2G32850	0.221116	0.007493
Chr2	16425150	16425246	+	AT2G39340	-0.13626	0.025974
Chr2	18028319	18028449	-	AT2G43410	0.153922	0.017982
Chr2	19360450	19360526	-	AT2G47160	0.143354	0.02997
Chr3	392355	392448	+	AT3G02150	0.072209	0.027473
Chr3	3777224	3777303	+	AT3G11930	0.057088	0.043956
Chr3	4304100	4304206	+	AT3G13300	0.256156	0.001499
Chr3	6201686	6201757	+	AT3G18100	0.258251	0.047952
Chr3	7461603	7461693	-	AT3G21250	0.106507	0.035964
Chr3	18690283	18690359	-	AT3G50380	0.088169	0.046953
Chr3	18690665	18690750	-	AT3G50380	0.088169	0.046953
Chr3	18698338	18698590	-	AT3G50380	0.113902	0.046953
Chr3	20177555	20177621	-	AT3G54500	-0.16787	0.006993
Chr3	20178021	20178123	-	AT3G54500	-0.0635	0.036364
Chr3	20178064	20178158	-	AT3G54500	0.094401	0.008159

Chr3	20178064	20178161	-	AT3G54500	0.150194	0.006993
Chr3	20270980	20271074	-	AT3G54760	0.116612	0.021978
Chr3	20748238	20748317	+	AT3G55940	0.1346	0.007992
Chr3	21660146	21660216	+	AT3G58570	0.116783	0.014985
Chr3	21829086	21829166	-	AT3G59060	-0.18318	0.014985
Chr3	21922917	21923011	-	AT3G59310	-0.20111	0.011239
Chr4	7125245	7125336	-	AT4G11840	0.146868	0.015984
Chr4	10696525	10696607	-	AT4G19660	0.139352	0.049451
Chr4	12612953	12613035	+	AT4G24380	-0.0421	0.040959
Chr4	13082090	13082160	-	AT4G25650	0.044811	0.007992
Chr4	15130346	15130419	+	AT4G31115	0.106341	0.024975
Chr4	15658272	15658397	+	AT4G32440	0.144156	0.042707
Chr4	17295297	17295398	-	AT4G36690	0.058352	0.032468
Chr5	4557631	4557707	+	AT5G14120	-0.06236	0.036963
Chr5	6062261	6062346	+	AT5G18310	0.061279	0.048452
Chr5	6834529	6834645	+	AT5G20250	-0.12996	0.037962
Chr5	14815843	14815938	-	AT5G37370	0.265868	0
Chr5	18844142	18844225	+	AT5G46470	0.178074	0.00999
Chr5	21243164	21243243	+	AT5G52310	0.266782	0.00999

4.3.4 Differential DNA methylation is associated with gene promotors and exons

Our DE and DAS results show that DNA hypomethylation increase up-regulated genes and decreases DAS gene upon cold stress compared to Ctrl plants (group C compared to D from figure 1 and 3). Additionally, DNA methylation changes affect gene expression and splicing upon temperature changes rather than steady temperature (contrast group A and B compared to contrast group E and F in figure 3.1 and 3.3). To investigate how genome-wide hypomethylation induced by 5-aza-dC affect s Arabidopsis genome including promoter, exons, and intronic regions under cold stress to modulate gene expression and splicing patterns., Illumina sequencing of bisulphite-converted genomic DNA (Supplementary table 21) of AzadC grown at 22°C and 4°C for 24 hours has been performed, followed by identification of DMRs having q-value higher than 0.05 and methylation difference less than 5%, 1319 DMRs has been identified in the CpG sequence context, of which, 1129 and 190 are hypo- and hyper- DMRs, respectively. In the in the CHG context, 337 DMRs were identified, of which 218 and 121 are hypo- and hyper- DMRs regions, respectively. Whereas in CHH DNA methylation context, only 19 DMRs are found of which 8 are identified as hyper- DMRs (Figure 4.13 A-C, table 4.13-4.15. The rest of the data presented in table 4.14-4.15 can be found in supplementary table 14). Annotation of DMRs with promoters, exons, introns, and intergenic regions shows that most DMRs are associated with promoters exons,

and intergenic regions yet, no DMRs regions were detected in the intronic regions which is likely due to the low level of DNA methylation in plants' intronic regions which might be more pronounced in the case of hypomethylated plants (Figure 4.13 D).

 Overall, DMRs analysis shows that 5-aza-dC treated plants are likely to induce gene upregulation upon cold stress through changes in DNA methylation mostly located in the promoter regions. Additionally, DMRs regions detected in exons and depleted from intronic regions clearly show that DNA methylation is likely to influence exon definition upon cold stress which may subsequently affect splice site recognition and splicing. Intergenic regions which were previously called as 'junk DNA' are now emerging as gene expression regulators in Arabidopsis, such as, transposable elements as well as enhancers of gene expression. Interestingly, 16% of the DMRs in the CpG context were identified in the intergenic regions (Figure 4.13 D) and may be involved in regulating gene expression under cold stress. Collectively, DMRs analysis shows that upon cold stress DNA methylation regulates gene expression probably through defining exons and promoter accessibility to the splicing and transcription machinery, respectively. This is mainly due to the interplay of DNA methylation and nucleosome occupancy as main epigenetic features.
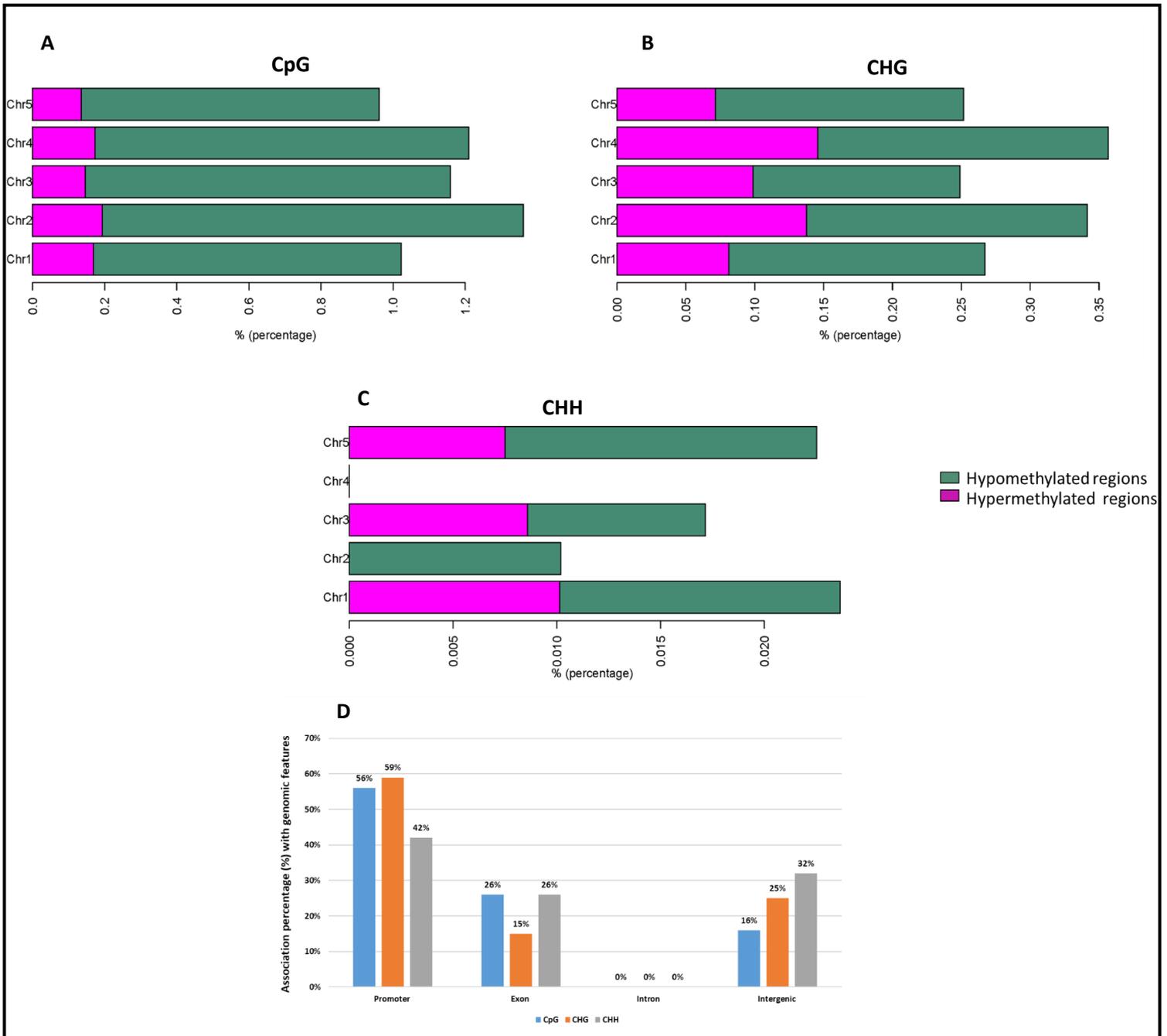
Figure 4.13. Differentially methylated regions of 1000bp window and 1000bp step size and their gene annotation. (A-C) Significant hyper- and hypo- differentially methylated regions (DMRs) in AzadC under cold stress compared to AzadC grown at normal temperature across all chromosomes (Ch1-Chr5) in the CpG (A), CHG (B), and CHH (C) sequence contexts. The x axis represents methylation percentage difference; the y axis represents the chromosome on which DNA methylation changes were detected. For all DNA methylation sequence contexts, hypomethylated regions were higher than hypermethylated ones for all chromosomes (except Chr4 in CHH sequence context that didn't show any significant changes in DNA methylation). **(D)** The percentage of DMRs overlapping with genomic features (promoter, exons, introns, and intergenic) in the three sequence contexts CpG, CHG, and CHH show that DMRs are majorly located in the promoter regions (56%, 59%, and 42% of the total DMRs detected in CpG, CHG, and CHH respectively), exons (26%, 15%, and 26% of the total DMRs detected in CpG, CHG, and CHH, respectively), and intergenic regions (16%, 25%, and 32% of the total DMRs detected in CpG, CHG, and CHH, respectively).

167

Table 4.13. The most significant (Top 50, qvalue < 0,05 and methylation difference > 5%) differentially methylated regions detected in AzadC under cold compared to AzadC under temperature in genomic window of 1000bp with a step size of 1000 bp in the three sequence CpG contexts.

Chromosome	Start	End	pvalue	qvalue	Methylation difference
Chr1	4621001	4622000	5.60E-61	6.19E-57	-13.9783
Chr1	7724001	7725000	1.66E-26	3.77E-23	-7.66923
Chr1	13640001	13641000	1.21E-23	2.22E-20	-8.04341
Chr1	13959001	13960000	2.24E-20	3.00E-17	-5.29231
Chr1	24275001	24276000	1.71E-19	2.19E-16	12.03302
Chr1	23316001	23317000	3.15E-19	3.87E-16	-8.68205
Chr2	2060001	2061000	8.64E-112	2.55E-107	18.25021
Chr2	3295001	3296000	2.91E-100	6.43E-96	-6.40746
Chr2	2282001	2283000	3.76E-64	5.54E-60	12.73773
Chr2	5431001	5432000	1.81E-62	2.29E-58	14.65602
Chr2	3188001	3189000	1.74E-58	1.71E-54	18.25921
Chr2	3348001	3349000	6.27E-44	3.96E-40	-5.83477
Chr2	3320001	3321000	1.67E-36	8.66E-33	-5.75695
Chr2	4496001	4497000	1.04E-34	4.61E-31	-6.02048
Chr2	8855001	8856000	1.27E-34	5.33E-31	-10.1852
Chr2	4993001	4994000	8.52E-34	3.28E-30	-6.84861
Chr2	3001	4000	3.25E-30	1.06E-26	-5.94176
Chr2	3189001	3190000	4.72E-29	1.44E-25	11.91688
Chr2	4212001	4213000	4.38E-28	1.25E-24	-7.19378
Chr2	5852001	5853000	3.78E-27	9.54E-24	-13.2937
Chr2	5367001	5368000	7.78E-26	1.68E-22	-7.46702
Chr2	1949001	1950000	3.00E-24	6.03E-21	-8.141
Chr2	3319001	3320000	9.38E-24	1.80E-20	-5.01542
Chr2	13560001	13561000	1.09E-23	2.05E-20	-10.1272
Chr2	6001	7000	1.65E-20	2.28E-17	-7.36975
Chr3	14195001	14196000	2.11E-94	3.73E-90	-7.82034
Chr3	13540001	13541000	1.63E-57	1.44E-53	11.16606
Chr3	13539001	13540000	1.58E-33	5.81E-30	9.001175
Chr3	14249001	14250000	7.99E-31	2.72E-27	-5.04829
Chr3	1555001	1556000	2.50E-27	6.69E-24	8.581966
Chr3	12112001	12113000	2.86E-27	7.43E-24	-5.5986
Chr3	17206001	17207000	5.58E-27	1.37E-23	-6.93754
Chr3	13541001	13542000	8.99E-27	2.09E-23	8.509359
Chr3	16280001	16281000	2.87E-26	6.34E-23	-6.80305
Chr3	7906001	7907000	2.32E-23	4.19E-20	-6.22814
Chr3	13538001	13539000	2.47E-23	4.36E-20	8.863283
Chr3	18028001	18029000	2.98E-23	5.16E-20	-8.95116
Chr3	22790001	22791000	3.31E-23	5.62E-20	-7.64726
Chr3	16322001	16323000	4.65E-23	7.61E-20	-7.32062

Chr3	8368001	8369000	1.95E-22	3.03E-19	-9.44101
Chr3	21662001	21663000	1.80E-20	2.45E-17	-10.6358
Chr4	6426001	6427000	8.85E-27	2.09E-23	-9.14494
Chr4	2049001	2050000	1.15E-22	1.85E-19	-5.42987
Chr4	132001	133000	4.98E-21	7.22E-18	6.189459
Chr4	1774001	1775000	1.01E-20	1.41E-17	-7.05365
Chr4	9509001	9510000	4.87E-19	5.90E-16	-5.54957
Chr5	11765001	11766000	4.86E-161	4.30E-156	19.26809
Chr5	10358001	10359000	1.51E-24	3.11E-21	-10.6506
Chr5	4371001	4372000	3.01E-22	4.59E-19	-6.54406

Table 4.14. The most significant (Top 50, qvalue < 0,05 and methylation difference > 5%) differentially methylated regions detected in AzadC under cold compared to AzadC under tempereature in genomic window of 1000bp with a step size of 1000 bp in the three sequence CHG contexts.

Chromosome	Start	End	pvalue	qvalue	Methylation difference
Chr1	4621001	4622000	1.8E-36	5.71E-32	-9.84182
Chr1	15095001	15096000	1.34E-16	6.69E-13	-5.80481
Chr1	6715001	6716000	1.6E-13	4.59E-10	-7.06136
Chr1	24275001	24276000	5.85E-13	1.39E-09	8.319987
Chr1	25343001	25344000	8.08E-13	1.83E-09	-7.29965
Chr1	27515001	27516000	2.82E-12	5.35E-09	-7.86371
Chr1	22829001	22830000	6.27E-11	8.9E-08	-6.26874
Chr1	11789001	11790000	1.3E-10	1.67E-07	6.752578
Chr1	18847001	18848000	5.2E-10	5.36E-07	6.244746
Chr2	3353001	3354000	5.75E-14	1.71E-10	-7.07166
Chr2	8855001	8856000	4.17E-13	1.07E-09	-7.67111
Chr2	5431001	5432000	1.15E-12	2.55E-09	5.933625
Chr2	752001	753000	1.64E-12	3.55E-09	-6.4736
Chr2	2865001	2866000	1.5E-11	2.37E-08	6.924962
Chr2	5566001	5567000	9.24E-11	1.22E-07	-6.48048
Chr2	4415001	4416000	4.73E-10	5.11E-07	-5.02331
Chr2	19414001	19415000	7.99E-10	7.83E-07	6.143098
Chr3	22790001	22791000	4.47E-16	2.12E-12	-5.17262
Chr3	1555001	1556000	3.93E-14	1.25E-10	6.842469
Chr3	760001	761000	5.06E-13	1.27E-09	-6.16472
Chr3	12520001	12521000	2.87E-12	5.35E-09	-6.5023
Chr3	12162001	12163000	5.46E-10	5.52E-07	-5.48641
Chr3	10514001	10515000	6.71E-10	6.71E-07	-6.43012
Chr4	13123001	13124000	2.03E-14	7.44E-11	-8.34941
Chr4	7611001	7612000	4.98E-14	1.52E-10	8.758672
Chr4	5540001	5541000	2.67E-13	7.45E-10	7.319985
Chr4	14283001	14284000	3.38E-13	8.92E-10	-7.58851

Chr4	5581001	5582000	7.68E-13	1.78E-09	-6.06396
Chr4	2878001	2879000	2.37E-12	4.9E-09	-5.91402
Chr4	5474001	5475000	2.74E-12	5.35E-09	5.949482
Chr4	7721001	7722000	8E-12	1.37E-08	-7.17363
Chr4	6426001	6427000	4.01E-11	5.86E-08	-5.23412
Chr4	4002001	4003000	4.98E-11	7.17E-08	-7.81767
Chr4	4041001	4042000	1.8E-10	2.2E-07	-5.61486
Chr4	3972001	3973000	2.28E-10	2.71E-07	7.151393
Chr4	7688001	7689000	3.85E-10	4.36E-07	-7.19902
Chr4	16058001	16059000	4.8E-10	5.13E-07	-5.10465
Chr4	8328001	8329000	8E-10	7.83E-07	-5.64926
Chr4	5422001	5423000	9.47E-10	8.82E-07	-5.9514
Chr5	8749001	8750000	2.39E-18	1.62E-14	-9.86857
Chr5	19198001	19199000	1.36E-15	6.13E-12	-6.56226
Chr5	18146001	18147000	1.81E-15	7.81E-12	-9.17415
Chr5	139001	140000	4.08E-15	1.62E-11	-7.13747
Chr5	11765001	11766000	2.5E-14	8.8E-11	6.617325
Chr5	15919001	15920000	2.78E-14	9.44E-11	-6.05371
Chr5	10358001	10359000	2.03E-10	2.44E-07	-6.51363
Chr5	17135001	17136000	2.47E-10	2.9E-07	6.19561
Chr5	9739001	9740000	4.64E-10	5.07E-07	-6.06126
Chr5	13455001	13456000	5.02E-10	5.3E-07	6.012129

Table 4.15. The most significant (Top 50, qvalue < 0,05 and methylation difference > 5%) differentially methylated regions detected in AzadC under cold compared to AzadC under tempereature in genomic window of 1000bp with a step size of 1000 bp in the three sequence CHG contexts.

Chromosome	Start	End	pvalue	qvalue	Methylation difference
Chr1	4621001	4622000	5.6E-61	6.19E-57	-13.9783
Chr1	7724001	7725000	1.66E-26	3.77E-23	-7.66923
Chr1	13640001	13641000	1.21E-23	2.22E-20	-8.04341
Chr1	13959001	13960000	2.24E-20	3E-17	-5.29231
Chr1	24275001	24276000	1.71E-19	2.19E-16	12.03302
Chr1	23316001	23317000	3.15E-19	3.87E-16	-8.68205
Chr2	2060001	2061000	8.6E-112	2.5E-107	18.25021
Chr2	3295001	3296000	2.9E-100	6.43E-96	-6.40746
Chr2	2282001	2283000	3.76E-64	5.54E-60	12.73773
Chr2	5431001	5432000	1.81E-62	2.29E-58	14.65602
Chr2	3188001	3189000	1.74E-58	1.71E-54	18.25921
Chr2	3348001	3349000	6.27E-44	3.96E-40	-5.83477
Chr2	3320001	3321000	1.67E-36	8.66E-33	-5.75695
Chr2	4496001	4497000	1.04E-34	4.61E-31	-6.02048
Chr2	8855001	8856000	1.27E-34	5.33E-31	-10.1852

Chr2	4993001	4994000	8.52E-34	3.28E-30	-6.84861
Chr2	3001	4000	3.25E-30	1.06E-26	-5.94176
Chr2	3189001	3190000	4.72E-29	1.44E-25	11.91688
Chr2	4212001	4213000	4.38E-28	1.25E-24	-7.19378
Chr2	5852001	5853000	3.78E-27	9.54E-24	-13.2937
Chr2	5367001	5368000	7.78E-26	1.68E-22	-7.46702
Chr2	1949001	1950000	3E-24	6.03E-21	-8.141
Chr2	3319001	3320000	9.38E-24	1.8E-20	-5.01542
Chr2	13560001	13561000	1.09E-23	2.05E-20	-10.1272
Chr2	6001	7000	1.65E-20	2.28E-17	-7.36975
Chr3	14195001	14196000	2.11E-94	3.73E-90	-7.82034
Chr3	13540001	13541000	1.63E-57	1.44E-53	11.16606
Chr3	13539001	13540000	1.58E-33	5.81E-30	9.001175
Chr3	14249001	14250000	7.99E-31	2.72E-27	-5.04829
Chr3	1555001	1556000	2.5E-27	6.69E-24	8.581966
Chr3	12112001	12113000	2.86E-27	7.43E-24	-5.5986
Chr3	17206001	17207000	5.58E-27	1.37E-23	-6.93754
Chr3	13541001	13542000	8.99E-27	2.09E-23	8.509359
Chr3	16280001	16281000	2.87E-26	6.34E-23	-6.80305
Chr3	7906001	7907000	2.32E-23	4.19E-20	-6.22814
Chr3	13538001	13539000	2.47E-23	4.36E-20	8.863283
Chr3	18028001	18029000	2.98E-23	5.16E-20	-8.95116
Chr3	22790001	22791000	3.31E-23	5.62E-20	-7.64726
Chr3	16322001	16323000	4.65E-23	7.61E-20	-7.32062
Chr3	8368001	8369000	1.95E-22	3.03E-19	-9.44101
Chr3	21662001	21663000	1.8E-20	2.45E-17	-10.6358
Chr4	6426001	6427000	8.85E-27	2.09E-23	-9.14494
Chr4	2049001	2050000	1.15E-22	1.85E-19	-5.42987
Chr4	132001	133000	4.98E-21	7.22E-18	6.189459
Chr4	1774001	1775000	1.01E-20	1.41E-17	-7.05365
Chr4	9509001	9510000	4.87E-19	5.9E-16	-5.54957
Chr5	11765001	11766000	4.9E-161	4.3E-156	19.26809
Chr5	10358001	10359000	1.51E-24	3.11E-21	-10.6506
Chr5	4371001	4372000	3.01E-22	4.59E-19	-6.54406

4.3.5 Splicing ratios are strongly modulated by nucleosome occupancy levels

To further explore the relationship between nucleosome occupancy and exon inclusion levels between different AS events, PSI values detected for each AS event type were grouped into four bins and aligned nucleosome peaks 200 bp upstream and downstream relative to the 3' splice sites of the exon or intron (Supplementary Data Set 7). It is notable that for ES, A3'SS and A5'SS, exons with higher inclusion levels display more nucleosome occupancy, whereas retained introns with higher inclusion levels display less nucleosome occupancy (Figure 4.14

A). Additionally, nucleosome occupancy levels decreased in AzadC plants compared to Ctrl and upon cold treatment regardless of exon inclusion level and the AS event type. For mCG profiles around exons (for A5'SS, A3'SS, ES events) and introns (for IR events including exitrons) for which the PSI values are calculated, we found strong associations with different PSI values among AS events (Figure 4.14 B). Interestingly, overall nucleosome patterns and mCG profiles are not affected by PSI values; however, there is clear difference between nucleosome occupancy levels. Alternatively, exons may be associated with specific epigenetic features and their levels are likely to influence local splicing events and abundance of transcripts.

B

Col-0 22°C

Col-0 4°C

AzadC.Col-0 22°C

AzadC.Col-0 4°C

Legend: PSI ≤ 20% | 20% <PSI≤50% | 50%<PSI ≤80% | PSI ≥80%

ES   A5'SS   CpG methylation   A3'SS   RI

AzadC.Col-0 22°C
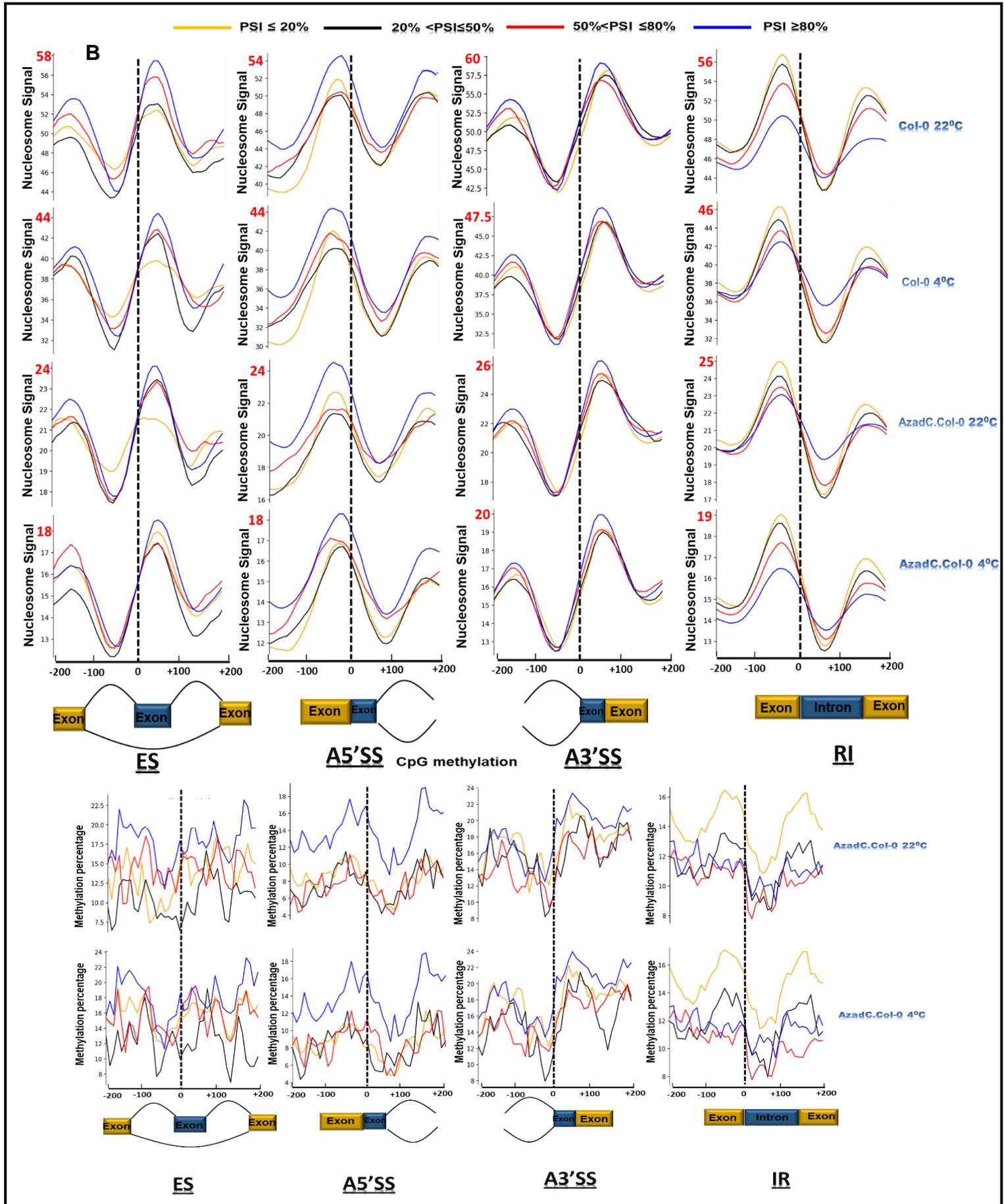
AzadC.Col-0 4°C

ES   A5'SS   A3'SS   IR

Figure 4.14. General nucleosome profiles (A) and DNA methylation level (B) aligned to the 3'SS of exons involved in different AS events grouped according to PSI index, and flanking sequences. The x-axis is the position relative to the acceptor site; the y-axis is the nucleosome signal density for A and CG percentage for B. ES: Exon skipping, A3'SS: Alternative 3'SS, A5'SS: Alternative 5'SS, and IR: Intron retention. Exons or introns for which the PSI value is represented are coloured in blue, whereas exons involved in the splicing event are coloured in yellow. Diagonal lines indicate a splicing event.

To further investigate the distinctive patterns of nucleosome occupancy for local AS with significant change in the mean distribution of PSI values between samples (p value ≤ 0.05), five AS events meeting these criteria were selected. Then nucleosome occupancy were profiled across their splice sites [I4R Chr4:15502301-15502512) and A5 (Chr4:15502446-15502512) of the MITOCHONDRIAL CALCIUM UPTAKE (*MICU)* gene, A3'SS (Chr5:4656173-4656498) of the SURGEIT LOCUS PROTEIN 2 *(SURF2)* gene, ES (Chr4:7885064-7885147) of the HTA4 gene, and retained EI (Chr2: 15,743,353-15, 743429 bp) (Figure 4.15). Interestingly, we found that AzadC plants display a significant increase (p ≤ 0.05) in the PSI for the different events compared to Ctrl at 4°C and 22°C (Box plots represented in figure 4.15). This was further confirmed by an increase of transcripts ratios involved in the AS event in AzadC compared to Ctrl (Bar plots represented in figure 4.15). Based on these observations, we reasoned that lower DNA methylation levels in AzadC may reduce nucleosome positioning resulting in "missplicing" of upstream introns and subsequently their retention as well as reduced exon definition leading to their skipping. Indeed, this was observed in the case of *MICU* gene where AzadC plants displayed an increase in the inclusion level of intron 4 in the (Figure 4.15 A [IR event box plot] group A and B), and increase in exon skipping pin HT4A gene compared to Ctrl at 22°C and 4°C (Figure 4.15 B [IR event box plot] group A and B). This was confirmed as well by our nucleosome profiles across the coordinates of these AS events (nucleosome profiles in Figure 4.15 A [IR] and B), showing that nucleosome occupancy levels are lower in AzadC compared to Ctrl plants. Interestingly, results from figure 4.9 show that alternatively spliced exons and introns display lower nucleosome occupancy and DNA methylation compared to constitutively spliced ones, which was true for the alternative spliced sites as well. We reason that the significant increase (p ≤ 0.05) in the inclusion levels of exons involved in the A5'SS and A3'SS events in the case of *MICU* and *SURF2,* respectively in the case of AzadC plants compared to Ctrl at 4°C and 22°C (Figure 4.15 A [A5] C2 and C4) are likely due to DNA hypomethylation in AzadC plants. This is further confirmed by nucleosome occupancy profiles around the alternative splice sites of A5'SS and A3'SS events (nucleosome profiles in figure 4.15 A [A5] and C), showing that nucleosome occupancy levels are lower in AzadC compared to Ctrl plants.

Interestingly, retained exitrons (Figure 4.15 D) show significant differential AS event in Ctrl plants shifted from normal to cold stress; however, this change is only marginally significant in the case of AzadC plants subjected to the same temperature shift. Importantly, this difference in AS of this EI abundance between AzadC and Ctrl for the same temperature shift was associated with different nucleosome profile and levels (Figure 4.15 D). This lead to the hypothesis that differences in epigenetic features mediated by AzadC treatment are not only sufficient to modulate alternative splice site selection, but potentially regulate the abundance of RNAPII accumulation to modulate the ratio of transcripts involved in these local events. Further experiments are needed to illuminate the relationship between nucleosome occupancy and RNAPII processivity to mediate splicing outcomes. Additionally, lower DNA methylation levels are likely to reduce exon definition and subsequently influence RNAPII processivity around splice sites, thus leading to exon skipping and intron retention.
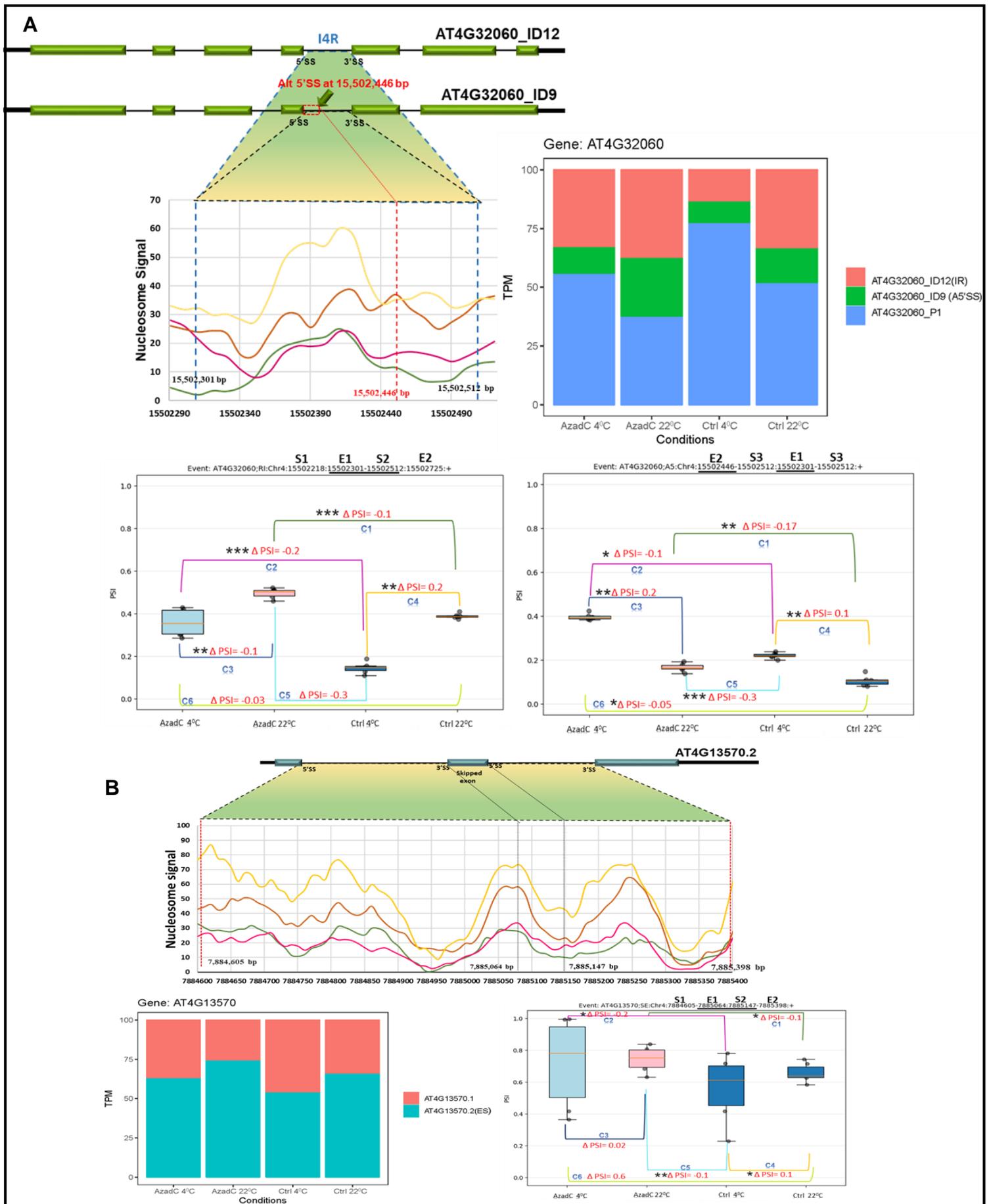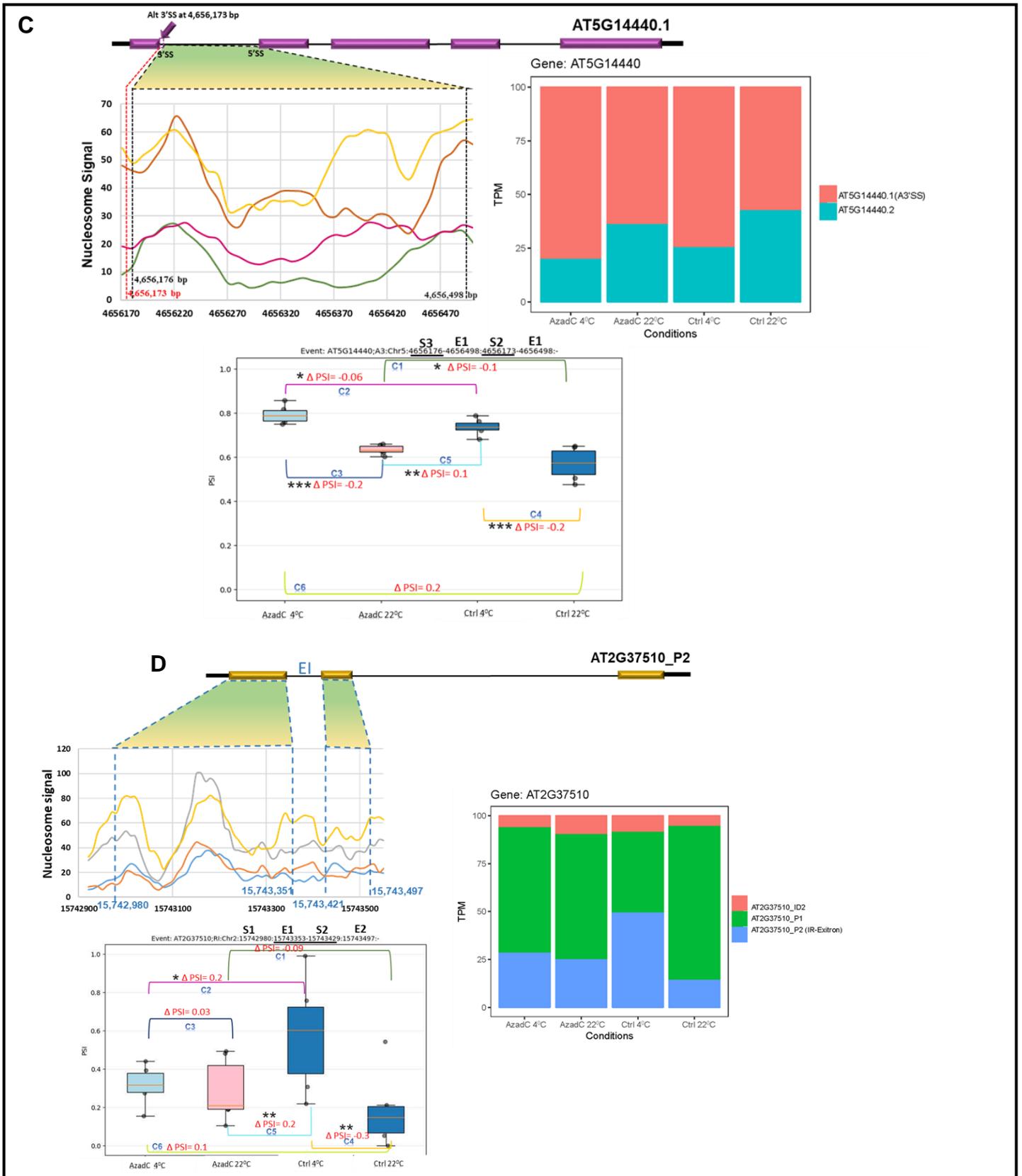
Figure 4.15. Illustration of nucleosome occupancy profiles extended from the donor and acceptor alternative splice sites for different AS events. Nucleosome profile across **(A)** the retained intron (Chr4:15,502,301-15,502,512 bp) of AT4G32060 (MICU) gene (Transcript AT4G32060_ID12 AT4G32060_ID11), and alternative 5'SS (Alt 5'SS) (Chr4:15,502,446-15,502,512 bp) of AT4G32060 (MICU) gene (Transcript AT4G32060_ID9), **(B)** Alternative 3'SS (Alt 3'SS) (Chr5:4,656,173-4,656,498 bp) of

AT5G14440 (SURF2) gene (Transcripts AT5G14440.1) and AT5G14440_ID4), **(C)** Exon skipping (Chr4:7,884,605-7,885,398bp) of AT4G13570 (HTA4) gene (Transcript AT4G13570.2), and **(D)** the retained exitron (Chr2: 15,743,353-15,743429 bp) of AT2G37510 (RNA binding protein) gene transcript (AT2G37510_P2). Authentic and alternative splice site regions are indicated by dotted black and red lines, respectively. Only transcripts involved in the AS event are presented. 3' and 5'UTR are represented by thick black lines. Exons and introns are presented by green boxes and thin black lines, respectively. Only transcripts involved in the alternative splicing events are represented. Below each nucleosome profile, a box-plot represents the relative inclusion level (PSI) of each AS event in each sample alongside the differential expression (DeltaPSI) in different contrast groups with their significance. Significant differences are labelled with asterisks (P > 0.05, *P < 0.05, **P < 0.01, and ***P < 0.001). Error bars indicate sd, n = 3 biological replicates. The x-axis is the PSI; the y-axis is the different conditions. The start (s) and end (e) coordinates for different exonic regions (1-3) involved in the event are indicated. The external coordinates of the event are only used for the RI (please refer to SUPPA manual for event details). The underlined regions are the ones for which PSI is given. (+/-) indicating the strand on which the event was detected. Alongside nucleosome profile, a bar plot is presented to indicate the transcript involved in each AS event and their ratios/variation across different conditions. C1 (AzadC 22°C vs Ctrl 22°C), C2 (AzadC 4°C vs Ctrl 4°C), C3 (AzadC 4°C vs AzadC 4°C), C4 (Ctrl 4°C vs Ctrl 22°C), C5 (AzadC 22°C vs Ctrl 4°C), and C6 (Ctrl 22°C vs AzadC 4°C).

4.4 Discussion

The analysis of RNA-seq data in chapter 3 shows that AzadC and Ctrl plants (Plants with differences in DNA methylation patterns) display differences in gene expression and AS profiles under normal as well as cold stress. Hence, this implies that differences in the transcriptome pool between AzadC and Ctrl under the same temperature conditions, as well as changes in gene expression/AS upon cold stress in AzadC and Ctrl are hypothetically associated with changes in epigenetic features (i.e nucleosome occupancy and DNA methylation). To investigate this hypothesis, in this chapter, MNase-seq of Ctrl and AzadC plants grown at 22°C and subject to 4°C for 24 hours has been performed, as well as, WGBS of AzadC grown at 22°C and subject to 4°C for 24 hours.

MNase digestion combined with high-throughput sequencing method is considered as strong technique to determine genome-wide nucleosome positioning. Additionally, WGBS data reveals genome-wide changes of methylation patterns. In this chapter, iNPS, DANPOS, as well as deeptools3 were the most reliable tools to perform WGBS data analysis. iNPS was used to detect nucleosome positioning due to its capacity of resolving technical problems of other nucleosome positioning detection tools. Furthermore, DANPOS was compulsory algorithmic module to integrate in the analysis presented in this chapter to obtain DPNs whereas, deeptools3 helped in profiling nucleosome profiles across various genomic features. Furthermore, WGBS data analysis has been done using Bismark, Methylkit, and deeptools. Bismark provides the best option to map WGBS paired-end reads to the reference genome; taking into consideration bisulphite conversion of the genomic DNA performed during the library preparation. Additionally, Methylkit was the tool of choice to extract methylation calls in the three methylation context, to detect and annotate DMRs.

The analysis of MNase-seq and WGBS data in this chapter is important to (1): Profile nucleosome occupancy patterns around SJs and exon/intron in AzadC and Ctrl, (2) Detect if nucleosome occupancy change in response to cold stress, and if these profiles differ between AzadC and Ctrl plants, (3) Detect if DNA methylation profiles are similar to those detected for nucleosome.

Collectively, this analysis will help to understand to which extent transcriptome changes upon cold stress are associated with genome-wide changes in epigenetic features, and will reveal the importance of chromatin remodelling upon environmental changes to modulate gene expression.

The most important result reported here is that nucleosome occupancy patterns remains the same while their levels change; nucleosome occupancy levels were lower in plants subject to cold stress and AzadC plants compared to Ctrl regardless of the temperature treatment. This provide an evidence nucleosome patterns around exons, introns, and the spice sites remains the same to maintain exon/intron definition. Yet, these patterns change their levels upon different environmental conditions, which could affect the accumulation of RNAPII and subsequently splice site selection as well as the recruitment of SFs. Indeed, this is further supported by the second attractive result of this chapter showing that different AS events display different levels of nucleosome occupancy in each sample, which could affect splice site selection and subsequently the frequency of AS event under certain environmental condition.

The strong association of nucleosome with exon and their depletion from intronic regions may act as a "speed bump", slowing RNAPII elongation and leading to an increase in the inclusion level of that exon. This was further demonstrated by the fact that exons with higher inclusion levels display more nucleosome occupancy compared to exons to less inclusion levels in ES, A5, and A3 events. Interestingly, DNPs strongly suggests that nucleosome positioning change under cold stress in AzadC and Ctrl plants as well as between plants displaying methylation differences, which implies the strong connection between DNA methylation and nucleosome occupancy. Indeed, this is further supported by the CpG DNA methylation profiles that mirror nucleosome occupancy around exons, introns, and SJs. The association between DNA methylation profiles and nucleosome occupancy was also true for the different AS events and the different inclusion levels. Finally, the individual examples presented in figure 4.15 shows that changes in the ratios of individual AS events between samples is associated with changes in nucleosome occupancy and patterns.

Collectively, the results presented in this confirm the hypothesis that changes in DNA methylation and nucleosome occupancy are likely to modulate Arabidopsis gene expression and AS profile is response to cold stress. This is supported by the fact that AzadC and Ctrl plants, which displayed in this chapter genome-wide differential nucleosome occupancy under normal as well as cold stress, show as well changes in gene expression and AS profiles. The demonstrations presented here showing that epigenetic features regulate cold stress responses in plants display the complexity of plants responses to stress and open up new horizons to discover the mechanistic details and the networks of such regulation. Future research will reveal more complex and dynamic changes of the chromatin structure such as histone modifications to regulate cold-induced gene expression and splicing in plants. This will bring

us a step forward towards the understanding of co-transcriptional regulation of AS in plants in response to cold stress through epigenetic modifications

References

Akalin, A. *et al.* (2012) 'MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles', *Genome Biology*, 13(10), p.R87 doi: 10.1186/gb-2012-13-10-R87.

Berget, S. M. (1995) 'Exon recognition in vertebrate splicing', *Journal of Biological Chemistry*, 270(6), pp. 2411–2414. doi: 10.1074/jbc.270.6.2411.

Carey M, S. ST. (2005) 'Micrococcal nuclease - Southern blot assay', *Nature Methods*, 2, pp. 719–720. doi: 10.1038/nmeth0905-719.

Chen, K. *et al.* (2013) 'DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing', *Genome Research*, 23(2), pp. 341–51. doi: 10.1101/gr.142067.112.

Chen, W. *et al.* (2014) 'Improved nucleosome-positioning algorithm iNPS for accurate nucleosome positioning from sequencing data', *Nature Communications*, 18(5), p. 4909. doi: 10.1038/ncomms5909.

Chen, W., Luo, L. and Zhang, L. (2010) 'The organization of nucleosomes around splice sites', *Nucleic Acids Research*, 38(9), pp. 2788–2798. doi: 10.1093/nar/gkq007.

Chodavarapu, R. K. *et al.* (2010) 'Relationship between nucleosome positioning and DNA methylation', *Nature*, 466(7304), pp. 388–392. doi: nature09147 [pii]\n10.1038/nature09147.

Churchman, L. S. and Weissman, J. S. (2011) 'Nascent transcript sequencing visualizes transcription at nucleotide resolution', *Nature*, 469(7330), pp. 368–373. doi: 10.1038/nature09652.

Chwialkowska, K. *et al.* (2016) 'Water-deficiency conditions differently modulate the methylome of roots and leaves in barley (Hordeum vulgare L.)', *Journal of Experimental Botany*, 67(4), pp. 1109–1121. doi: 10.1093/jxb/erv552.

Cokus, S. J. *et al.* (2008) 'Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning', *Nature*, 452(7184), pp. 215–219. doi: 10.1038/nature06745.

Dubin, M. J. *et al.* (2015) 'DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation', *eLife*, 4, p. e05255. doi: 10.7554/eLife.05255.

Ehrlich, M. *et al.* (1982) 'Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells', *Nucleic Acids Research*, 10(8), pp. 2709–2721. doi: 10.1093/nar/10.8.2709.

Fincher, J. A. *et al.* (2013) 'Genome-Wide Prediction of Nucleosome Occupancy in Maize (Zea mays L.) Reveals Plant Chromatin Structural Features at Genes and Other Elements at Multiple Scales', *plant physiology*, 162(2), pp. 1127–1141.

Gelfman, S. *et al.* (2013) 'DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure', *Genome Research*, 23(5), pp. 789–799. doi: 10.1101/gr.143503.112.

Hollister, J. *et al.* (2011) 'Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata', *Proceedings of the*, 108(6), pp. 2322–7.

Kalyna, M. *et al.* (2006) 'Evolutionary conservation and regulation of particular alternative

splicing events in plant SR proteins', *Nucleic Acids Research*, 34(16), pp. 4395–4405. doi: 10.1093/nar/gkl570.

Kornblihtt, A. R. (2015) 'Transcriptional control of alternative splicing along time: ideas change, experiments remain.', *RNA (New York, N.Y.)*, 21(4), pp. 670–2. doi: 10.1261/rna.051151.115.

Li-Byarlay, H., Li, Y. and Stroud, H. (2013) 'RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee', *Proceedings of the National Academy of Sciences of the United States of America*, 110(31), pp. 12750–12755. doi: 10.1073/pnas.1310735110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1310735110.

Li, G. *et al.* (2014) 'ISWI proteins participate in the genome-wide nucleosome distribution in Arabidopsis', *Plant Journal*, 10(5), p. e1004378. doi: 10.1111/tpj.12499.

Lister, R. *et al.* (2008) 'Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis', *Cell*, 133(3), pp. 523–536. doi: 10.1016/j.cell.2008.03.029.

Liu, J. *et al.* (2015) 'Genetic and epigenetic control of plant heat responses', *Frontiers in Plant Science*, 06. doi: 10.3389/fpls.2015.00267.

Liu, M.-J. *et al.* (2015) 'Determinants of nucleosome positioning and their influence on plant gene expression', *Genome Research*, 25(8), pp. 1182–1195. doi: 10.1101/gr.188680.114.

Lu, X. *et al.* (2017) 'Single-base resolution methylomes of upland cotton (Gossypium hirsutum L.) reveal epigenome modifications in response to drought stress', *BMC Genomics*, 18(1), p. 297. doi: 10.1186/s12864-017-3681-y.

M. Frommer *et al.* (1992) 'A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands', *Proc. Natl. Acad,* 89(5), pp.1827-31

Marquez, Y. *et al.* (2015) 'Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity', *Genome Research*, 25(7), pp. 995–1007. doi: 10.1101/gr.186585.114.

Mavrich, T. N. *et al.* (2008) 'Nucleosome organization in the Drosophila genome', *Nature*, 453(7193), pp. 358–362. doi: 10.1038/nature06929.

Mei, W. *et al.* (2017) 'Evolutionarily conserved alternative splicing across monocots', *Genetics*, 207(2), pp. 465–480. doi: 10.1534/genetics.117.300189.

Nahkuri, S., Taft, R. J. and Mattick, J. S. (2009) 'Nucleosomes are preferentially positioned at exons in somatic and sperm cells', *Cell Cycle*, 8(20), pp. 3420–3424. doi: 10.4161/cc.8.20.9916.

Peckham, H. E. *et al.* (2007) 'Nucleosome positioning signals in genomic DNA', *Genome Research*, 17(8), pp. 1170–1177. doi: 10.1101/gr.6101007.

Rauch, H. B. *et al.* (2014) 'Discovery and expression analysis of alternative splicing events conserved among plant SR proteins', *Molecular Biology and Evolution*, 31(3), pp. 605–613. doi: 10.1093/molbev/mst238.

Regulski, M. *et al.* (2013) 'The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA', *Genome Research*, 23(10), pp. 1651–1662. doi: 10.1101/gr.153510.112.

Schwartz, S., Meshorer, E. and Ast, G. (2009) 'Chromatin organization marks exon-intron structure.', *Nature Structural & Molecular Biology*, 16(9), pp. 990–995. doi: 10.1038/nsmb.1659.

Shayevitch, R. *et al.* (2018) 'The Importance of DNA Methylation of Exons on Alternative Splicing', *RNA*, 24(10), pp. 1351–1362. doi: 10.1261/rna.064865.117.

Shukla, S. *et al.* (2011) 'CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.', *Nature*, 479(7371), pp. 74–9. doi: 10.1038/nature10442.

Sibley, C. R., Blazquez, L. and Ule, J. (2016) 'Lessons from non-canonical splicing', *Nature Reviews Genetics*, 17(7), pp. 407–421. doi: 10.1038/nrg.2016.46.

Singh, J. and Padgett, R. A. (2009) 'Rates of in situ transcription and splicing in large human genes', *Nature Structural and Molecular Biology*, 16(11), pp. 1128–1133. doi: 10.1038/nsmb.1666.

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

Tilgner, H. *et al.* (2009) 'Nucleosome positioning as a determinant of exon recognition', *Nature Structural and Molecular Biology*, 16(9), pp. 996–1001. doi: 10.1038/nsmb.1658.

Tolstrup, N., Rouzé, P. and Brunak, S. (1997) 'A branch point consensus from Arabidopsis found by non-circular analysis allows for better prediction of acceptor sites', *Nucleic Acids Research*, 25(15), pp. 3159–3163. doi: 10.1093/nar/25.15.3159.

Yang, L. *et al.* (2011) 'Lowly expressed genes in arabidopsis thaliana bear the signature of possible pseudogenization by promoter degradation', *Molecular Biology and Evolution*, 28(3), pp. 1193–203. doi: 10.1093/molbev/msq298.

Y, Li. and O, Tollefsbol. (2011) 'DNA methylation detection: Bisulfite genomic sequencing analysis,', *Methods Mol Biol*, 791, pp. 11-21. doi: 10.1007/978-1-61779-316-5_2.

Zhang, C., Yang, H. and Yang, H. (2015) 'Evolutionary Character of Alternative Splicing in Plants.', *Bioinformatics and biology insights*, 9(1), pp. 47–52. doi: 10.4137/BBI.S33716.

Zhang, R. *et al.* (2017) 'A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing', *Nucleic Acids Research*, 45(9), pp. 5061–5073. doi: 10.1093/nar/gkx267.

Zhang, Y. *et al.* (2008) 'Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq', *BMC Genomics*, 9, p. 537. doi: 10.1186/1471-2164-9-537.

Zhang, Z. *et al.* (2011) '*Arabidopsis* Floral Initiator SKB1 Confers High Salt Tolerance by Regulating Transcription and Pre-mRNA Splicing through Altering Histone H4R3 and Small Nuclear Ribonucleoprotein LSM4 Methylation', *The Plant Cell*, 23(1), pp. 396–411. doi: 10.1105/tpc.110.081356.

Zhou, Y., Lu, Y. and Tian, W. (2012) 'Epigenetic features are significantly associated with alternative splicing', *BMC Genomics*, 13, p. 123. doi: 10.1186/1471-2164-13-123.

Chapter 5. General Discussion and Future Perspectives

5.1 Discussion

Recent evidence from Arabidopsis shows that transcription and the splicing process are coupled (Dolata et al., 2015; Hetzel et al., 2016; Jabre et al., 2019; Ullah et al., 2018) and that DNA methylation and nucleosome occupancy may modulate these processes in a time- and condition-specific manner (L. T. Chen, Luo, Wang, & Wu, 2010; Ullah et al., 2018). Epigenetic features in plants regulate transcriptional activity and differentially mark exons, introns as well as cassette and constitutively spliced exons. Furthermore, RNAPII elongation speed has been found to be slower in nucleosome-rich exons allowing more time for the splicing process to take place (Chodavarapu et al., 2010; J. Zhu et al., 2018). The relationship between DNA methylation, nucleosome occupancy and transcriptional control has been demonstrated in recent years (Chodavarapu et al., 2010; Xutong Wang et al., 2016); however, how the chromatin environment influences the splicing/AS process under variable growth and stress conditions remains elusive in plants. Since splicing/AS regulation is achieved by the context of the *cis*-regulatory sequence as well as the chromatin environment (Reddy et al., 2013), it is important to understand the relative contributions of the genetic and the epigenetic landscape. To interrogate and reveal these contributions towards AS, Ctrl and 5-aza-dC-treated hypomethylated plants have been used to demonstrate that differential DNA methylation and nucleosome occupancy in identical genetic backgrounds are sufficient to modulate gene expression and AS in Arabidopsis (Figure 5.1). Remarkably, 5-aza-dC and cold treatment resulted in a reduction in DNA methylation and nucleosome occupancy levels across all five chromosomes of Arabidopsis in a uniform manner. This reduction in nucleosome occupancy was more pronounced around TSS and TTS and was strongly associated with gene expression levels among groups of genes showing low, medium or high expression levels. Intriguingly, there was very little overlap between DE and DAS genes across all treatments and groups, indicating that despite coupling with the transcriptional and the splicing machinery, SFs recruitment and RNAPII dynamics through differential chromatin context may be important in various growth conditions. Indeed, significant DAS between Ctrl and AzadC plants was detected when shifted to cold temperature, which was accompanied by further downregulation of DNA methylation and nucleosome occupancy levels. The results presented in this work show that epigenetic features are not only involved in modulating the AS type in temperature

changes, but nucleosome and DNA methylation levels are also associated with the abundance of differentially spliced transcripts as higher PSI values are associated with an elevated level of nucleosome occupancy. Interestingly, ΔPSI values for various AS types in different contrast groups indicate significant splicing variation among groups with differences in epigenetic signatures but identical DNA sequences. RNA-seq analysis (DAS, DE, and ΔPSI of multiple AS events) show that DNA methylation is more likely to modulate the transcriptome upon temperature shifts rather than steady temperature, indicating that chromatin signatures are malleable to environmental changes and modulate splicing events in Arabidopsis (Zeng et al., 2019). These results demonstrate that although the chromatin environment provides the context through which splicing is modulated, the crosstalk of the splicing and transcriptional machinery, in a condition-dependent (cold stress in this study) manner, is important. For example, a recent study demonstrated that RNAPII speed can be influenced by growth conditions (light quality) and affect splicing patterns in Arabidopsis (Godoy Herz et al., 2019). The results presented in chapter 3 are also consistent with previous findings in rice where only 7% of AS events are influenced by global changes in methylation, hence may play a fine-tuning role under normal conditions as is evident from the splicing pattern differences between Ctrl and hypomethylated plants in this study. RNA-seq analysis presented in this work also reinforces previous findings that AS is more prevalent under stress and/or variable growth conditions in plants to fine-tune their gene expression patterns and/or knock-down a significant proportion of their constitutive transcripts *via* the production of nonsense transcripts, which would be targets of the nonsense-mediated decay (NMD) pathway (Chaudhary et al., 2019; Jabre et al., 2019; Maria Kalyna et al., 2012). These strategies would allow plants to minimize their energy expenditure for protein production under stress conditions to maintain energy homeostasis (Chaudhary et al., 2019). Although Ctrl and AzadC show pronounced variation in their DNA methylation and nucleosome occupancy profiles (along with splicing differences), it is remarkable that mostly genes that confer cold tolerance show expression and splicing differences. These observations support the hypothesis that despite global differences in nucleosome occupancy levels, only cold-responsive genes show expression and splicing differences. It is tempting to reason that plants may remember previous episodes of stresses via chromatin signatures but largely modulate the expression and splicing of those genes which are actively transcribing and/or whose expression needs to be reduced via the production of non-productive mRNA species (Chaudhary et al., 2019; Ling et al., 2018).

Previous studies have shown that higher nucleosome occupancy promotes exon definition, prevents their skipping, and helps intron removal (Chodavarapu et al., 2010; Lev Maor, Yearim, & Ast, 2015; S. Schwartz et al., 2009). Exon recognition is mainly achieved through accumulating RNAPII in a context-dependent manner around splice sites to enhance SF recruitment and allowing more time for splicing to take place (Chodavarapu et al., 2010; J. Zhu et al., 2018). The results presented in chapter 4 support this notion and it is likely that higher DNA methylation and nucleosome occupancy regulate RNAPII accumulation around splice sites and enable SFs recruitment to facilitate and/or modulate splicing variation. For example, when elongating RNAPII reaches a 3'SS it will encounter speed bumps because of higher nucleosome occupancy in an exon and this reduction in elongation speed may help to recognize the 3'SS. Indeed, kinetic experiments in yeast suggest that RNAPII transiently accumulates at 3'SS, facilitating SFs recruitment, and serves the role of a checkpoint associated with co-transcriptional splicing (Alexander et al., 2010). Interestingly, RNAPII elongation speed in Arabidopsis would be much slower after clearing a 3'SS, and may not provide sufficient time (because of higher speed in plant introns) for RNAPII to recognise the 5'SS. Arguably, therefore, 5'SS splicing dynamics are much more complicated and the scanning splicing machinery has to travel to the branch point/ polypyrimidine tract to complete lariat formation and process 5'SS. Beggs and colleagues proposed that in yeast, initial propensity of splicing is low but increases subsequently to allow accumulation of splicing precursors to improve splicing propensity in subsequent and/or successive reactions (Aitken, Alexander, & Beggs, 2011). Mutations at the 3'SS and 5'SS impact transcription initiation and a mutant 3'SS reduces the first step of co-transcriptional splicing in yeast (Aitken et al., 2011). Similarly, splicing dynamics of the human beta-globin gene which fails to form lariat formation and complete 5'SS when a deletion removes the polypyrimidine tract and AG dinucleotide at the 3'SS (Reed & Maniatis, 1985). It is tempting to speculate that nucleosome occupancy and/or histone decoration may be more important in the 5' regions of exons providing a checkpoint to the elongating RNAPII to help recognise 5'SS, lariat formation and cleavage at the 5'SS and 3'SS. It is evident that efficient splicing/AS is dependent on an optimum RNAPII elongation speed and any variation (slow or fast) results in changes in splicing patterns in humans (Dujardin et al., 2014; Fong et al., 2014). However, it is difficult to define the optimum speed of RNAPII because multiple factors including the genetic and the epigenetic context can influence RNAPII processivity and splicing dynamics. Therefore, further work using lines which differ in nucleosome occupancy but have identical DNA sequence is needed to measure elongation kinetics and splicing dynamics of 3'SS and 5'SS in a time- and chromatin-dependent manner.

Nucleosome occupancy and WGBS data also demonstrate that plants displaying changes in DNA methylation define differential nucleosome occupancy levels and exhibit identical nucleosome patterning (not levels) around splice sites and exons under normal and cold conditions. Remarkably, exitrons also show differential nucleosome occupancy and CG methylation levels, which help to differentiate them from flanking exon and intronic regions. Although nucleosome patterns among exons are similar in Arabidopsis, the nucleosome occupancy levels change upon decrease in methylation and temperature, affecting the transcriptional and splicing outcomes. This is remarkable because if exons can be differentiated after hypomethylation and temperature changes from intervening introns, the splicing process should remain relatively robust, considering plant exons are relatively GC rich and facilitate nucleosome formation via DNA methylation (Chodavarapu et al., 2010). However, Mnase-seq data show that perturbations in nucleosome occupancy around splice sites may affect RNAPII processivity and influence splicing/AS patterns. It is possible that chromatin-mediated modulation of RNAPII processivity may also affect its phosphorylation status and SF recruitment to influence splicing/AS decisions (J. Zhu et al., 2018); however, this needs to be explored in future studies. Overall, the results presented in chapter 3 and 4 show that DNA methylation and nucleosome occupancy are connected and work in concert with each other to regulate local AS events (IR, A3'SS, ES, and A5'SS). The results presented here demonstrate that chromatin context is an important controller for the transcriptional and the AS dynamics; however, growth conditions, metabolism and physiology of plants may exert a tight control over desirable expression and splicing patterns. Therefore, reprogramming and preservation of chromatin structure in multiple generations and in response to diverse environmental cues may be more meaningful and provide a context to modulate gene expression patterns (Chaudhary et al., 2019; Jabre et al., 2019). Indirectly, these results also highlight that it is the crosstalk with the transcription and the splicing machinery, in a context-dependent manner, which would ultimately influence the expression and the splicing patterns. This is evident from recent findings that plants possess splicing memory for high temperature conditions, which may be also defined by the chromatin context (Ling et al., 2018). Recent data also shows that temperature-induced differentially spliced genes are enriched in histone H3 lysine 36 tri-methylation (H3K36me3) and any perturbation in these marks affect flowering in Arabidopsis (Pajoro et al., 2017; Steffen and Staiger, 2017). Therefore, chromatin mapping (DNA methylation, nucleosome occupancy, histone modifications) for plants grown under different and recurrent growth and stress conditions needs to be undertaken to reveal the relationship between observed gene expression and splicing patterns to fully understand the underlying

molecular mechanism. Chromatin mapping and splicing analyses of plants growing in diverse conditions and recurring stresses may reveal the extent to which reproducible chromatin patterns are associated with observed AS patterns and have biological significance. Furthermore, the availability of genome-wide profiles of DNA methylation, nucleosome occupancy and associated splicing profiles could open new avenues to engineer desirable crop plants without changing the genetic background.
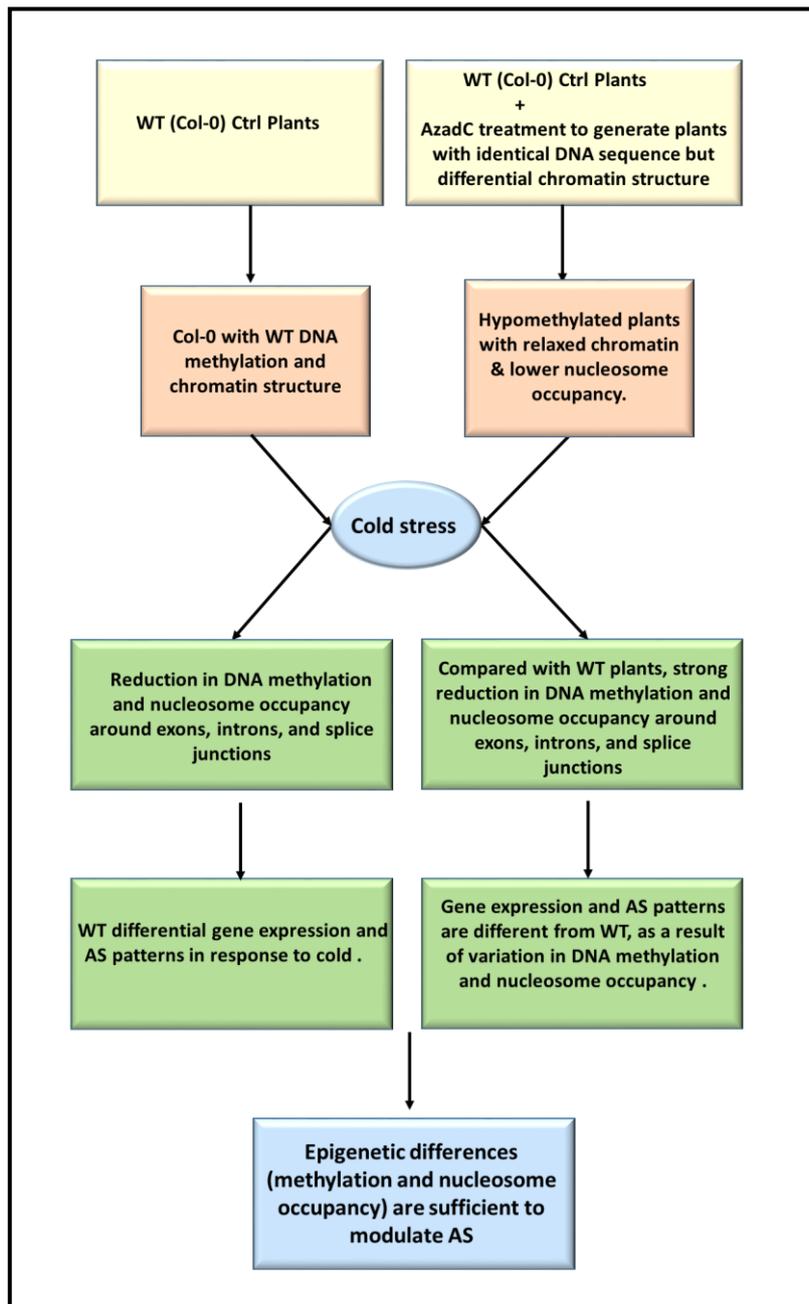
Figure 5.1. Scheme representing the role of epigenetic landscape in modulating cold responsive alternative splicing. Compared to Arabidopsis wild type (WT) Ctrl, AzadC plants display lower DNA methylation and nucleosome occupancy levels in identical genetic background. Upon cold stress, AzadC plants display chromatin re-organisation as a result of hypomethylation and lower nucleosome occupancy, which is coincident with tremendous variation in the expression and splicing patterns of hundreds of genes. Epigenetic differences in genetically identical plants are sufficient to modulate gene expression and AS patterns.

5.2 Future perspective

5.2.1 Engineering splicing variation

RNA Interference (RNAi) has been a gold standard for silencing targeted genes (Fang & Qi, 2016; Mohr, Bakal, & Perrimon, 2010), however, the advent of CRISPR/Cas9 driven strategies have revolutionised the way we could now modulate the expression (and possibly splicing) of single or multiple genes at the DNA level with greater target specificity (Zaidi, Mahfouz, & Mansoor, 2017). Fortunately, the recent development of tissue-specific RNAPII-driven promoter systems, coupled with self-cleaving ribozyme and tRNAs flanking the desired guide RNAs (gRNAs), have made it possible to express gRNAs from any desirable promoter, providing unprecedented cell and tissue specificity (Mahas, Neal Stewart, & Mahfouz, 2018; L. Xu, Zhao, Gao, Xu, & Han, 2017; T. Zhang, Gao, Wang, & Zhao, 2017). Development of Cas9 and Cas13 systems to modulate transcriptional and post-transcriptional dynamics opens up exciting new possibilities to engineer desirable transcriptomes. It is possible that epigenetic context in plants that leads to the expression of a certain number or type of genes may be dependent on the DNA methylation status and chromatin context of genes exhibiting differential expression and splicing patterns. Modulating gene expression patterns in a given generation or at a specific time point is important, however, the challenge is to develop CRISPR arrays, which could modulate the expression and splicing of many genes through multiple generations. Stable inheritance of differentially methylated regions has been demonstrated to mediate extensive phenotypic variation in many traits in plants and contribute to a component of the observed heritability which is explained by the epi-alleles (Johannes et al., 2009). It has now become possible to modulate methylation of target loci using CRISPR/deadCas9 systems coupled with demethylation enzymes to engineer important traits like flowering (Gallego-Bartolomé et al., 2018). Epigenetic engineering could produce the desirable methylation and chromatin context to not only modulate expression and splicing differences, but also maintain the spatiotemporal order and desirable co-regulatory functionality defined by chromatin scaffolds. Since DNA methylation and histone modifications modulate splicing outcomes in concert with RNAPII speed in many species (Carrillo Oesterreich et al., 2016; Dujardin et al., 2014; Shukla & Oberdoerffer, 2012; Xutong Wang et al., 2016), modulating desirable splicing and expression patterns in a tissue and growth specific manner would be desirable and now feasible in plants.

In addition, SJ or exon-specific crRNA could be used to target specific splice isoforms (Mahas et al., 2018; Zaidi et al., 2017) Furthermore, different splicing enhancers and suppressors could be used in multiple arrays to fine-tune desirable splicing variation. Recently, CRISPR/Cas9 mediated homology-directed repair was performed by using a single-stranded oligodeoxynucleotide (ssODNs) to modify the 5' splice site of Xist (encodes a long noncoding RNA affecting chromosomal inactivation of X-chromosomes in females) intron 7 to modulate its splicing efficiency (Yue & Ogawa, 2017). Recent data also show that CRISPR/Cas9 indels could also alter splicing and larger deletions may cause exon skipping (Kapahnke, Banning, & Tikkanen, 2016). Single or multiple gene arrays could be driven by the circadian clock-associated or other tissue- and time-specific promoters to drive metabolic pathways in one or the other direction For example, more vigorous or hybrid plants are efficient at utilizing their starch reserves to fuel growth and usually exhaust them before the onset of dawn (Graf et al., 2010). Genes or spliced isoforms involved in starch metabolism pathway (P. Seo, Kim, Ryu, Jeong, & Park, 2011), could also be targeted via Cas9 or Cas13 systems in single or multiple arrays. Flowering is another important life history trait that could be manipulated to achieve desirable flowering time to reap maximum yields or fit plants into a particular crop rotation system. For example, the FLOWERING LOCUS M (FLM) gene exhibits temperature-dependent AS and inhibits flowering in Arabidopsis and could be targeted to manipulate flowering time (Posé et al., 2013). Recently, CRISPR/Cas9 technology was used to demonstrate the role of the two splice variants of FLM (FLM- β and FLM- δ) by deleting exon 2 and 3 that characterize splice variant β and δ, respectively without any overall change of FLM expression (Capovilla, Symeonidi, Wu, & Schmid, 2017). This scheme resulted in FLM- β that was not able to produce FLM- δ to display late flowering and FLM- β that was able to produce FLM- δ, to display early flowering, indicating the role of splice variant β as a flowering suppressor, whereas the contribution of the second splice variant in flowering time control is improbable (Capovilla et al., 2017). Silencing of particular genes or undesirable spliced isoforms is exciting; however, activation of genes and using their expression as a proxy for translation should be treated with caution because higher expression levels do not correlate well with protein abundance. Future work should explore their turnover and how these levels change in different cells, growth and development conditions (Vogel & Marcotte, 2013).

Since translation and ribosomal loading of transcripts is mediated by the circadian clock and photoperiodic length in plants (Missra et al., 2015; Seaton et al., 2018), the timing of expression should also be taken into consideration to design CRISPR arrays which would coincide with

their natural or wild-type expression context to reap maximum benefits. Even if the translation of a particular protein is desired at the so-called wrong or different time compared to the natural world, Cas13 systems coupled with RNA methylation readers, writers or erasers could be combined to carve desirable methylation patterns to enhance or suppress their translation (Figure 5.2 a and b) (Slobodin et al., 2017). Further refinement of CRISPR/Cas strategies and the availability of versatile vectors and arrays will allow us to target multiple genes for different outcomes simultaneously (Figure 5.2 c) (Cermak et al., 2017; Cong et al., 2013; Mahas et al., 2018; L. Xu et al., 2017; T. Zhang et al., 2017). Finally, although CRISPR systems have revolutionized the way we could edit genomes on a global basis, the chromatin context which may provide timing and regulatory context will remain relevant and the chromatin language need to be understood (Berger, 2007) before engineering biological networks at will.
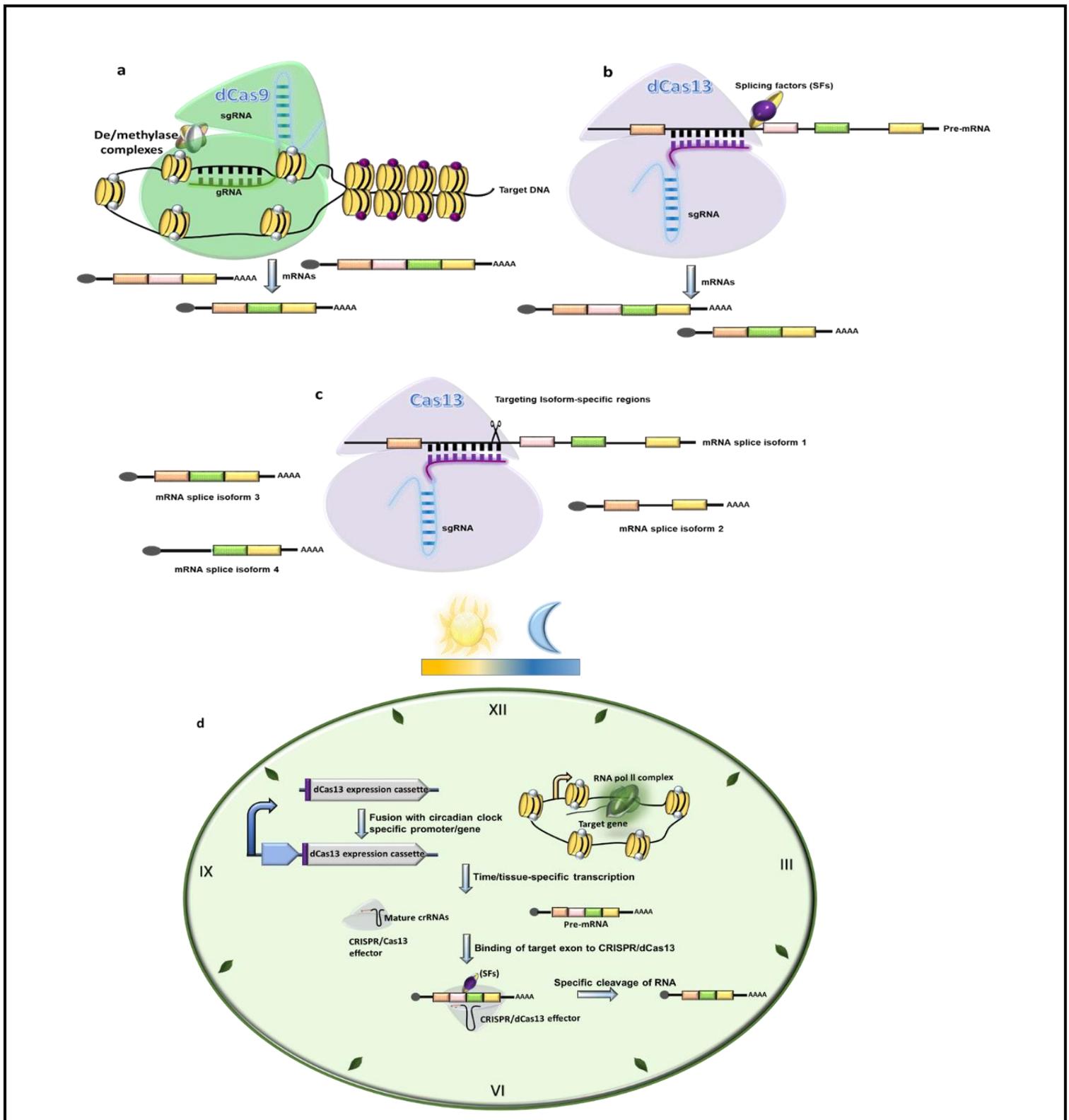
Figure 5.2. Targeted modulation of gene expression and splicing at the co-/post-transcriptional level using CRISPR/Cas9 and CRISPR/Cas13 systems. Transcription and splicing tuning using dCas9 **(a)** dCas9 fused to de/methylase complexes or chromatin re-modellers can target stress-responsive genes to induce relaxed/compact chromatin structure and mediate desirable changes in expression and splicing isoforms/ratios. **(c-b)** Coupling dCas13 to SFs to reprogram pre-mRNA alternative splicing to produce desirable type and quantity of different splice isoforms or targeting specific splice isoform using Cas13. **(d)** Fusion of CRISPR/dCas13 to circadian clock promoter/gene (or tissue-specific promoters) can be used to express target genes/isoforms in a spatiotemporal manner.

194

5.2.2 Isolating normal and stress-specific spliceosomal complexes from constitutive and alternative splice junctions

So far in plants, a few studies show the importance of histone PTMs in recruiting reader complexes, however, it is unknown how changes in histone modifications or slow or fast moving RNAPII recruit different effector proteins to coordinate splice site switching/selection and assemble splicing complexes (H. Li et al., 2006; S. Zhao, Zhang, Yang, Zhu, & Li, 2018). RNA sequencing, the yeast two-hybrid system and co-immunoprecipitation strategies have uncovered some details on how transcriptional and splicing machineries are coupled in space and time (Barash et al., 2010; Churchman & Weissman, 2011; Görnemann, Kotovic, Hujer, & Neugebauer, 2005; Tardiff & Rosbash, 2006). However, isolating spliceosomal complexes in an SJ-specific manner has been always a difficult task to achieve. Here, an alternative strategy is proposed to understand how different histone modifications could help in recruiting different effector proteins to orchestrate different splicing outcomes (Figure 5.3). To investigate how specific histone marks can mediate spliceosome formation on particular SJ, chromatin immunoprecipitation (ChIP) using antibodies to selected histone marks under normal and stressful conditions can be performed as a first step. For this purpose, the approach of photo cross-linking by an amine reactive cross-linking agent SPB (succinimidyl-[4-(psoralen-8-yloxy)]-butyrate) can be used (Wilson, Roebuck, & High, 2008). SPB is an NHS-ester and psoralen cross-linker that effectively conjugates primary amines to DNA via light-activated intercalation next to pyrimidine bases. The photo-coupling of samples will occur after a short exposure to long UV light (Wilson et al., 2008). The advantage of this method is that SPB is stable at high temperatures (Melting point: 177-178°C) allowing the DNA protein complexes to be warmed up to open up DNA structure to trap gene specific DNA-protein complexes. Since the same histone mark can be distributed at different positions of the genome, purification of splicing complexes along a given SJ can be achieved via biotin-labelled full length, as well as smaller amplicons, to cover the entire length of a gene as baits to capture specific DNA-protein complexes. It is possible that all amplicons may not have access to targeted regions, however, cross-linking with SBP may allow experimentation with different temperatures to gain access to a desired region. Captured protein complexes can then be separated using SDS-PAGE followed by in-gel digestion and analysis using mass spectrometry (LC-MS/MS). In contrast, biotin labelled fragments can directly target specific splice SJs after warming up the DNA in case condition-dependant histone marks were not previously identified. In this instance, further analysis can reveal what histone PTM(s) are associated with the purified

complexes. This strategy will illuminate how different protein partners from various time points and stresses can be associated to achieve splice site selection and splicing dynamics of different genes/SJs. Additionally, such a technique can give insights into how protein complexes recruited on a given gene can potentially mediate different splicing outcomes, however, the next challenge will be to find the order in which different effector proteins are recruited in a single gene/SJ specific manner.
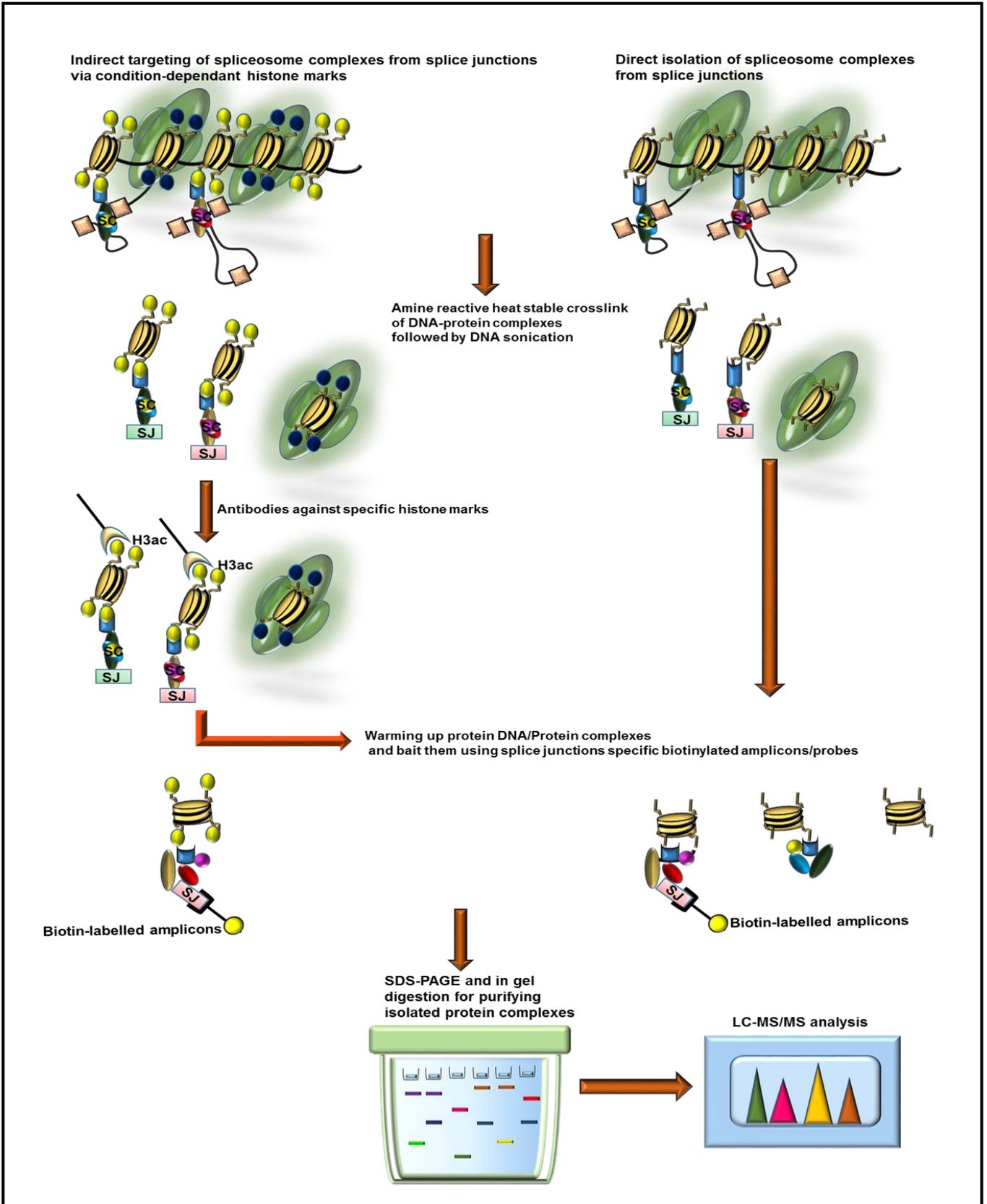
Figure 5.3. Isolation of splice junction specific spliceosome complexes. An illustration of the proposed technique to explain how specific spliceosome complexes formed on exon junctions (EJs) under different conditions can be purified and detected in case of previously known **(a)** or unknown **(b)** condition-dependant histone marks that may mediate splicing.

5.3 Conclusion

DNA methylation and nucleosome occupancy are intimately associated with each other and influence global changes in gene expression and AS patterns in Arabidopsis upon cold stress. The relationship between DNA methylation and nucleosome occupancy is also true for exitrons as they exhibit different profiles compared to flanking exonic regions, which harbour them. Although, nucleosome levels in hypomethylated plants (AzadC) were much lower compared with Ctrl plants; nonetheless, they exhibited similar patterns (not levels) under normal and cold conditions. Even after global hypomethylation, exons could be differentiated from flanking introns due to relatively higher DNA methylation and nucleosome occupancy when compared with Ctrl. These results demonstrate that relative nucleosome occupancy between exons and introns may influence RNAPII processivity and SFs recruitment to modulate AS profiles. Taken together, these results show that epigenetic difference in plants with identical DNA sequence can modulate global expression and AS dynamics in Arabidopsis under variable temperature conditions.

5.4 Outstanding questions

1. To which extent alternatively spliced transcripts are engaged with the ribosomal machinery (partly known) and translated into proteins?

2. How do plants couple their AS events to photoperiodic changes to modulate their proteome upon physiological need through IDPs?

3. What is the impact of chromatin state on transcriptional dynamics, alternative splicing, epitranscriptome and translational efficiency of transcripts in plants?

4. To which extent PTC+ transcripts make truncated but functional proteins?

5. Similar to yeast, is there any regulatory role of plant spliceosomal introns under stress conditions?

References

Aitken, S., Alexander, R. D. and Beggs, J. D. (2011) 'Modelling reveals kinetic advantages of co-transcriptional splicing', *PLoS Computational Biology*, 7(10), p. e1002215. doi: 10.1371/journal.pcbi.1002215.

Alexander, R. D. *et al.* (2010) 'Splicing-Dependent RNA polymerase pausing in yeast', *Molecular Cell*, 40(4), pp. 582–593. doi: 10.1016/j.molcel.2010.11.005.

Barash, Y. *et al.* (2010) 'Deciphering the splicing code', *Nature*, 465(7294), pp. 53–59. doi: 10.1038/nature09000.

Berger, S. L. (2007) 'The complex language of chromatin regulation during transcription', *Nature*, 447(7143), pp. 407–412. doi: 10.1038/nature05915.

Capovilla, G. *et al.* (2017) 'Contribution of major FLM isoforms to temperature-dependent flowering in Arabidopsis thaliana', *Journal of Experimental Botany*, 68(18), pp. 5117–5127. doi: 10.1093/jxb/erx328.

Carrillo Oesterreich, F. *et al.* (2016) 'Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II', *Cell*, 165(2), pp. 372–381. doi: 10.1016/j.cell.2016.02.045.

Cermak, T. *et al.* (2017) 'A multi-purpose toolkit to enable advanced genome engineering in plants', *The Plant Cell*, 29(6), pp. 1196–1217. doi: 10.1105/tpc.16.00922.

Chaudhary, S. *et al.* (2019) 'Perspective on Alternative Splicing and Proteome Complexity in Plants', *Trends in Plant Science*, 24(6), pp. 496–506. doi: 10.1016/j.tplants.2019.02.006.

Chen, L. T. *et al.* (2010) 'Involvement of Arabidopsis histone deacetylase HDA6 in ABA and salt stress response', *Journal of Experimental Botany*, 61(12), pp. 3345–3353. doi: 10.1093/jxb/erq154.

Chodavarapu, R. K. *et al.* (2010) 'Relationship between nucleosome positioning and DNA methylation', *Nature*, 466(7304), pp. 388–392. doi: nature09147 [pii]\n10.1038/nature09147.

Churchman, L. S. and Weissman, J. S. (2011) 'Nascent transcript sequencing visualizes transcription at nucleotide resolution', *Nature*, 469(7330), pp. 368–373. doi: 10.1038/nature09652.

Cong, L. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/VCas Systems', *Science*, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.Multiplex.

Dolata, J. *et al.* (2015) 'NTR1 is required for transcription elongation checkpoints at alternative exons in Arabidopsis', *The EMBO Journal*, 34(4), pp. 544–558. doi: 10.15252/embj.201489478.

Dujardin, G. *et al.* (2014) 'How Slow RNA Polymerase II Elongation Favors Alternative

Exon Skipping', *Molecular Cell*, 54(4), pp. 683–690. doi: 10.1016/j.molcel.2014.03.044.

Fang, X. and Qi, Y. (2016) 'RNAi in Plants: An Argonaute-Centered View', *The Plant Cell*, 28(2), pp. 272–285. doi: 10.1105/tpc.15.00920.

Fong, N. *et al.* (2014) 'Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate', *Genes and Development*, 28(23), pp. 2663–2676. doi: 10.1101/gad.252106.114.

Gallego-Bartolomé, J. *et al.* (2018) 'Targeted DNA demethylation of the *Arabidopsis* genome using the human TET1 catalytic domain', *Proceedings of the National Academy of Sciences*, 115(9), p. 201716945. doi: 10.1073/pnas.1716945115.

Godoy Herz, M. A. *et al.* (2019) 'Light Regulates Plant Alternative Splicing through the Control of Transcriptional Elongation', *Molecular Cell*, 73(5), p. 1066–1074.e3. doi: 10.1016/j.molcel.2018.12.005.

Görnemann, J. *et al.* (2005) 'Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex', *Molecular Cell*, 19(1), pp. 53–63. doi: 10.1016/j.molcel.2005.05.007.

Graf, A. *et al.* (2010) 'Circadian control of carbohydrate availability for growth in Arabidopsis plants at night.', *Proceedings of the National Academy of Sciences*, 107(20), pp. 9458–9463. doi: 10.1073/pnas.0914299107.

Hetzel, J. *et al.* (2016) 'Nascent RNA sequencing reveals distinct features in plant transcription', *Proceedings of the National Academy of Sciences*, 113(43), pp. 12316–12321. doi: 10.1073/pnas.1603217113.

Jabre, I. *et al.* (2019) 'Does co-transcriptional regulation of alternative splicing mediate plant stress responses?', *Nucleic Acids Research*. Oxford University Press, 47(6), pp. 2716–2726. doi: 10.1093/nar/gkz121.

Johannes, F. *et al.* (2009) 'Assessing the impact of transgenerational epigenetic variation on complex traits', *PLoS Genetics*, 5(6), p. e1000530. doi: 10.1371/journal.pgen.1000530.

Kalyna, M. *et al.* (2012) 'Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis', *Nucleic Acids Research*, 40(6), pp. 2454–69. doi: 10.1093/nar/gkr932.

Kapahnke, M., Banning, A. and Tikkanen, R. (2016) 'Random splicing of several exons caused by a single base change in the target exon of CRISPR/cas9 mediated gene knockout', *Cells*, 5(4), p. 45. doi: 10.3390/cells5040045.

Lev Maor, G., Yearim, A. and Ast, G. (2015) 'The alternative role of DNA methylation in splicing regulation', *Trends in Genetics*, 31(5), pp. 274–280. doi: 10.1016/j.tig.2015.03.002.

Li, H. *et al.* (2006) 'Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF', *Nature*, 442(7098), pp. 91–95. doi: 10.1038/nature04802.

Ling, Y. *et al.* (2018) 'Thermopriming triggers splicing memory in Arabidopsis', *Journal of Experimental Botany*, 69(10), pp. 2659–2675. doi: 10.1093/jxb/ery062.

Mahas, A., Neal Stewart, C. and Mahfouz, M. M. (2018) 'Harnessing CRISPR/Cas systems for programmable transcriptional and post-transcriptional regulation', *Biotechnology Advances*, 36(1), pp. 295–310. doi: 10.1016/j.biotechadv.2017.11.008.

Missra, A. *et al.* (2015) 'The Circadian Clock Modulates Global Daily Cycles of mRNA Ribosome Loading', *The Plant Cell*, 27(9), pp. 2582–2599. doi: 10.1105/tpc.15.00546.

Mohr, S., Bakal, C. and Perrimon, N. (2010) 'Genomic Screening with RNAi: Results and Challenges', *Annual Review of Biochemistry*, 79(1), pp. 37–64. doi: 10.1146/annurev-biochem-060408-092949.

Pajoro, A. *et al.* (2017) 'Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants', *Genome Biology*, 18(1), p. 102. doi: 10.1186/s13059-017-1235-x.

Posé, D. *et al.* (2013) 'Temperature-dependent regulation of flowering by antagonistic FLM variants', *Nature*, 503(7476), pp. 414–417. doi: 10.1038/nature12633.

Reddy, A. S. N. *et al.* (2013) 'Complexity of the alternative splicing landscape in plants.', *The Plant cell*, 25(10), pp. 3657–83. doi: 10.1105/tpc.113.117523.

Reed, R. and Maniatis, T. (1985) 'Intron sequences involved in lariat formation during pre-mRNA splicing', *Cell*, 41(1), 95-105. doi: 10.1016/0092-8674(85)90064-9.

Schwartz, S., Meshorer, E. and Ast, G. (2009) 'Chromatin organization marks exon-intron structure.', *Nature Structural & Molecular Biology*, 16(9), pp. 990–995. doi: 10.1038/nsmb.1659.

Seaton, D. D. *et al.* (2018) 'Photoperiodic control of the *Arabidopsis* proteome reveals a translational coincidence mechanism', *Molecular Systems Biology*, 14(3), p. e7962. doi: 10.15252/msb.20177962.

Seo, P. *et al.* (2011) 'Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism', *Nature communications*, 2(303), p. 303.

Shukla, S. and Oberdoerffer, S. (2012) 'Co-transcriptional regulation of alternative pre-mRNA splicing', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(7), pp. 673–683. doi: 10.1016/j.bbagrm.2012.01.014.

Slobodin, B. *et al.* (2017) 'Transcription Impacts the Efficiency of mRNA Translation via

Co-transcriptional N6-adenosine Methylation', *Cell*, 169(2), p. 326–337.e12. doi: 10.1016/j.cell.2017.03.031.

Tardiff, D. F. and Rosbash, M. (2006) 'Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly.', *RNA*, 12(6), pp. 968–979. doi: 10.1261/rna.50506.

Ullah, F. *et al.* (2018) 'Exploring the relationship between intron retention and chromatin accessibility in plants', *BMC Genomics*, 19(1), p. 21. doi: 10.1186/s12864-017-4393-z.

Vogel, C. and Marcotte, E. M. (2013) 'Insights into regulation of protein abundance from proteomics and transcriptomis analyses', *Nature reviews genetics*, 13(4), pp. 227–232. doi: 10.1038/nrg3185.Insights.

Wang, X. *et al.* (2016) 'DNA Methylation Affects Gene Alternative Splicing in Plants: An Example from Rice', *Molecular Plant*, 9(2), pp. 305–307. doi: 10.1016/j.molp.2015.09.016.

Wilson, C. M., Roebuck, Q. and High, S. (2008) 'Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core.', *Proceedings of the National Academy of Sciences of the United States of America*, 105(28), pp. 9534–9. doi: 10.1073/pnas.0711846105.

Xu, L. *et al.* (2017) 'Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA', *Nucleic Acids Research*, 45(5), p. e28. doi: 10.1093/nar/gkw1048.

Yue, M. and Ogawa, Y. (2017) 'CRISPR/Cas9-mediated modulation of splicing efficiency reveals short splicing isoform of Xist RNA is sufficient to induce X-chromosome inactivation', *Nucleic Acids Research*, 46(5), pp. 1–11. doi: 10.1093/nar/gkx1227.

Zaidi, S. S. e. A., Mahfouz, M. M. and Mansoor, S. (2017) 'CRISPR-Cpf1: A New Tool for Plant Genome Editing', *Trends in Plant Science*, 22(7), pp. 550–553. doi: 10.1016/j.tplants.2017.05.001.

Zeng, Z. *et al.* (2019) 'Cold stress induces enhanced chromatin accessibility and bivalent histone modifications H3K4me3 and H3K27me3 of active genes in potato', *Genome Biology.*, 20(1):123. doi: 10.1186/s13059-019-1731-2.

Zhang, T. *et al.* (2017) 'Production of Guide RNAs in vitro and in vivo for CRISPR Using Ribozymes and RNA Polymerase II Promoters', *Bio-protocol*, 7(4), p. 2148. doi: 10.21769/BioProtoc.2148.

Zhao, S. *et al.* (2018) 'Systematic Profiling of Histone Readers in Arabidopsis thaliana', *Cell Reports*, 22(4), pp. 1090–1102. doi: 10.1016/j.celrep.2017.12.099.

Zhu, J. *et al.* (2018) 'RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis', *Nature Plants*. Springer US, 4(12), pp. 1112–1123. doi: 10.1038/s41477-018-

0280-0.