Characterising RNA 5-methylcytosine in *A. thaliana*: functions and regulation pathways

A dissertation submitted for the degree of Doctor of Philosophy (Bioscience)

Xingyu Wu

Supervisor: A/Prof Iain Searle

The University of Adelaide

Faculty of Sciences, Engineering and Technology

School of Biological Sciences

March 2022

Acknowledgments

First and foremost, I am sincerely grateful to my principal supervisor, A/Prof Iain Searle, for his continuous support, invaluable advice and meticulous care during my Master's and PhD journey. His immense knowledge and great experience in the plant molecular area have encouraged me through my academic research and career. Also, his excellent character and rich experience have guided me daily. Thank you, Iain. Your kind help and advice will be an invaluable treasure in my life.

I want to thank my co-supervisor, Dr Steven Polyak and mentor, Dr Dan Peet, for technical support and advice during my project. Thank you, Steven and Dan. Your wise suggestions and kind support helped me to overcome various difficulties.

I want to thank my best friend and workmate, Jun Li， for the bioinformatic analysis of Chapter 2. I cherished the time we lived together, worked in the laboratory, and being with each other. Thank you, Jun, for being my friend forever.

Last but not least, I would like to express my great appreciation to my dear husband， Ziyi Zhao and my family for their unrequited love and unconditional support during my PhD. I am so lucky to have you all in my past, present and future. Thank you.

Many thanks to my colleagues and lab mates who contributed to my research and provided invaluable comments and suggestions.

Many thanks to the University of Adelaide for providing me with a PhD scholarship.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that the copyright of published works contained within this thesis resides with the cop those works' copyright holder(s) so give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signature of student

28/05/2022

Publications

Li J, Wu X, Do T, Nguyen V, Zhao J, Ng PQ, et al. Quantitative and Single-Nucleotide Resolution Profiling of RNA 5-Methylcytosine. Methods Mol Biol. 2021;2298:135–51.

Posters

1. Xingyu Wu, JL, Alice Burgess, Rakesh David, Iain Searle 2017, 'RNA 5-methylcytosine is required for oxidative stress tolerance in *A. thaliana*', poster presented at Lorne Genome 2017, Lorne, Victoria, 3232.

2. Xingyu Wu, JL, Iain Searle 2019, 'RNA 5mC participates in heat shock response in *Arabidopsis thaliana*', paper presented at GSA conference, Melbourne, Victoria, 3010.

Glossary

Ψ	Pseudouridine
2'O-Me	2'-O-methylation
3mC	3-methylcytosine
5caC	5-carboxylcytosine
5mC	5-methylcytosine
5mC-RIP-seq	5-methylcytosine-RNA Immunoprecipitation sequencing
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
7mG	7-methylguanosine
AdoMet	S-adenosyl L-methionine (SAM)
AGO	Argonaute protein
ALKBH	α-ketoglutarate-dependent dioxygenase alkB homolog
ALYREF	Aly/REF export factor protein
A. thaliana	*Arabidopsis thaliana*
ATPCAP1	plasma-membrane-associated cation-binding protein 1

BER	Base excision repair
B. napus	*Brassica napus*
bp	Base pair
BS-RNA-seq	Bisulfite RNA sequencing
C. reinhardtii	*Chlamydomonas reinhardtii*
C5mU	5-carboxymethyluridine
CCS	*Copper chaperone for superoxide dismutase*
cDNA	Complementary DNA
CDS	Coding sequence
C. elegans	*Caenorhabditis elegans*
CGIs	CpG islands
Col-0	Columbia-0 ecotype of *A. thaliana*
CR domain	cysteine-rich domain
CSD	*copper/zinc superoxide dismutase gene*
DLR	Dual-Luciferase Reporter Assay
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase

DSBH	Double-stranded β-helix
dsDNA	Double-stranded DNA
E. coli	*Escherichia coli*
ELAVL1	ELAV Like RNA Binding Protein 1
ES	Embryonic stem cell
F-Luc	Firefly luciferase
FTO	Fat mass and obesity-associated protein
g	Gram
G. barbadense	*Gossypium barbadense*
GEO	Gene Ontology Expression analysis
G. hirsutum	*Gossypium hirsutum*
GLYI4	*Glyoxalase I4*
G. max,	*Glycine max*
GO	Gene ontology
GSPs	Gene-specific primers
H3/4ac	histone H3/4 acetylation
H3K4	Histone H3 lysine 4

H4K5ac	Histone H4 lysine 5 acetylation
HCl	Hydrochloric acid
HEK293 cells	Human embryonic kidney 293 cells
HeLa	Henrietta Lacks
HDGF	*Heparin Binding Growth Factor gene*
HPLC	High-performance liquid chromatography
HO2	*Heme oxygenase 2*
HSP	heat shock protein
HSF	heat shock factor
hr	Hour
Hz	Hertz
IDAX	*Inhibition of the Dvl and Axin complex*
JBP	J-binding protein
Kan	Kanamycin
kV	Kilovolt
LB	Luria-Bertani media
lncRNA	long non-coding RNA

LP	Left border primer
LSU	Large subunit
LTPG5	Glycosylphosphatidylinositol-anchored lipid protein transfer 5
M	Mole
MAG5	*MAIGO 5*
mc5mU	5-methoxycarbonylmethyluridine
MEA	Maximum expected accuracy
MeRIP	Methyl-RNA immunoprecipitation
METTL	Methyltransferase like protein
MG	Methylglyoxal
mg	Microgram
min	Minute
miR156	microRNA156
miRNA	MicroRNA
M. jannaschii	*Methanocaldococcus jannaschii*
mL	Millilitre
mM	MilliMole

mm	Millimetre
mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein
N. benthamiana	*Nicotiana benthamiana*
N1-mA	*N1*-methyladenosine
N1-mG	*N1*-methylguanosine
*N*6-mA	*N*6-methyladenosine
NaAc	Sodium acetate
NaOCl	Sodium hypochlorite
NOP2	Nucleolar protein 2
ncRNA	Non-coding RNA
NPC	Nuclear pore complex
ng	Nanogram
nt	Nucleotide
NSUN	NOP2/Sun domain family member
O. Sativa	*Oryza sativa*
ORF	Open reading frame

PAL1	*Phenylalanine Ammonia-Lyase 1*
PC analysis	Principal Component analysis
PCR	Polymerase chain reaction
PDF2	*Prefoldin 2*
PLB	Passive lysis buffer
Poly(A)	Polyadenylation
PRC2	Polycomb Repressive Complex 2
PS-II	photosystem-II
RBS-seq	RNA Bisulfite sequencing
RBS-amplicon-seq	RNA Bisulfite amplicon sequencing
RBP	RNA binding protein
RCMT	RNA (cytosine-5)-methyltransferase
RFM	Rossmann fold motif
RIP-seq	RNA Immunoprecipitation-sequencing
RISC	RNA-induced silencing complex
R-Luc	Renilla Luciferase
RNA	Ribonucleic acid

RNAi	RNA interference
ROS	Reactive Oxygen Species protein
RP	Right border primer
RPL22A	Ribosomal Protein L22
rRNA	Ribosomal RNA
rt-PCR	Reverse-transcript PCR
S. bicolor	*Sorghum bicolor*
s2U	2-thiouridine
SAM	S-Adenosylmethionine
sec	Second
SF3b complex	Spliceosome factor 3b
SHY2	*Short Hypocotyl 2 gene*
siRNA	Small interfering RNA
snRNA	Small nuclear RA
snoRNA	Small nucleolar RNAs
SPL	*SQUAMOSA promoter-binding protein-like gene*
ssDNA	Single strand DNA

SSU	small subunit
svRNA	Small viral RNA
TCTP1	*Translationally Controlled Tumor protein 1*
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation
TRDMT1	tRNA aspartic acid methyltransferase
tRNA	Transfer RNA
TREX	THO/TREX complex
TRM	tRNA-specific methyltransferase
TSS	Transcriptional start site
Um	2’-O-methyluridine
UTR	Untranslated region
vtRNA	vault ncRNA
XIST	X-inactive specific transcript
YBX1	Y-Box Binding Protein
YTHDF2	YTH domain-containing family protein 2
αKG	α-ketoglutarate

β-OsLCY	*Lycopene β-Cyclase*
µg	Microgram
µL	Microliter
µmol	Micromole

Abstract

5mC is a common post-transcriptional modification identified on cellular mRNAs, tRNAs and ncRNAs in both prokaryotes and eukaryotes. Despite the modification's prevalence, many facets of RNA 5mC remain to be understood. My thesis describes the RNA 5mC landscape after heat shock stress in the plant *Arabidopsis thaliana*. In Chapter 2, by performing RBS-seq, I identified over 5,500 highly confident 5mC sites in the transcriptome-wide after heat shock stress, compared to only 3,000 5mC sites under control conditions. Overall 5mC levels are increased after heat shock stress, especially on mRNAs. I then tested the correlation of 5mC in regulating steady-state mRNA and found no correlation in either the control or heat shock samples which is consistent with other observations. I also investigated the potential roles of heat-response 5mC sites in regulating the expression of Firefly Luciferase (F-Luc). Two 5mC clustered regions on the 5'UTR and exon 2 of *AT1G05340* have roles in maintaining the expression of F-Luc to resist heat shock. In Chapter 3, I attempted to identify 5mC demethylases by identifying mutants of *Alpha-ketoglutarate-dependent dioxygenase AlkB-like* genes and examining the effects of these mutants on the 5mC level of *MAG5* C_{3349} and *AT2G36120* C_{78} under both control and heat shock treatments through the RNA-Bisulfite-amplicon-sequencing (RBS-amplicon-seq). No *atalkb-like* mutant was identified that directly affected 5mC abundance at these two sites, although this experiment was incomplete due to COVID-19 interruptions. Interestingly, *AtALKBH6* (*AT4G20350*) and *AtALKBH8* (putative *AT1G31600*) were found to modulate 5mC levels of *MAG5* C_{3349}, suggesting that RNA 5mC level is under multiple layers of regulation. Taken together, my research reveals a complex regulatory mechanism of RNA 5mC in responding to heat stress in plants, which is worth further investigation.

Table of contents

List of Figures

List of Tables

1. Chapter 1: An overview of cellular RNA 5-methylcytosine

1.1. Overview

Epigenomic regulations play vital roles during multicellular development, cellular differentiation, and adaptation to environmental variation in eukaryotes. Epigenetic processes include DNA modifications, histone modifications, chromatin remodelling, non-coding RNAs, RNA silencing, and RNA modifications. Until recently, RNA modifications, or epitranscriptomics, lagged behind other fields due to the low abundance of these RNA modifications and the lack of techniques to readily detect the modifications transcriptome-wide at single nucleotide resolution. However, this has changed due to the rapid development of high-throughput RNA sequencing, like Oxford Nanopore, and the development of antibodies specific to an RNA modification (Meng et al. 2019; Saletore et al. 2012; Xiong, Yi & Peng 2017). RNA modifications have been demonstrated to participate in post-transcriptional regulation. For example, RNA modifications can directly alter RNA structure by promoting or disrupting intramolecular interactions, changing the flexibility of an RNA molecule and regulating RNA-protein interactions (Decatur & Fournier 2002; Lewis, CJT, Pan & Kalsotra 2017; Motorin, Y. & Helm 2010; Tuorto, Francesca et al. 2012). Although the regulatory pathways and biological functions of RNA modifications are yet to be fully elucidated, early data from animals and plants imply a complex, precise and vital role in growth and development.

The methyl group, $-CH_3$, is a covalent modification that frequently occurs on DNA, RNA and proteins, including histones. Methylation on histones is a highly conserved modification, and all four types of histones (H2A, H2B, H3, H4) can be methylated at specific amino acid residues. For example, the histone 3 lysine 4 (H3K4) can be mono- (H3K4me1), di- (H3K4me2) or trimethylated- (H3K4me3), leading to different cellular functions in plants and animals (Adhikary, Bakos &

Biggar 2019; Martin & Zhang 2005). On DNA, all four bases can be methylated in response to environmental perturbations or cellular programs. Methylation on cytosine, specifically at the fifth carbon, 5-methylcytosine (5mC), is linked to genomic imprinting, chromosome inactivation, transposon silencing and gene transcription in animals and plants (He, XJ, Chen & Zhu 2011; Jaenisch & Bird 2003; Liu, R & Lang 2020; Zhang, Huiming, Lang & Zhu 2018). Like DNA, all four RNA nucleotide bases can be methylated. Examples include; *N*6-methyladenosine (*N*6-mA), 3-methylcytosine (3mC), 7-methylcytosine (7mG), and 2'-O-methyluridine (Um). Cellular RNA 5mC was first identified in Henrietta Lacks (HeLa) and hamster cells in the 1970s (Dubin & Taylor 1975; Salditt-Georgieff et al. 1976).

With the development of high-throughput sequencing technologies, only recently has the full repertoire of 5mC RNAs been elucidated. These RNA classes include not only transfer RNAs (tRNAs) but also ribosomal RNAs (rRNAs), small nucleolar RNAs (snoRNAs), messenger RNAs (mRNAs), and long non-coding RNA (lncRNAs, Squires, Jeffrey E et al. 2012). More than 10,000 5mC sites have been described in human RNAs, and over 1,000 5mC sites on plant RNAs (Cui et al. 2017; David et al. 2017; Squires, Jeffrey E et al. 2012). RNA 5mC is an essential regulator of RNA export, ribosome assembly, translation efficiency and RNA structure stability (Gigova et al. 2014; Trixl, Lukas & Lusser 2018; Warren et al. 2010; Yang, X et al. 2017).

In this chapter, I provide an overview of common chemical modifications on RNA, focusing on methylation and an in-depth analysis of 5mC on different cellular RNAs.

1.2. RNA modifications and distributions

Over 200 modified RNA bases have been described in different organisms (Boccaletto et al. 2021). Examples of abundant or well-characterized RNA modifications are shown in Figure 1.1. This section discusses the distribution of common modifications on cellular RNAs and their potential functions.

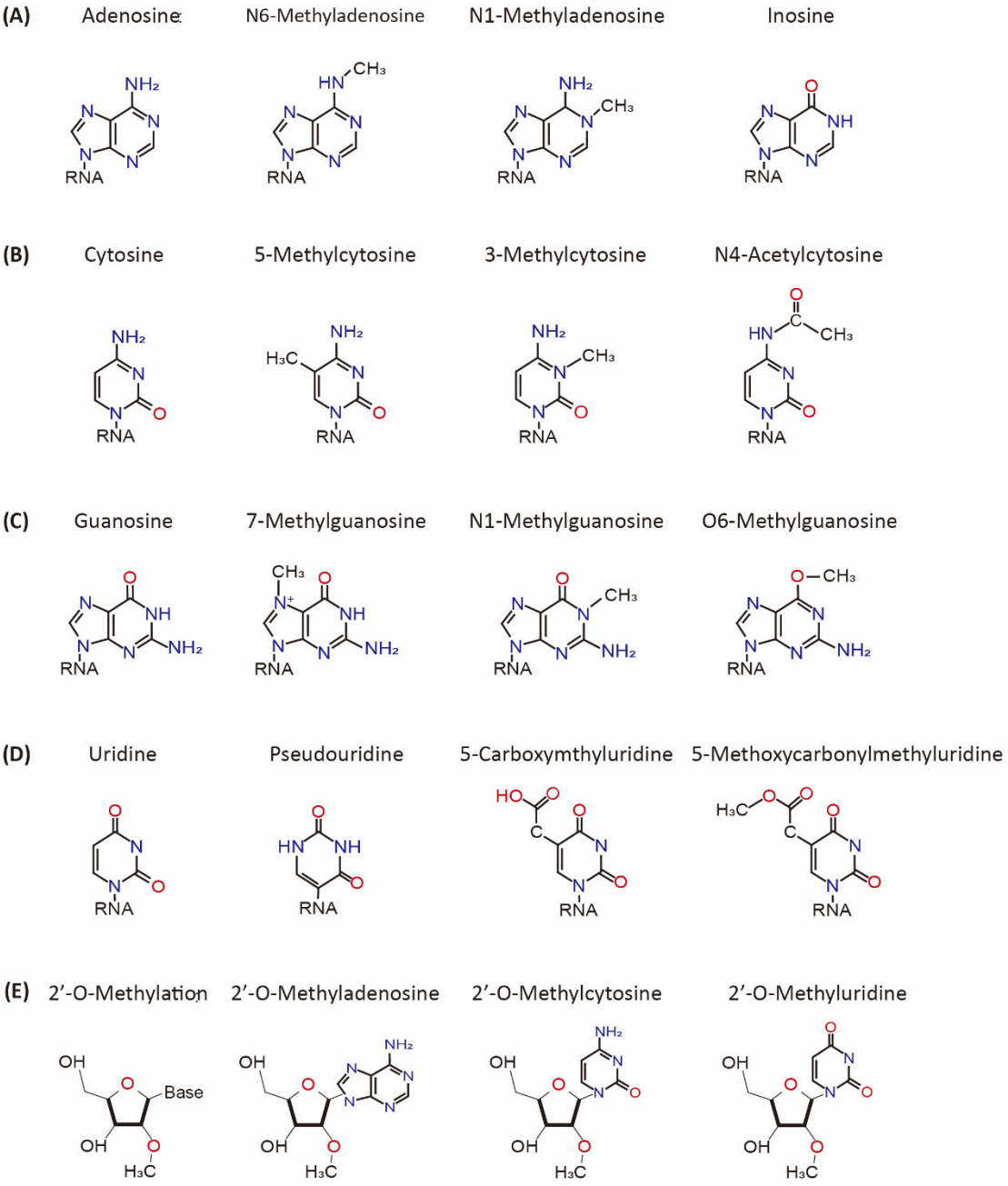


Figure 1.1: Commonly modified ribonucleosides present on cellular RNAs

(A) Examples of common chemical modifications on the four ribonucleosides: are **(A)** adenosine, **(B)** cytosine, **(C)** guanosine, and **(D)** uridine. 2'-O-methylation modification on **(E)** adenosine, cytosine and uridine bases.

1.2.1. mRNA ribonucleotides modifications

mRNAs contain numerous chemical modifications. *N*6-mA is one of the most common and abundant mRNA modifications, mainly in the coding region, 3'UTR and stop codon in humans and plants (Li, X., Xiong & Yi 2016; Niu et al. 2013). Its core role appears to affect mRNA metabolism, including mRNA translation, splicing, nuclear export, translation ability and mRNA structure stability (Camper et al. 1984; Carroll, Narayan & Rottman 1990; Dominissini et al. 2012; Niu et al. 2013; Tuck, Wiehl & Pan 1999). *N*6-mA was the first described reversible RNA modification with methyltransferase-like (METTL) proteins METTL3, METTL14 and METTL16 being active methyltransferases and the α-ketoglutarate-dependent dioxygenase alkB homolog protein (ALKBH5) and the fat mass and obesity-associated protein (FTO) being demethylases (Cao et al. 2016; Fu, Y et al. 2013). In contrast to *N*6-mA, *N1*-methyladenosine (*N*1-mA) was recently found near the 5' untranslated region (5'UTR) and the start codon region with potential roles in promoting translation and stress responses (Dominissini et al. 2016; Li, Xiaoyu et al. 2016). 5mC is a widely distributed modification that regulates gene expression and mRNA export. Its further details will be discussed in the following section. In the 5' UTR, 2'-O-methylated ribonucleotides (2'OMe) play roles in initiating transcription (Tschudi & Ullu 2002), and the 7mG cap improves the efficiency of gene expression, the stability of transcripts and translation (Barik 1993; Cowling 2010). There are also other modifications present on mRNAs. For example, pseudouridine (Ψ) exists in the coding region

and 3'UTR of mRNAs with some enigmatic functions (Li, X., Xiong & Yi 2016). Also, the generation of poly (A) tail at the 3' UTR is a crucial characteristic of mature mRNAs and has significant functions in transcript stability and nuclear export (Dubin & Taylor 1975; Elkon, Ugalde & Agami 2013). Detailed descriptions of the distribution and function of ribose base modifications were extensively reviewed by Cowling (2010); Gilbert, Bell and Schaening (2016); Nachtergaele and He (2018) and more recently by Chmielowska-Bąk, Arasimowicz-Jelonek and Deckert (2019).

1.2.2. tRNA ribonucleotide modifications

tRNAs are probably the most heavily modified cellular RNAs, with more than 100 covalent ribose modifications described at different positions among different organisms (Boccaletto et al. 2021; Jackman & Alfonzo 2013; Rościsław, Karol & Sebastian 2019). Some examples include adenosine to inosine (A-I), methylation (*N*6-mA, 5mC, Cm, 7mG), acetylation (4acC), and isomerisation of uridine to pseudouridine (Ψ, Björk & Hagervall 2014; Helm 2006; Koh & Sarin 2018). Interestingly, the prevalence of modifications can vary between different tRNAs, the metabolic state of the cell, and the organism, making it difficult to demonstrate the function of a single modification on a tRNA (Björk & Hagervall 2014; Jackman & Alfonzo 2013). The main functions of tRNA modifications are thought to affect the overall structure and translation (Alexandrov et al. 2006; Koh & Sarin 2018; Motorin, Y. & Helm 2010).

Chemical modifications can regulate the stability and flexibility of tRNA structure. Regarding tRNA structures, the 'cloverleaf' and 'L-shape' structures are commonly known, although gross changes of the 'cloverleaf' structure, such as the loss of the entire D-stem and loop, are observed in some mitochondrial tRNAs (Jühling et al. 2018). Maintenance of the L-shape tRNA structure requires

RNA modifications that affect the hydrogen bonding potential of a nucleotide base. For example, $1mA_9$ of human mitochondrial $tRNA^{Lys}$ shifts the equilibrium toward the canonical cloverleaf by disrupting the formation of a Watson–Crick base pair in another non-canonical conformation (Voigts-Hoffmann et al. 2007). Modifications in the core of the folded tRNA can adjust the rigidity and flexibility of the tRNA structure. One example of increased rigidity is the pseudouridine residue, which participates in the 3'-endo sugar associated pucker with A-form RNA helices (Durant et al. 2005). While on the contrary, dihydrouridine increases tRNA conformational flexibility by destabilising the 3'-endo sugar conformation and promoting the 2'-endo sugar conformation (Dalluge, Joseph J, Hashizume & McCloskey 1996; Dalluge, Joseph J. et al. 1996).

tRNA modifications also play critical roles in translational fidelity and efficiency. Nucleotides at position 37 and wobble position 34 are frequently modified in almost all tRNA, regardless of the organism (Jackman & Alfonzo 2013). Modifications at position 37 function in maintaining and opening loop conformation, disrupting Watson–Crick pairing and enhancing the anticodon–codon pairing during decoding (Jackman & Alfonzo 2013). For example, in most tRNAs, G_{37} is methylated to form $1mG_{37}$. The lack of $1mG_{37}$ leads to incorrect +1 frameshifting, which reduces the accuracy of translation and, as a result, severely damages the growth of bacteria (Björk et al. 2001). Other modifications at position 34 include pseudouridylation, ribose methylation and acetylation, and their functions are to improve the efficiency of codon-anticodon pairing and combined with modifications at position 37 to enhance the stability of the tRNA–mRNA interaction during decoding (Agris, P. F., Vendeix & Graham 2007). For example, U_{34} in $tRNA^{Gln}$, $tRNA^{Glu}$, and $tRNA^{Lys}$ is modified to 2-thiouridine (s2U), which can increase anticodon rigidity, translation efficiency and fidelity (Agris, P. F., Vendeix & Graham 2007; Madore et al. 1999; Rodriguez-Hernandez

et al. 2013; TISNÉ et al. 2000).

Of course, many more tRNA modifications cannot be enumerated here. The presence and role of tRNA modifications have been extensively reviewed by Motorin, Y. and Helm (2010), Jackman and Alfonzo (2013) and more recently by Rościsław, Karol and Sebastian (2019).

1.2.3. rRNA ribonucleotide modifications

In eukaryotes, the major rRNA modifications are U to Ψ conversion, 2'O-Me and ribonucleotide methylations. 2'O-Me and Ψ are the most abundant modifications in rRNAs (Birkedal et al. 2015). The main roles of these two modifications are to maintain the accuracy and efficiency of protein synthesis. For example, in yeast, the loss of 2'O-Me and Ψ modifications at the A- and P-sites of the decoding centre led to a significant reduction of amino acid incorporation rates and a strong delay of pre-rRNA processing (Liang, Liu & Fournier 2009). Also, rRNA ribose modifications generally stabilise rRNA secondary and tertiary structures (Decatur & Fournier 2002; Helm 2006; Sloan et al. 2017). Specifically, 2'O-Me stabilises the helix structure by increasing base accumulation. Correspondingly, Ψ, which has a greater hydrogen bonding potential than uridine, also enhances the rigidity of the glycolate-phosphate backbone. Also, ribonucleotide methylations maintain rRNA structure stabilisation through other aspects. For example, 7mG promotes the ionic interaction between RNAs and proteins through the increased positive charge of its quaternary nitrogen (Agris, Paul F., Sierzputowska-Gracz & Smith 1986). Further information about the distribution and function of rRNA modifications was reviewed by Decatur and Fournier (2002), Baxter-Roshek, Petrov and Dinman (2007) and more recently by Sloan et al. (2017).

1.2.4. ncRNA ribonucleotide modifications

Chemical modifications are also found on other non-coding RNAs (ncRNAs), such as small nuclear RNA (snRNA), microRNA (miRNAs) and lncRNA. For example, 2'O-Me methylations on G_{25} of U2 snRNA 5'Cap and G_{22} of U12 snRNA 5'Cap are specific recognition sites for the binding of the spliceosome factor 3b (SF3b) complex in metazoans (Tycowski, Aab & Steitz 2004). Also, in plants, 2'O-Me at the 3'-end of miRNA stabilise miRNAs and prevent 3'-uridylation and degradation (Backes et al. 2012; Borges & Martienssen 2015). Moreover, 5mC sites in the 5' A-region of lncRNA disrupt the binding of X-inactive specific transcript (XIST, Amort et al. 2013), while *N6*-mA is involved in XIST-mediated transcriptional repression (Patil et al. 2016). The presence and role of other ncRNA modifications have been extensively reviewed by Lewis, CJ, Pan, T and Kalsotra, A (2017) and Rong et al. (2021).

1.3. The distribution and function of 5mC on cellular RNAs

1.3.1. The presence and function of 5mC on mRNA

The abundance and distribution of 5mC on mRNAs vary among species and tissues. In HeLa cells and mice, 5mC sites are highly concentrated around the start codon in 5'UTR and coding sequencing (CDS), followed by the 3'UTR, CDS and 5'UTR (Figure 1.2A, Amort et al. 2017; Yang, X et al. 2017). While in plants, 5mC sites mainly accumulate in the CDS, especially near the stop codon, followed by the 3'UTR. A low abundance of 5mC sites is distributed in the 5'UTR (Figure 1.2A, Cui et al. 2017; David et al. 2017). Besides, the distribution of mRNA 5mC is sequence specific. In human, 5mC predominantly exist in CG

regions, followed by CHG and CHH regions (where H = A, C, U, Yang, X et al. 2017). Some 5mC sites are tissue specific. For example, in Archaea, 5mC is associated with the conserved motif AU**C**GANGU, while this motif has not been associated with 5mC sites in animals and plants (David et al. 2017; Edelheit et al. 2013; Hussain et al. 2013; Squires, Jeffrey E et al. 2012).

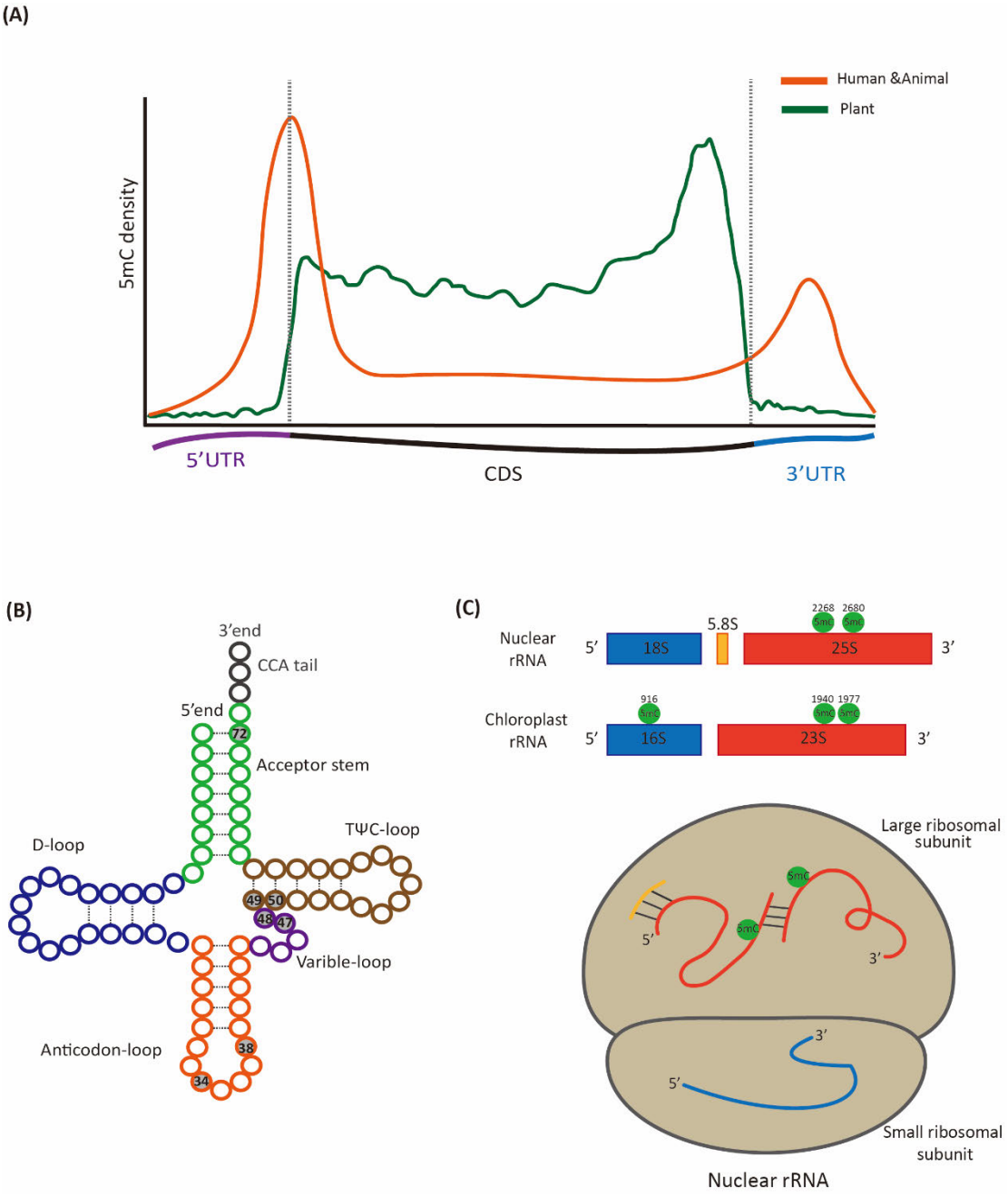


Figure 1.2: The distribution of 5mC on tRNA, mRNA and ncRNAs in animals and plants.

(A) The distribution of 5mCs on mRNAs in animals and plants. In HeLa cells and mice, 5mC sites are concentrated around the start codon, 5'UTR and 3' UTR, and have reduced abundance in the coding regions (CDS). In *A. thaliana*, the distribution of 5mC to animals as cites is more concentrated in the coding sequence and low abundance in the UTRs. **(B)** The distribution of 5mCs on tRNAs. 5mC is highly conserved throughout evolution and is present in the anticodon loop (C_{34}), the anticodon stem (C_{38}), the junction of the variable loop (C_{47}-C_{50}) and the acceptor stem (C_{72}). **(C)** The distribution of 5mCs in nuclear and chloroplast rRNAs in the Plant. In *A. thaliana*, 5mC exists at positions C_{2268} and C_{2860} of the nuclear 25S but not in 18S RNA; C_{1940} and C_{1977} in chloroplast 23S rRNA and C_{916} in 16S chloroplast rRNAs.

To date, the functions of most 5mCs on mRNAs have not been unrevealed in animals and plants. A clear correlation between the occurrence of 5mC and the overall transcript abundance is controversial. In some cases, 5mC regulates the binding of RNA binding proteins; examples include mRNA export from the nucleus, splicing and decay (Amort et al. 2017; Yang, X et al. 2017). 5mC regulates mRNA export from the nucleus through binding an RNA 5mC 'reader', the Aly/REF export factor protein (ALYREF, Yang, X et al. 2017). A detailed description of how 5mC regulates mRNA export is described in section 3.1.1. A second reader protein is Y-Box Binding Protein (YBX1) that stabilises the *Heparin Binding Growth Factor gene (HDGF)* mRNA by recruiting the ELAV Like RNA Binding Protein 1 (ELAVL1) to targeting 5mC sites (Chen, X et al. 2019). Another YBX protein, YBX2, can specifically bind to 5mC to undergo liquid-liquid phase separation (LLPS) in mammals (Wang, X et al. 2021). In addition, Warren et al. (2010) found that synthetic 5mC methylated mRNAs are more stable and have a higher half-life than those unmethylated. Perhaps less clear is 5mC on mRNA 3' UTRs associated with Argonaute proteins 1-4 (AGO1–4) binding in HeLa cells, which is the central component of the miRNA/RISC complex (Squires, Jeffrey E et al. 2012). This implies that RNA 5mC functions in the miRNA-guided decay of mRNA and inhibition of translation. In saying this, other groups have not

reported the same observations (personal communication). In plants, the long-distance transport of some mRNAs is vital for coordinated growth and development; however, how this process is achieved remains controversial (Thieme et al. 2015; Xu, H et al. 2013). Recently, a role of 5mC on long-distance mRNA transport was demonstrated in *A. thaliana,* which is outlined in detail in section 1.5.2 (Yang, L et al. 2019).

1.3.1.1. 5mC guides mRNA transport from the nucleus to the cytoplasm

In eukaryotes, transporting mature mRNA from the nucleus to the cytoplasm is an elaborate process requiring high-level coordination and cooperation of mRNA and export receptors (RNA-binding proteins). However, mature mRNAs are different in length, sequence, and structure, making them hard to be directly identified by export receptors. There should be some consistent characteristics of mature mRNAs responsible for binding export receptors, and 5mC modification is one of them.

In 2017, RNA 5mC was reported to promote the ALY/REF-mediated mRNA export from the nucleus to the cytoplasm in mice (Yang, X et al. 2017). ALY/REF is a component of the highly conserved co-transcriptional THO/TREX complex (TREX) that functions as an export adaptor during transcription, pre-mRNA processing and mRNA export (Katahira 2012; Pfaff et al. 2018). ALYREF acts as a 5mC 'reader' in the nucleus by binding to NOP2/Sun domain family member (NSUN2)-dependent 5mC sites distributed at the beginning and the end of mRNA exons (Yang, X et al. 2017). When exporting mRNAs, ALY/REF-mRNAs are bound to the export receptor (TAP–p15 complex) to form export-competent messenger ribonucleoproteins (mRNPs, Kohler & Hurt 2007; Segref et al. 1997;

Taniguchi & Ohno 2008). After export through the nuclear pore complex (NPC) to the cytoplasm, the ALY/REF-TAP–p15 complex is released from the mRNAs and then re-imported to the nucleus through an NPC (Koffa et al. 2001; Kohler & Hurt 2007). During nuclear export, the interaction and release of the ALY/REF complex with mRNAs may be dynamically controlled by adding and removing 5-methyl groups at specific cytosines. Yang, X et al. (2017) showed that NSUN2-dependent 5mC could regulate nuclear-cytoplasmic shuttling by regulating the RNA binding affinity of ALYREF to mRNAs. Cells deficient in NSUN2 lead to reduced ALYREF-mRNA binding, enhanced retention of ALYREF and mRNAs in the nucleus and subsequently decreased levels of ALYREF and mRNAs in the cytoplasm. These findings highlight the importance of RNA 5mC in mRNA export.

1.3.2. The presence and function of 5mC on tRNAs

In tRNAs, 5mC sites are most commonly distributed at C_{34} of the anticodon loop and C_{38} of the anticodon stem, the junction of variable loop and C_{47}-C_{50} of the TΨC loop, and C_{70} of the acceptor stem (Figure 1.2B, Björk & Hagervall 2014; Burgess, David & Searle 2015), and their main functions are to stabilise tRNA secondary structure, regulate the stability and specificity of tRNA in translation and respond to cellular stress. $5mC_{38}$ at the anticodon stem-loop, for example, on $tRNA^{Asp}_{GTC}$, $tRNA^{Gly}_{GCC}$ and $tRNA^{Val}_{AAC}$, play roles in regulating Mg^{2+} binding, promoting interaction between codon and anticodon, promoting structure stability and protecting tRNAs from stress-induced fragmentation (Chen, Y et al. 1993; Motorin, YURI & Grosjean 1999; Motorin, Y. & Helm 2010; Schaefer et al. 2010; Tuorto, Francesca et al. 2012). The 5mC on C_{34} and C_{48} have a role in up-regulating the expression of oxidative-induced protein and enhancing cell survival under oxidative stress. In yeast, $tRNA^{Leu}_{CAA}$ 5mC is uniquely and dynamically present at the wobble position (C_{34}) of the anticodon loop and C_{48}

of the variable loop. Upon cellular oxidative stress, C_{34} and C_{48} 5mCs are significantly increased, leading to a significant translational bias for the oxidative response protein Ribosomal Protein L22 (RPL22A) has TTG-enriched codons in the mRNA (Chan et al. 2012; Hou, Gamper & Yang 2015). The function of 5mC at C_{47}-C_{50} is to stabilise the structure of the tRNA three-dimensional core and improve base stacking (Nobles et al. 2002; Sowers, Shaw & Sedwick 1987). In plants, C_{48}, C_{49} and C_{50} are the most frequently methylated sites (Burgess, David & Searle 2015). Nucleoside 48 in the V loop can unconventionally pair with nucleoside 15 in the D loop to generate an L-shaped three-dimensional structure. For those tRNAs with G_5 and C_{48}, over 26% of the C_{48} are 5-methylated (Oliva, Tramontano & Cavallo 2007). Although the physicochemical role of $5mC_{48}$ has not been fully elucidated, $5mC_{48}$ increases the hydrophobicity of the nucleoside and promotes base stacking to maintain the geometry structure and stabilise tRNA tertiary interactions (Oliva, Cavallo & Tramontano 2006; Oliva, Tramontano & Cavallo 2007; Vare et al. 2017). In contrast to other organisms, higher eukaryotes such as humans often have an additional tRNA 5mC site at position C_{72} (Burgess, David & Searle 2015; Khoddami & Cairns 2013; Motorin, Yuri, Lyko & Helm 2009). In plants, 5mC at C_{72} was only found in $tRNA^{Asp}_{GTC}$. The particular function of the $5mC_{72}$ remains unclear.

tRNA 5mC was demonstrated to regulate cell division and differentiation and stress responses in different organisms. The lack of tRNA 5mC sites in mice results in smaller body size and reduced division in many tissues, and some die soon after birth (Tuorto, Francesca et al. 2012). Similar observations were also made in zebrafish (Goll et al. 2006; Rai et al. 2007). Interestingly, mutations in the 5mC methyltransferase NSUN2 causes inherited intellectual disability, growth obstacle and skin disorders in humans (Blanco et al. 2014). These phenotypes may be due to increased cleavage of tRNAs, and as a result,

reduced translation rates and stress pathways are activated. In Drosophila, the absence of the tRNA 5mC results in increased sensitivity to oxidative stress, while overexpression of the 5mC methyltransferase improves oxidative stress tolerance (Schaefer et al. 2010). *Arabidopsis thaliana* (*A. thaliana*) *trm4b* mutants have less 5mC level at positions C_{39} and C_{40} of tRNAs, reduced steady-state levels of $tRNA^{Asp}_{GTC}$ and reduced cell division in the root meristem that leads to shorter primary roots (Burgess, David & Searle 2015; David et al. 2017).

1.3.3. The presence and function of 5mC on rRNAs and ncRNAs

Due to the high abundance of rRNAs in cells, many aspects of rRNA 5mC methylation and molecular function are well understood in many organisms. The ribosome comprises two subunits, the large subunit (LSU) and the small subunit (SSU). The LSU usually has two or three rRNA (e.g., 25-28S and 5.8S rRNA in eukaryote, 23s in Prokaryote and chloroplast), and the SSU contains one rRNA (e.g., 18S rRNA in eukaryote, 16s in Prokaryote and chloroplast, Figure 1.2C, Alberts et al. 2003). Because of the inconsistency of rRNA types, the distribution of rRNA 5mCs varies among organisms. *Escherichia coli* (*E.coli*) contains one conserved single 5mC site at C_{1962} in 23S rRNA and two 5mC sites at C_{965} and C_{1407} in 16S rRNA (Kowalak et al. 1993; Smith, Cooperman & Mitchell 1992). In yeast, rRNA 5mCs exist at positions C_{2870} and C_{2278} of 25S rRNA (Gigova et al. 2014; Sharma et al. 2013). In human, conserved 5mC sites were identified at the C_{3782} and C_{4417} of 28S rRNA (Schaefer et al. 2008; Squires, Jeffrey E. et al. 2012) and C_{841} of 12S mitochondrial rRNA (Squires, Jeffrey E. et al. 2012). In *A. thaliana*, 5mCs are present at C_{2268} and C_{2860} in nuclear 25S rRNA; C_{1940} and C_{1977} in chloroplast 23S rRNA and C_{916} in 16S chloroplast rRNAs; and C_{1586} of in mitochondrial 26S rRNA and C_{960} in mitochondria 18S rRNA (Burgess, David &

Searle 2015). The current known functions of rRNA 5mCs are related to translation efficiency, rRNA processing and rRNA structure conformation. For example, the deficiency of 5mC at C_{2278} changes the structural conformation of ribosome near C_{2278} and the fidelity of translation and recruit a distinct subset of mRNAs oxidative stress-responsive into polysomes in yeast cell (Schosserer et al. 2015), indicating the significant roles of 5mC in rRNA-mediated translational regulation and stress response.

5mCs also play essential roles in ncRNAs. The lack of 5mC in vault ncRNA (vtRNA) results in the generation of a miRNA and subsequently incorporates into AGO complexes to regulate target gene expression (Hussain et al. 2013). A second example is that five 5mC sites in the 5' A-region of lncRNA X-inactive specific transcript (XIST) prevented the binding of the Polycomb repressive complex 2 (PRC2) *in vitro* and as a result, stopped the silence of X chromosome (Amort et al. 2013).

1.4. Enzymatic regulation of RNA 5mC

1.4.1. RNA 5mC methyltransferases

The formation and maintenance of RNA 5mC are controlled by a series of so-called 'writer' enzymes or methyltransferases, which specifically transfer a methyl group from AdoMet to the fifth carbon of cytosine (Figure 1.3). To date, two classes of RNA 5mC methyltransferases have been identified. The largest class are a superfamily of Rossmann fold motif-containing enzymes (RFM), including nine RNA (cytosine-5)-methyltransferase (RCMT) subfamilies (RCMT1-RMCT9, Foster et al. 2003; Pavlopoulou, Athanasia & Kossida, Sophia 2009). All these subfamilies are highly conversed at the primary amino sequence level and have functions from single-celled to multicellular organisms (Pavlopoulou,

Athanasia & Kossida, Sophia 2009; Reid, Greene & Santi 1999). RCMT1 and RCMT3-6 are only present in prokaryotes. Four other subfamilies, RCMT2, RCMT7, RCMT8 and RCMT9, are strictly responsible for catalysing different RNA cytosine sites in eukaryotes.

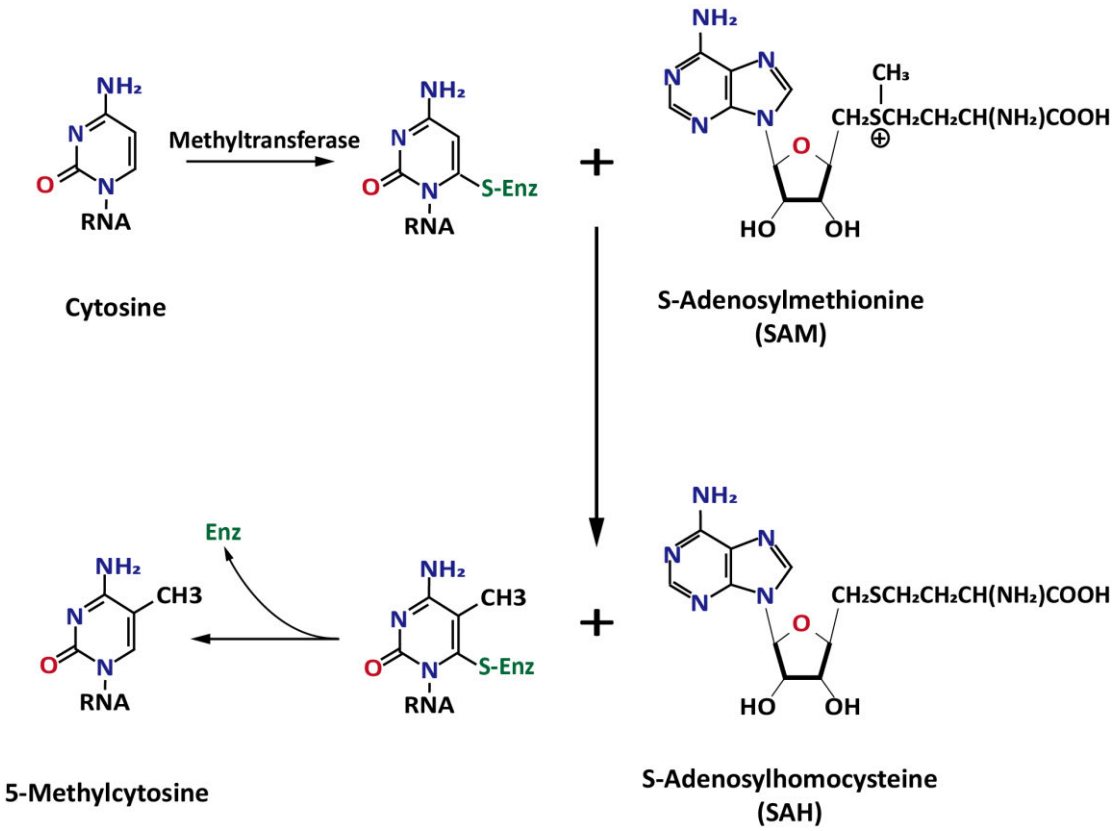


Figure 1.3: Enzymatic methylation of cytosine to 5-methylcytosine.

In the process of cytosine methylation, first, the methyltransferase covalently binds to the C_6 site of the cytosine ring; the C_5 site is activated to allow the transfer of a methyl group from S-Adenosylmethionine (SAM) to the C_5; after methylation, methyltransferase in the cytosine base is released, methylated cytosine is generated, and SAM is conversed to SAH. S-Enz: the combination of a sulphur atom (the conserved cysteine residue) and methyltransferase enzyme; Enz: methyltransferase.

1.4.1.1. RCMT2 family protein

NOP2/NSUN1 is the only characterised RCMT2 protein that is localized to the nucleolus (Hong et al. 1997; Pavlopoulou, A. & Kossida, S. 2009). In yeast, NOP2 catalyses the methylation of C_{2870} in 25S rRNA (Gigova et al. 2014; Hong et al.

1997; Sharma et al. 2013). Three paralogs of NOP2, namely OLI2 (also known as NOP2A), NOP2B and NOP2C, have been identified in *A. thaliana* (Burgess, David & Searle 2015; Maekawa & Yanagisawa 2021; Pavlopoulou, Athanasia & Kossida, Sophia 2009). Interestingly, single mutants do not reduce methylation at C_{2860} (the orthologous positions of C_{2870}) of rRNA (Burgess, David & Searle 2015), but recently RNA interference (RNAi) NOP2 knockdown lines were shown to have reduced CXXX methylation (Wu, Haijun et al. 2020). No NOP2/NSUN1-dependent 5mC site has been found in humans. Further study is required to demonstrate its RCMT capacity on RNA 5mCs in animals.

1.4.1.2. RCMT7 family proteins

RCMT7 family proteins, including NSUN2 (also known as TRM4), NSUN3, and NSUN4, are the main methyltransferases of 5mC on tRNAs, mRNAs and mitochondrial RNAs (Bujnicki et al. 2004; Motorin, YURI & Grosjean 1999; Motorin, Y. & Helm 2010; Schaefer 2015; Sibbritt, Patel & Preiss 2013).

To date, NSUN2 is the most studied RNA 5mC methyltransferase in either animals or plants. NSUN2 was firstly demonstrated as a tRNA methyltransferase, which was mainly located near the nucleolus where tRNA processing occurs (Frye & Watt 2006; Sakita-Suto et al. 2007) and methylates C_{34} in the anticodon loop and the C_{47}-C_{50} of the TΨC stem junction of the flexible loop (Figure 1.2B). Some molecular analyses showed that *nsun2 mutant* mice lost methylation in these positions of $tRNA^{Gly}_{GCC}$, $tRNA^{Leu}_{TAG}$, $tRNA^{Asp}_{GTC}$, and $tRNA^{Val}_{AAC}$ but no difference in the abundance of other tRNA 5mC sites (Tuorto, Francesca et al. 2012). Similar results were also published in zebra fish (Goll et al. 2006; Rai et al. 2007). In *A. thaliana*, one NSUN2 ortholog, TRM4B, has 5mC methylation activity towards C_{48}-C_{50} of $tRNA^{Gly}_{GCC}$, $tRNA^{Asp}_{GTC}$, $tRNA^{Asp}_{GTC}$, $tRNA^{Glu}_{TTC}$ (Burgess, David & Searle 2015). These data support the significance of NSUN2 in cytosine

methylation in tRNAs. Subsequent studies also confirmed the methylation activity of NSUN2 on mRNAs and other ncRNAs in different organisms (Burgess, David & Searle 2015; Hussain et al. 2013; Yang, X et al. 2017). In HeLa cells, NSUN2 is responsible for the 5mC methylation on over 2000 cytosine sites on mRNAs and ncRNAs (Yang, X et al. 2017). In *O. Sativa*, the deficiency of the NSUN2 lead to a significant reduction of 5mC level on mRNAs (Tang, Yongyan et al. 2020). Furthermore, NSUN2 can catalyse 5mC methylation on 6 cytosine sites in the vtRNA, thereby affecting the processing of a precursor vtRNA into small viral RNAs (svRNAs) in Human embryonic kidney 293 cells .

Other RCMT7 family proteins regulate RNA 5mC levels. NSUN3 is required for methylation of mitochondrial (mt)-$tRNA^{Met}$ in human cells and mouse embryonic stem cells (Haag, Sara et al. 2016; Nakano et al. 2016; Van Haute et al. 2016). Unlike other RCMT7 members, NSUN4 is an rRNA-specific methyltransferase targeting C_{911} on the mitochondrial small subunit 12S rRNA in mice (Metodiev et al. 2014). The importance of NSUN4 in regulating mitochondrial ribosome assembly has been highlighted by mitochondrial dysfunction in *nsun4* deficiency mice (Metodiev et al. 2014).

1.4.1.3. RCMT8 family protein

RCMT8 subfamily NSUN5 (also known as RCM1P in yeast) is the second 5mC methyltransferase known to target the 25S rRNA at C_{2278} in yeast. Similar to NOP2/NSUN1, NSUN5 is mainly localized to the nucleolus (Gigova et al. 2014; Sharma et al. 2013). The deficiency of NSUN5/RCM1P in yeast impacts the binding of several ribosomal proteins and results in a decreased stability of the 60S ribosomal subunit (Gigova et al. 2014). In *A. thaliana*, *nsun5* mutant plants have strongly reduced methylation at C_{2268} (the orthologous position of C_{2278}) of the nuclear LSU 25S rRNA (Burgess, David & Searle 2015), indicating that NSUN5

had the role in plants.

1.4.1.4. RCMT9 family protein

Phylogenetic analysis demonstrated a new RCMT subfamily, RCMT9, in higher plants, such as *A. thaliana*, *O. sativa*, *Zea mays* (*Z. mays*, maize) and *Sorghum bicolor* (*S.bicolor*). RCMT9 seems to be a paralogue of RCMT7 and contains a SAM binding domain and a Rossmann fold structure. However, the cellular substrates of RCMT9 remain unknown.

1.4.1.5. TRDMT1 protein

Another class of RCMTase is the tRNA aspartic acid methyltransferase (TRDMT1), the first known RNA 5mC methyltransferase (Goll et al. 2006). TRDMT1 was previously named DNA methyltransferase2 (DNMT2) because of its similar primary amino acid sequence to DNA methyltransferases. However, several studies demonstrated that TRDMT1/DNMT2 methylates site C_{38} of $tRNA^{Asp}_{GTC}$, $tRNA^{Gly}_{GCC}$ and $tRNA^{Val}_{AAC}$ in animals and C_{38} of $tRNA^{Asp}_{GTC}$ in *A. thaliana* (Burgess, David & Searle 2015; Goll et al. 2006; Schaefer et al. 2010; Tuorto, Francesca et al. 2012). All these tRNAs share the same sequence motif around C_{38} (5' CA-5mC-GCG 3'), indicating that TRDMT1 has a sequence and structural recognition mechanism (Goll et al. 2006). *Trdmt1*-knockout mice, flies and plants do not have an obvious phenotype under controlled conditions (Burgess, David & Searle 2015; Goll et al. 2006). However, further analysis revealed that new-born *trdmt1*-mutated mice had delayed endochondral ossification, decreased haematopoiesis stem and progenitor cell population and a cell-autonomous differentiation defect due to the loss of $5mC_{38}$ in $tRNA^{Asp}$ and $tRNA^{Glu}$ (Tuorto, F. et al. 2015). Similarly, tissue differentiation defects were also found in *dnmt2* zebrafish (Rai et al. 2007).

1.4.2. RNA 5mC demethylases

In contrast to methylation, the demethylation pathway is the process of removing methyl groups from substrates. It is also an important process in regulating methylation levels. For a long time, RNA 5mC has been considered a unidirectional and constant modification. The discovery of RNA 5-hydroxymethylcytosine (5hmC) as an oxidative derivative of 5mC and Ten-eleven translocation 1 (TET1) as the oxidase demonstrated that RNA 5mC could be further oxidized and reversed to unmodified cytosine in mammals (Fu, L et al. 2014). However, no TET-like enzyme or RNA 5mC demethylase has been found in plants. If and how the methyl group is removed from RNA 5mC in plants is unclear. Below, TET family proteins are shown as an example to introduce the demethylation pathway and mechanism of RNA 5mC.

1.4.2.1. TET family proteins are cytosine demethylases in mammals

TET family proteins, including TET1, TET2 and TET3, belong to the TET/J-binding protein (JBP) family of AlkB-like Fe (II)/α-ketoglutarate (αKG)-dependent dioxygenases (Iyer et al. 2009; Tan & Shi 2012). TETs were first identified as demethylases of DNA 5mC, which can dynamically reverse DNA 5mC to unmodified cytosine via the oxidative demethylation pathway (Kriaucionis & Heintz 2009; Tahiliani et al. 2009). Briefly, TETs repeatedly oxidize DNA 5mC to generate 5hmC, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Both 5fC and 5caC can be removed by thymine DNA glycosylase (TDG), and then the unmodified cytosine is filled into the vacancy site by the DNA base excision repair (BER) pathway (He, Y-F et al. 2011; Ito et al. 2011). Less clear is the process of demethylating RNA 5mC by TET1.

Though TET1 has demethylation activity on DNA and RNA 5mCs, the efficiency

of TET1- demethylation in RNA is much lower than on DNA substrates. *In vitro* experiments by Fu, L et al. (2014) showed that 5hmC is the main oxidation product of RNA 5mC catalysed by TET1. When double-stranded DNA (dsDNA) was employed as the substrate, a rapid formation of 5hmC and then 5fC was observed, and subsequently, 5hmC and 5fC almost entirely conversed to 5caC. In mouse embryonic stem (ES) cells, *TET* triple mutants only decrease RNA 5hmC (Fu, L et al. 2014), implying that TET1 might not be the only RNA 5mC demethylase in mice. Therefore, other RNA 5mC demethylases are likely involved and present in mice.

1.4.2.2. Demethylation mechanism of TET enzymes

Although there is still a large of unclear how RNA 5mC is returned to unmodified C, the demethylation pathways and related enzymes of other methylated bases, such as DNA 5mC, have been well studied. The *Escherichia coli* AlkB family proteins of iron (II)/ αKG-dependent dioxygenases are significant regulators of methylation level that remove methyl groups from different bases via oxidative dealkylation pathway (Chen et al. 2014; Shen et al. 2014; Ye et al. 2014). AlkB is the *E. coli* archetype to reverse DNA methylation damage. Nine human homology proteins, ALKBH1-8 and FTO, together with TET proteins, are well known to demethylate different kinds of DNA and RNA-methylated bases. All AlkB-like proteins share a very similar catalysation pathway. Here, TETs are used as an example to introduce the detailed mechanism of the oxidative demethylation pathway.

TET enzymes have a highly conserved carboxyl-terminal core catalytic domain consisting of a unique CR domain and a double-stranded β-helix (DSBH) domain as other AlkB-like enzymes (Figure 1.4, Pavlopoulou, A. & Kossida, S. 2009; Tahiliani et al. 2009). The DSBH fold contains eight conserved antiparallel β-

strands (I–VIII, McDonough et al. 2010; Yi, Yang & He 2009). Also, a large, unique, non-conserved, low-complexity region is inserted between β-strands IV and V with some unknown functions (Figure 1.4, An, Rao & Ko 2017; Hu, L et al. 2015; Saletore et al. 2012; Shen et al. 2014). In the centre of the DSBH domain, a key catalytic motif: His–Xaa–Asp–$(Xaa)_n$–His (Xaa means any amino acid, HXD…H motif, Figure 1.4) cooperates with Fe (II) and αKG to active the oxidation reaction (Shen et al. 2014; Stephanie & Timothy 2013; Yi, Yang & He 2009). The unique and conserved CR domain is adjacent to the N terminus of the DSBH (Figure 1.4, An, Rao & Ko 2017; Tahiliani et al. 2009). It contains two subdomains, at least eight conserved cytosines, and one histidine (An, Rao & Ko 2017). However, the function of the CR domain is still largely unknown. One of its potential roles is to modulate the chromatin targeting of TET protein (An, Rao & Ko 2017; Yamagata & Kobayashi 2017). The CR domain of TET2 preferentially binds chromatin at the histone H3 tail by recognising H3 lysine 36 mono (H3K36me1) and demethylation (H3K36me2, Yamagata & Kobayashi 2017).

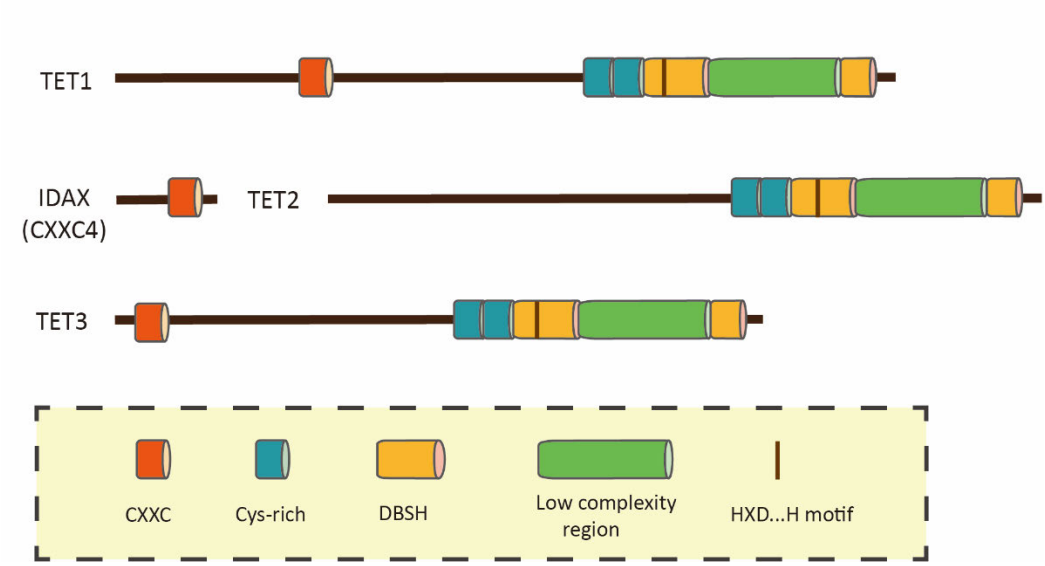


Figure 1.4: functional domains of TET family proteins

All TET enzymes have a highly conserved carboxyl-terminal core catalytic domain consisting of two cysteine-rich regions shown in red, a DSBH domain shown in yellow, an HXD...H motif and a low complexity region shown in green. Both TET1 and TET3 have a

DNA-binding CXXC domain at their N terminus. However, the CXXC domain of TET2 was encoded separately by the neighbouring gene, IDAX.

At the N terminus, TET1 and TET3 have a DNA-binding CXXC domain consisting of two Cys4-type zinc finger motifs (Figure 1.4, Iyer et al. 2009; Tahiliani et al. 2009). However, the CXXC domain of TET2 was separated from the catalytic domain due to a chromosomal inversion during evolution and is now encoded by a neighbouring gene called *inhibition of the Dvl and Axin complex* (IDAX, also known as CXXC4, Figure 1.4, Ko et al. 2013; Long, Blackledge & Klose 2013). CXXC domain and IDAX are highly conserved and preferentially bind to the unmethylated CpG region (Ko et al. 2013; Xu, Y et al. 2012; Zhang, H. et al. 2010). TET1 is preferentially enriched at the promoter CpG islands (CGIs) or enhancers in mouse embryonic stem cells (ESCs), while TET2 without the CXXC domain is primarily found in gene bodies or enhancer regions (Hon et al. 2014; Huang, Y et al. 2014; Wu, Hao et al. 2011).

TETs can specifically recognize and capture the targeted 5mC into their DSBH pocket using a base-flipping mechanism (Shen et al. 2014). The oxidation reaction can be divided into two steps: dioxygen activation and oxidation. Firstly, Fe (II) and αKG donate two electrons to react with dioxygen to generate a bridged preoxo and then form a highly charged active Fe (IV)-oxo compound. Subsequently, the charged Fe(IV)-oxo compound oxidizes the C-H bond on the CH_3 substrate to form -CH_2OH and returns itself to Fe (II) (Li, S & Mason 2014). Because the –CH_3 of 5mC is linked with the cytosine by a chemically stable C-C bond, the oxidized 5hmC can be further oxidized to 5fC and even 5caC (Shen et al. 2014). Briefly, TET repeatedly oxidizes 5mC to generate 5hmC, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). TDG can remove both 5fC and 5caC, and the unmodified cytosine is filled into the vacancy site by the BER pathway.

1.4.2.3. RNA 5hmC is present in diverse organisms

5-hydroxymethylcytosine (5hmC), a possible oxidative derivative of 5mC, was observed in diverse organisms, including *Methanocaldococcus jannaschii* (*M. jannaschii*), *E. coli, A. thaliana, Caenorhabditis elegans* (*C. elegans*), *HEK293T* cells and mouse brains (Huber et al. 2015). Surprisingly, RNA 5hmC has also been identified in species that do not contain TET homologues, such as *A. thaliana* and *C. elegan,* suggesting that RNA 5mC demethylation may be catalysed by other AlkB-like enzymes other than TETs in plants and worms.

The abundance of RNA 5hmC varies across different organisms. *A. thaliana* has the highest RNA 5hmC level, about 0.07% of cytosines, while in *C. elegans,* RNA 5hmC is less than 0.0005% (Fu, L et al. 2014). The variation of 5hmC level among different organisms suggests some species-specific functions. In Drosophila*,* over 3,000 5hmC sites were identified in S2 cells, most of which were in mRNAs (Delatte et al. 2016). Like RNA 5mC, 5hmC is most prevalent in the coding regions, followed by the 5'UTRs, intron boundaries and 3'UTRs of mRNAs. RNA 5hmC on mRNAs is associated with increased translation activity (Delatte et al. 2016).

Until recently, the exploration of 5mC demethylation to 5hmC was limited by technology. Due to the low abundance of 5mC and 5hmC *in vitro*, it was difficult to detect the dynamic change of these two modifications accurately. The difficulty in distinguishing the two modifications is due to the similar size and molecular weights of 5mC and 5hmC. However, with the development of the third-generation Nanopore sequencing technology, how RNA 5hmC is demethylated, generated and regulated will quickly be revealed.

1.5. RNA 5mC in plants

Understanding the function of RNA 5mC in plants lags behind that of animals and yeast. At the start of my PhD, about 1,000 RNA 5mC sites were identified in the model plant *A. thaliana* using either RBS-seq or 5mC RNA Immunoprecipitation-sequencing (5mC-RIP-seq, Cui et al. 2017; David et al. 2017; Maekawa & Yanagisawa 2021). David et al. (2017) detected more than 1,000 highly-confidence 5mC sites (Figure 1.5A), and these sites were in the shoots (201), roots (859), and siliques (128) using the RBS-seq (Figure 1.5B). Most of these sites were tissue-specific, and only a small proportion was found in all three tissues. Another group found more than 6,000 RNA 5mC sites in young seedlings by applying 5mC-RIP-seq (Cui et al. 2017). Despite differences in 5mC abundance and tissue distribution, data from both groups showed that mRNAs occupied over 90% of 5mC sites, and only a small proportion was in tRNAs and other ncRNAs (Cui et al. 2017; David et al. 2017). Unlike in animals, mRNA 5mCs mainly exist in the CDS region, enriched before the stop codon and after the start codon, followed by 3'UTR and 5'UTR (Cui et al. 2017), implying that 5mC may have different context-specific functions. Interestingly, poorly translated mRNAs enriched with 5mC were associated with non-translating monosomes, while translating polysomes had a low abundance of 5mC levels (Cui et al. 2017).

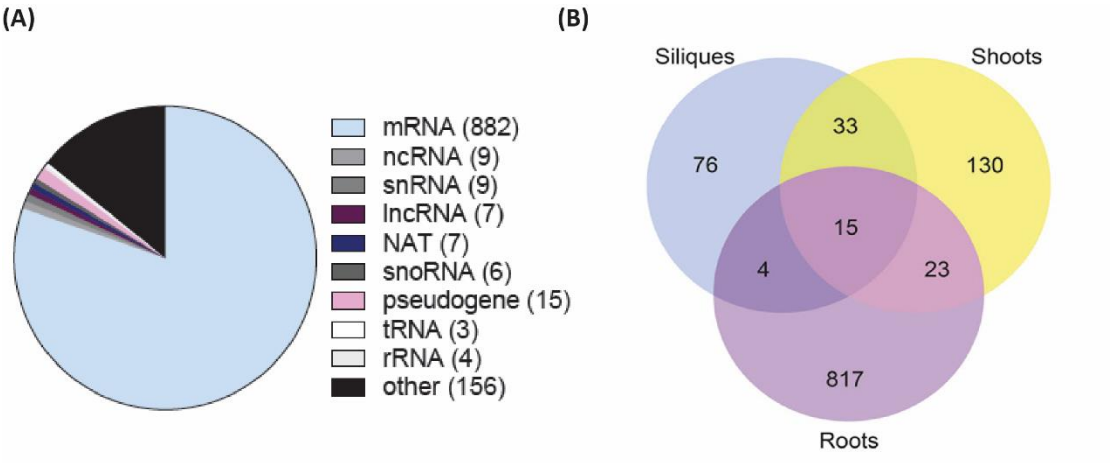


Figure 1.5: Distribution of RNA 5mC in the root, shoot and silique transcriptomes of *A. thaliana* (David et al. 2017)

(A) RNA 5mC sites were identified transcriptome-wide by using RBS-seq. More than a thousand sites were identified in mRNA, ncRNAs, rRNAs, tRNAs, snoRNAs, and NATs. The majority of the sites were in mRNAs. **(B)** A Venn diagram showing the numbers of 5mC sites amongst the three tissues; siliques, shoots and roots. Overlapping circles represent common 5mC sites amongst the tissues.

1.5.1. The distribution and functions of TRM4B-dependent 5mC sites

RNA 5mC methyltransferase tRNA-specific methyltransferase 4B (TRM4B/NSUN2) is an essential methyltransferase in *A. thaliana*. In *trm4b* mutants, 5mC sites are abolished on mRNAs, tRNAs and ncRNAs (Burgess, David & Searle 2015; David et al. 2017). Most TRM4B-dependent mRNA 5mC sites exist in the coding region (Cui et al. 2017; David et al. 2017). Like animals and yeast, TRM4B methylates cytosine at positions C_{48}, C_{49} and C_{50} in tRNAs. While only one C_{72} site in $tRNA^{Asp}_{GTC}$ depends on TRM4B in plants (Burgess, David & Searle 2015). TRM4B is required for 5mC methylation in shoots, roots and siliques. The *A. thaliana trm4b* seedlings have reduced 5mC sites in the root, reduced cell division rates of stem cells in the root apical meristem and subsequently shorter primary roots than the wild type (David et al. 2017). It is possible that the proliferative capacity of stem cells was reduced due to the loss of TRM4B. Another explanation is that the loss of 5mC sites in the CDS caused the decreased mRNA expression levels and transcripts levels of root genes, like *Short Hypocotyl 2* (*SHY2*) and *Indoleacetic Acid-Induced Protein 16* (*IAA16*) that are involved in root development (Cui et al. 2017). Also, *trm4b* seedlings display increased sensitivity to oxidative stress (David et al. 2017). In humans and mice, tRNAs 5mC protect tRNA from oxidative stress-induced cleavage (Blanco et al. 2014).

These findings highlight the significant status of RNA 5mC in regulating gene expression, cell development and stress response in plants. Plants might have their own regulatory pathways on RNA 5mC to perform special biological functions. However, so far, the exploration of the functions of RNA 5mC is still limited. Although the methyltransferase activity of TRM4B has been proved, there are a large number of TRM4B-independent 5mC sites existing in mRNAs, tRNAs and ncRNAs. How these cytosine sites are methylated and what their methyltransferase is remaining unclear. In addition, as mentioned above, whether RNA 5mC could be dynamically reversed to unmodified cytosine through some unknown demethylation pathways is still a mystery in the plant. With the development of advanced technologies and gradual emphasis on post-transcriptional modifications, it can be said that these questions and unknowns will be answered in following studies.

1.5.2. RNA 5mC guides mRNA over graft junctions

Long-distance transport of mRNA transcripts is an important and poorly understood mechanism in regulating distal plant growth and development. For example, phosphate homeostasis regulatory *At4* (AT5G03545) is a directionally mobile mRNA that controls the Pi transport and Pi homeostasis by sequestrating the shoot-to-root movement of miRNA *miR-399*, a regulator of the Pi throughput from the root to shoot *(Franco-Zorrilla et al. 2007)*. The mechanism of mobile mRNA loading, transport and unloading are still unclear.

Recently, 5mC was linked to the transport of mobile mRNAs in *A. thaliana* (Yang, L et al. 2019). 5mC sites are highly enriched in over 300 mobile mRNAs. The deficiency of 5mC in *dmnt2 nsun2b* double mutants inhibited the mobile *Translationally Controlled Tumor protein 1* (*TCTP1*) mRNA transcript over graft junctions and, as a result, abolished shoot-derived *TCTP1* in the roots of grafted

plants. The transport of *TCTP1* mRNA is mediated by 5mC at C_{141}, and the mutation of C_{141} to T_{141} prevented the long-distance movement of *TCTP1*. In conclusion, this study showed that 5mC is a crucial factor in triggering the long-distance transport of mRNAs. The requirement in 5mC on long-distance mRNA movement in different plant species is still to be explored.

1.5.3. 5mC is involved in environmental stress response

Recently, the significance of RNA 5mC in responding to environmental stresses was highlighted in *O. sativa* and *A. thaliana*. The deficiency of RNA 5mC methyltransferases *trm4b* and *trdmt1* in *A. thaliana* resulted in reduced 5mC levels on tRNA and for *trm4b* also in other cellular RNAs and increased sensitivity to oxidative stress (Burgess, David & Searle 2015). RNA 5mC was also shown to play a role in heat tolerance in rice. Heat stress enhances 5mC abundance on several mRNAs involved in the photosynthesis and detoxification of the cytotoxic molecule methylglyoxal (MG). The loss of RNA 5mC in *OsNSUN2* mutants led to increased sensitivity to high temperatures and reduced efficiency of photosynthesis upon heat stress (Tang, Yongyan et al. 2020). In contrast, Cui et al. (2017) detected reduced 5mC levels after heat or drought treatments. The opposing conclusions may be caused by differences between the organisms or heat treatment conditions.

A clear understanding of RNA 5mC is still to be elucidated as a field, and the regulatory mechanisms and specific functions of most stress-responsive 5mC sites remain unclear. The studies outlined above have laid the foundation for further exploration of RNA 5mC in environmental stress responses.

1.5.4. Potential demethylation pathway of RNA 5mC in plants

Although no RNA 5mC demethylase has been discovered in plants, some evidence, like High-performance liquid chromatography (HPLC) identification of 5hmC, implies that RNA 5mC is demethylated. Interestingly, our laboratory identified a methylated cytosine at position C_{3349} of *MAIGO 5* (*MAG5*) in *A. thaliana* that varied in methylation levels across different tissues: 55% in siliques, 33% in seedlings and 26% in roots (David et al. 2017) suggesting an unknown regulatory mechanism. In saying this, it cannot be ruled out that other cellular factors influence C_{3349} *MAG5* methylation as opposed to RNA 5mC demethylation. During my master's research, I discovered 5 heat-sensitive 5mC sites in mRNAs of *A. thaliana*. The 5mC site at position C_{3349} of *MAG5* (*AT5G47480*) was increased from 48% to 71% after heat shock treatment (40°C, 20 mins). An increased methylation level after heat shock was also observed at other 5mC sites, including position C_{328} of *AT2G36120*, C_{121} of *At5g40395*, C_{389} of *AT1G71350* and C_{295} of *AT4G2376*. Unlike in *MAG5*, there was a reduced level of methylation under oxidative stress conditions in C_{295} of *AT4G23760*, only 12% compared to 29% under control conditions. The dynamic alteration of the 5mC level at these specific sites suggests multiple regulatory mechanisms and a demethylation pathway. Overall, the discovery of RNA 5hmC and dynamic change of RNA 5mC level advises that the oxidative demethylation pathway of RNA 5mC may exist in plants. While no TET-like protein was recognised in plants. If RNA 5mC can be demethylated, it will most likely be catalysed by one or more AlkB-like proteins. However, further studies are highly encouraged to uncover the mystery of the RNA 5mC demethylation pathway.

1.6. Conclusion

Thanks to the rapid advances in molecular technology, especially low-cost, high-throughput sequencing, research on RNA modifications has become a booming field. However, our understanding of this field is still in its infancy. There is still a long way to comprehensively reveal the functions and regulatory mechanisms of these elaborate, transient and dynamic RNA modifications. Much remains unknown about the molecular and biological functions and regulation pathway(s) of 5mC, especially in plants. Undoubtedly, these mysteries will be solved using improved methodology and technology (e.g., Nanopore sequencing) to detect high-confidence 5mC and 5hmC. The elucidation of corresponding methyltransferase and demethylase of each 5mC site is only the first step to decipher the 5mC epitranscriptome code. The elucidation of reader proteins will significantly enhance our understanding of the 5mC function.

1.7. Research aims and Objectives

Although RNA 5mC in responding to environmental stress (e.g., oxidative stress) has been gradually highlighted in plants, no clear evidence has pointed out the role of RNA 5mC in heat stress responses. More broadly, no RNA 5mC demethylase has been identified and characterized in plants. This project focused on addressing these questions. The main aim of this project is to investigate the role of RNA 5mC in heat stress response and the potential demethylation regulation pathway in *A. thaliana*.

The objectives of my research are:

(1) To identify transcriptome-wide RNA 5mC sites that are responsive to heat shock,

(2) To investigate the role of specific RNA 5mC sites on mRNA expression, and

(3) To identify AlkB-like candidates that regulate RNA 5mC in *A. thaliana*.

2. Chapter 2: The RNA 5-methylcytosine landscape after heat shock treatment in *A. thaliana*

2.1. Abstract

Plants have evolved intricate mechanisms to adapt to environmental perturbations, optimising their fitness. A largely unexplored regulatory mechanism is RNA modifications and their role in heat stress tolerance. Of the numerous RNA modifications, 5mC is highly abundant in eukaryotic cellular mRNAs and non-coding RNAs. However, the functional roles of RNA 5mC in plants are still largely unknown. In this Chapter, we investigated the impact of heat shock treatment on transcript abundance, RNA 5mC and translation. Our RBS-seq data showed that heat shock-induced changes in abundance of 3,367 mRNAs, 1,385 were increased, and 1,982 were decreased. A total of 3,141 RNA 5mC sites before and 5,728 sites after heat shock treatment were detected. After heat shock treatment, 3,657 new 5mC sites were observed, and 1,070 5mC sites were lost. For the first time, we also observed that 70-80% of mRNAs with 5mCs contained clustered sites under both control and heat conditions. Using a Firefly bioluminescence reporter, we tested the effects of RNA 5mC on translation and observed that most 5mC sites could not directly affect the reporter bioluminescence. However, two hypermethylated regions derived from the 5'UTR and exon 2 of *AT1G05340* were found to increase the reporter compared to the control after heat shock treatment. In summary, our findings highlight the critical role of RNA 5mC in post-transcriptional regulation after

heat shock treatment.

2.2. Introduction

Global warming has increased the frequency of extremely high temperatures, which causes reduced crop growth and productivity in some agricultural regions (Hasanuzzaman et al. 2013; Lesk, Rowhani & Ramankutty 2016; Ohama, N. et al. 2017; Pagamas & Nawata 2008). As sessile organisms, plants cannot escape environmental stresses like high temperatures, and the leaves are often the most thermally damaged tissue, as observed by leaf curling, reduced leaf water content, or even dying. Heat stress can significantly impact the biochemical and physiological processes and alter gene expression in plants (Driedonks et al. 2015; Nishad & Nandi 2021). For example, excessive heat negatively affects photosynthesis, particularly the inactivation of photosystem-II (PS-II), and as a result, reduces total photosynthetic yield (Tang, Yunlai et al. 2007; Zhao, Q et al. 2018). At high temperatures, the chloroplast and mitochondrial electron transport system can be damaged by oxidative stress leading to chloroplast and mitochondria dysfunction (Hu, S, Ding & Zhu 2020; Nishad & Nandi 2021). The outer plant cell membrane is also susceptible to high temperatures leading to changes in membrane fluidity that can result in electrolytic leakage. Reactive oxygen species, like superoxides, hydrogen peroxide, and hydroxyl radicals, can cause lipid peroxidation of the cell membrane (Choudhury et al. 2017; Ruelland & Zachowski 2010; Savchenko et al. 2002). These reactive oxygen species can also damage DNA and RNA.

Heat stress responses in plants involve the induction of gene expression. For example, the well-known *HSF* and *HSP* genes are induced immediately after heat shock, and *HSFs* further promote *HSP* transcription (Hu, W, Hu & Han 2009; Tian et al. 2021; Xue et al. 2013). *HSPs* mainly act as molecular chaperones and function to stabilize the structure of other proteins, preventing protein

aggregation and repairing denatured proteins during heat stress conditions (Molinier et al. 2006; Tripp, Mishra & SCHARF 2009).

Heat stress also influences the epigenome in plants (Liu, J et al. 2015; Ueda & Seki 2019). In heat-sensitive genotypes of *G. barbadense*, *B. napus* and *G. max*, heat stress leads to the genome-wide reduction of DNA 5mC levels, especially at CHH sites (Ma et al. 2018). While in heat-tolerant lines of *G. hirsutum*, DNA 5mC increased after heat stress (Min et al. 2014). In *C. reinhardtii*, under heat stress, HSF1 promotes histone H3/4ac and represses H3K4 mono-methylation at the *HSP22F* promoter to active *HSP22F* transcription (Strenkert et al. 2011). In *G. hirsutum*, heat stress induces a short-term increase of H3K4me2 level and a long-term increase of Histone H4 lysine 5 acetylation (H4K5ac) level, which leads to nucleosome relaxation and the transcriptional activation of HS-response genes, such as *GhHSFA1a*, *GhHSFA2*, *GhHSP3*, *GhRBCS*, *GhERF1A* and *GhHXK1* (He, S et al. 2022). Chromatin architecture is also influenced by heat stress. For example, HSF1 and CRR1 (Dihydrodipicolinate reductase-like protein) open the closed chromatin conformation after heat stress at the *HSP22F* promoter and actives transcription in *C. reinhardti (Strenkert et al. 2011)*. While in *A. thaliana*, an ATP-dependent chromatin-remodelling gene *AtCHR12* mediates the growth arrest of flower buds and primary stems after heat stress (Mlynárová, Nap & Bisseling 2007). ncRNAs also play a role in heat stress and heat stress memory. For example, miR156 is induced by heat stress, leading to AGO-dependent cleavage of *SQUAMOSA promoter-binding protein-like* (*SPL*) genes, and miRNA156 mutants have impaired HS memory (Stief et al. 2014). Conversely, another heat-induced microRNA, miR398, negatively regulates the expression of *CSD1/2* (*copper/zinc superoxide dismutase 1/2*) and *CCS* (*Copper chaperone for superoxide dismutase*) genes leading to the accumulation of Reactive Oxygen Species (ROS) protein and increase of HSF and HSP levels (Guan

et al. 2013). Heat stress has also been demonstrated to have a role in RNA modifications.

The role of RNA modifications during heat stress has largely been unexplored. RNA *N6*-mA is one RNA modification shown to be involved in heat stress response (Zhou et al. 2015). Within the 5'UTR of stress-induced mRNAs, for example, *HSP70*, adenosine bases are methylated to *N6*-mA, promoting the initiation of cap-independent translation in HeLa cells (Zhou et al. 2015). The 5'UTR *N6*-mA of heat-induced transcripts is bound by and maintained by the nuclear *N6*-mA reader protein YTH domain-containing family protein 2 (YTHDF2), which limits the activity of the *N6*-mA 'eraser' protein FTO (Zhou et al. 2015). RNA 5mC was recently linked to heat stress tolerance in rice (Tang, Yongyan et al. 2020). High temperature enhanced 5mC abundance of several mRNAs, including *β-LCY* (*Lycopene β-Cyclase*), *HO2* (*Heme oxygenase 2*), *PAL1* (*Phenylalanine Ammonia-Lyase 1*), and *GLYI4* (*Glyoxalase I4*), that are involved in photosynthesis and detoxification systems of cytotoxic molecule methylglyoxal (Tang, Yongyan et al. 2020). The *osnsun2* mutant was sensitive to high temperatures and reduced efficiency of photosynthesis upon heat stress, presumably due to the loss of RNA 5mC (Tang, Yongyan et al. 2020). This chapter explored the transcriptome-wide RNA 5mC response to heat shock in the model plant *A. thaliana*.

2.3. Materials and methods

2.3.1. Plant materials

The *A. thaliana* accession used throughout this project was Columbia-0 (Col-0). *Nicotiana benthamiana* (*N. benthamiana*) *was* used throughout this project, and the transgenic line used was the RDR6i line (kindly provided by Prof. Sir David Baulcombe, University of Cambridge). All *A. thaliana* and *N. benthamiana* seeds were desiccated after harvest and stored at 4°C in sealed 1.5 mL Eppendorf tubes for over two weeks to reduce seed dormancy.

2.3.2. Plant growth conditions

2.3.2.1. Media-based growth conditions for *A. thaliana*

A. thaliana seeds were sterilized by treating the seeds with chlorine gas for 16 hrs. The seeds were placed in an enclosed glass chamber with a beaker containing 2 mL of concentrated Hydrochloric acid (HCl, 25%) and 100 mL of Sodium hypochlorite (NaOCl, 12.5%). Sterilized seeds were plated on ½ MS media plates supplemented with 1% sucrose, stratified at 4°C in the dark for 3 days, and then grown at 21°C under long-day conditions (16 hr light and 8 hr darkness) for 10 days in Phoenix Biosystem controlled environment room with metal halide lights (100 µmol/m^2/s).

In order to induce heat shock, 10-day-old *A. thaliana* seedlings on ½ MS media plates were placed in a 40°C water bath for 30 min under the same light conditions. The water temperature was measured by using a thermometer. After heat shock treatment, 10 seedlings (about 100 mg) were harvested immediately into an RNase-free Eppendorf tube by snap frozen in liquid

nitrogen and stored at -80°C ultra-freezer. For the control group, 10-day-old seedlings without heat treatment were harvested directly into an RNase-free Eppendorf tube, snap-frozen in liquid nitrogen and stored at -80°C.

2.3.2.2. Soil-based growth conditions for *N. benthamiana*

N. benthamiana seeds were planted on soil mixture (2 portions of Debco Seed Raising Soil and 1 portion of Debco® Seed & Cutting Mix), and the pots were placed in a Phoenix Biosystem temperature-controlled room at 21°C with metal halide lights (photosynthetic-active radiation: 100 µmol/m^2/s). The plants were grown under long-day photoperiod conditions (16 hr light, 8 hr darkness).

2.3.3. RNA extraction, quantification and quality assessment

About 100 mg of leaf tissues were snap-frozen in liquid nitrogen and ground with two 3mm steel beads using a TissueLyser II (Retsch) at a shaking frequency of 25Hz for 1 min. Total RNA was purified, and DNA was digested using a Spectrum™ Plant Total RNA Kit, following the manufacturer's instructions. All RNA samples were stored in a -80°C ultra-freezer to prevent degradation.

RNA concentrations were measured using a Qubit™ RNA HS Assay Kit on a Qubit® 2.0 Fluorometer, following the manufacturer's instructions. RNA quality was assessed after RNase-free agarose gel electrophoresis. 500 ng of total RNA was mixed with equal volumes of 2x RNA loading buffer, incubated at 65°C for 5 min to denature the RNA and remove secondary RNA structure, and the denatured samples were loaded on a 2% agarose gel. Samples were separated at 100 V for 90 min and stained with Red Safe.

2.3.4. cDNA synthesis of RNA samples

500 ng of total RNA was reverse-transcribed to first-strand cDNA (complementary DNA) using the SuperScript™ III Reverse Transcriptase (ThermoFisher Scientific) using OligodT, following the manufacturer's instructions.

2.3.5. RNA Bisulfite-Sequencing (RBS-seq)

2.3.5.1. rRNA depletion

Total RNA samples were subjected to ribosome RNA depletion by using the Ribo-Zero rRNA Removal Kit (Plant leaf, Illumina). Successful rRNA depletion was verified using a Bioanalyzer 2001 (Agilent Technologies).

2.3.5.2. RNA sodium bisulfite treatment

2.5 µg of rRNA-depleted RNA samples were denatured at 75°C for 5 min and then reacted with pre-heated sodium bisulfite solution (40% sodium metabisulfite and 0.6 mM hydroquinone) in the dark for 4 hr. Sodium bisulfite treated RNA samples were passed through BioRad™ Micro Bio-Spin Columns (Bio-Gel P-6 in Tris Buffer) twice to remove the remaining salts, following the manufacturer's instructions. An equal volume of Tris buffer (pH 9.0) was added to the desalted RNA solutions and incubated at 75°C for 1 hr to allow the generation of uridine. 1/10 volume of 3 M NaAc, 3 volumes of 100% ethanol and 2 µL of 10 mg/ml glycogen were added to the RNA samples and incubated at -20°C overnight to precipitate the RNA. On the following day, RNA samples were precipitated by centrifugation in a benchtop centrifuge at 13,000 rpm, 4°C for 25 min and then washed with 70% ethanol twice. Cleaned RNA samples

were resuspended in 30-40 µL of ultra-pure water and stored at -80°C.

2.3.5.3. Library construction for RBS-seq

The RBS-library is constructed using the TruSeq Stranded Total RNA-Plant Kit (Illumina). As BS-treated RNA is already fragmented, the fragment step of the library preparation was not undertaken, and the first strand cDNA was quickly synthesised to prevent RNA degradation. The remaining steps of RNA library construction were performed as described in the manufacturer's instructions.

2.3.5.4. RBS-sequencing

RBS-library samples were sequenced on a HiSeq (2x 150nt paired-end) platform at the ACRF, Adelaide. For each group, three biological replicate samples were prepared and sequenced.

2.3.6. Data analysis of RBS-seq

2.3.6.1. Quality control and read alignment

The raw Illumina sequence was trimmed by TrimGalore! (version 0.4.2) to remove adaptor contamination and low-quality reads using a Phred score threshold of 20 (Krueger 2012). Trimmed reads were then aligned to the *A. thaliana* genome (TAIR 11) using the splicing-aware RBS-seq alignment tool, meRanGs, available as part of the meRanTK package (version 1.2.1, Rieder et al. 2016).

2.3.6.2. Gene expression analysis

Gene expression analysis was performed using the SAM or BAM files generated from meRanGs-mapped reads. Read counts annotated as exons were collected and normalised using the featureCounts model from the *subreads* package (Liao, Smyth & Shi 2013). A differential abundance analysis was performed using DESeq2 (version 1.34.0, Love, Huber & Anders 2014). In the DESeq results, genes associated with a p-value <0.05 and absolute value of log2 fold change >1 were considered as considered to have significant expression differences. During the DESeq2 process, quality control steps were performed, including PCA analysis for replicates clustering and the correlation of gene TPM (transcripts per million) in sample pairwise comparisons. Following gene differentiation expression, the expression of partial genes was performed by Pheatmap, an R package (version 1.0.12, Kolde 2012).

2.3.6.3. RNA 5mC calling

RNA 5mC sites on aligned reads were recognised by performing the meRanCall of the meRanTK (version 1.2.1, Rieder et al. 2016). The methylation calling of candidate 5mC sites was set as follows: the FDR parameter of 0.01, the minimum read coverage of 20, the base-calling quality score of >30 (for paired-end reads), the maximum number of duplicated reads of 10 (avoid the impact of PCR artefacts), the methylation rate of > 10%, and 5mCs existing in 2 or more replicates of the same group. The BS conversion of unmethylated cytosine can be prevented by the local secondary structure of RNA, leading to false-positive recognition of some 5mC sites. Hence, any RNA regions with the potential secondary structure were predicted and excluded at 75°C (temperature of BS conversion) using ViennaRNA (version 2.4.11, Lorenz et al. 2011), and only the

5mC sites on the single-stranded RNA were used for the following analyses.

2.3.6.4. 5mCs comparison after heat treatment

5mCs from the control and heat groups were compared using meRanCompare of the meRanTk package (version 1.2.1, Rieder et al. 2016). Gene normalization was performed for each sample using HTSeq (version 1.99.1, Anders, Pyl & Huber 2015) to validate the following analysis. The Cochran-Mantel-Haenszel test was proceeded to compare 5mC level differences between the two groups statistically. We only considered the candidate 5mCs that were either unique to one condition or a significantly different methylation level of another condition (p-value < 0.05 and FDR = 0.02) while being present in at least two of three replicates (Rieder et al. 2016).

2.3.6.5. Methylated cytosine annotation and GO analysis

Methylated cytosines were annotated with the associated transcripts and relevant gene IDs using the *intersect* function in the bedtools package (Quinlan & Hall 2010). The distribution of 5mCs was analyzed by RNAmod, a convenient, online-only tool (Liu, Q & Gregory 2019). Gene ontology (GO) analysis was performed by inputting the gene IDs into the Gene Functional Annotation Tool at the DAVID website (https://david.ncifcrf.gov/summary.jsp; version 6.8).

2.3.7. Constructing Luciferase reporters

2.3.7.1. Plasmid pGrDL-SPb

The dual luciferase assay system and plasmid pGrDL-SPb was previously described by Moyle et al. (2017).

2.3.7.2. Construction of plasmid pGrDL-SPb-5’

The plasmid pGrDL-SPb-5’ was constructed by modifying pGrDL-SPb. First, the SalI and PstI restriction sites were removed by digesting pGrDL-SPb with EcoRI (NEB) and PstI-HF® (NEB) restriction enzymes and then ligating to a double-stranded adapter Mu-SalIPstI using T4 DNA ligase (NEB). The adaptor Mu-SalIPstI contained two nucleotide changes, one in the SalI site and the other in the PstI site and was made by annealing two complementary oligonucleotides (IDT) that generated EcoRI and PstI overhang sequences (Supplementary Table 2.8). The constructed plasmid, pGrDL-SPb-3musp, was confirmed by enzyme digestion using SalI-HF® (NEB) and PstI-HF® (NEB) separately and Sanger sequencing of the inserted adapter using the primer: 5’_F-Luc_Seq_R (Supplementary Table 2.9).

To produce pGrDL-SPb-5’, SalI and PstI restriction sites were added to the 5’ UTR of *F-Luc* of pGrDL-SPb-3musp. A double-stranded fragment, In-SalIPstI (gBlock, IDT), was designed to have the same sequence between KpnI (2,989nt) and the KasI restriction site (3,461nt) except for SalI and PstI sequences between the 35S promoter and the *F-Luc* (3,402nt, Supplementary Table 2.8). pGrDL-SPb-3musp and In-SalIPstI were double digested using KpnI-HF® (NEB) and KasI (NEB), gel and PCR purified (Wizard® SV Gel and PCR Clean-Up System), respectively and then ligated together using T4 DNA ligase (NEB), following the manufacturer’s instructions. The subsequent plasmid pGrDL-SPb-5’ was confirmed by enzyme digestion using SalI-HF® (NEB) and PstI-HF® (NEB) and Sanger sequencing of the inserted region by using the primer: 5’_F-Luc_Seq_R (Supplementary Table 2.9).

2.3.7.3. Construction of double-stranded sensor fragments by annealing oligonucleotides

Double-stranded adaptors less than 90 bp used as methylation sensors were made by annealing two complementary oligonucleotides to produce adapters with complementary SalI and PstI overhangs. Oligonucleotides were ordered with a 5' phosphate (Supplementary Table 2.5**,** synthesized by IDT). 10 µL volume, 100 µmol of the top and complementary bottom oligonucleotides were mixed with 10 µL 2x annealing buffer (20 mM Tris pH 8, 2 mM EDTA, 100 mM NaCl). The mix was heated to 95°C for 5 min and then slowly cooled down to 12 °C by running the AutoDelta thermal program at a cooling rate of 0.1°C/ 40 sec in a 96-well Thermo Cycler (Applied Biosystems). Annealed double-stranded sensor fragments were diluted 100 times with Ultra-pure H_2O and stored at -20°C.

2.3.7.4. Construction of sensor fragments by synthesising double-stranded DNA

Double-stranded sensor fragments greater than 90 bp were synthesized as double-stranded gBlocks gene fragments (Supplementary Table 2.7). gBlocks (IDT) were double-digested by SalI-HF® (NEB) and PstI-HF® (NEB) restriction enzymes to generate complementary overhanging sites, following the manufacturer's instructions. Digests were purified by using a PCR Clean-Up System (Wizard®). Purified double-stranded fragments were diluted 10 times with Ultra-pure H_2O and stored at -20°C.

2.3.7.5. Ligation of sensor fragments to the 3'UTR of *firefly Luciferase*

Plasmid pGrDL_SPb was double digested by SalI-HF® (NEB) and PstI-HF® (NEB) restriction enzymes and gel purified as described above. Double-digested and purified pGrDL_SPb was ligated with diluted candidate adaptors using T4 DNA ligase (NEB) at room temperature for 1 hr. Ligated vectors were heat-shock transformed into One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen, Themo Fisher SCIENTIFIC) cells and incubated on kanamycin (50 mg/L) LB plates at 37°C overnight. Single colonies were selected and cultured in 5 mL LB with kanamycin at 37°C overnight. The recombinant plasmids were verified by PCR amplification using primers 3'_F-Luc_PCR_LP and RP (Supplementary Table 2.9) and Sanger sequencing of the inserted region using primer: 3'_Fluc_seq_R (Supplementary Table 2.9).

2.3.7.6. Ligation of sensor fragments to the 5'UTR of *firefly Luciferase*

Plasmid pGrDL_SPb-5' was double-digested by SalI-HF® (NEB) and SalI-HF® (NEB) restriction enzymes and ligated with candidate adaptors as described in section 2.3.5. The recombinant plasmids were verified by PCR amplification using primers: 5'_F-Luc_PCR_LP and RP (Supplementary Table 2.9) and Sanger sequencing of the inserted region using primer: 5'_Fluc_seq_R (Supplementary Table 2.9).

2.3.8. Transient expression by agroinfiltration in *N. benthamiana* leaves

Dual luciferase plasmids were transformed into electrocompetent *Agrobacterium tumefaciens* strain USDA (containing pSOUP) by electroporation (250 µFD resistance, 25 µFD capacitance and 1.8 k Volts) using 2mm cuvettes. Transformed Agrobacterium cultures were plated onto LB media plates supplemented with rifampicin (25 mg/L) and kanamycin (50 mg/L) and grown at 28°C for 48 hr. To prepare starter cultures, single colonies were selected and inoculated into 2 mL of LB media with rifampicin and kanamycin and grown overnight at 30°C in a shaking incubator (180 rpm). To produce working cultures, the 2mL starter culture was added to 25 mL LB media with rifampicin and kanamycin and incubated for 4-6 hr at 30°C in a shaking incubator until the OD_{600} reached 0.5. Working cultures were harvested by centrifugation at 24°C and 4,000 rpm for 20 min. Pelleted cells were resuspended in infiltration buffer (1X MS salt, 10 mM MES, 3% sucrose and 200 µM acetosyringone, pH 5.6) to achieve 0.5-0.6 OD_{600}. The resulting cell cultures were stored in the dark at room temperature for at least 1 hr before infiltration into leaves.

N. benthamiana seedlings were grown for approximately 3–4 weeks before agroinfiltration. The Agrobacterium suspension was infiltrated into the leaf apoplast of three expanded leaves per plant by gently applying pressure to the plunger of the disposable syringe while it was held against the abaxial surface of the leaf. Infiltrated plants were covered with a lid to prevent wind stress and grown for 72 hr. Each infiltrated leaf was treated as a biological replicate.

2.3.8.1. Quantitative dual Luciferase assays

N. benthamiana leaves were laterally divided into halves along the mid-vascular rib. The left half of the leaf was used as a control and infiltrated with the pGrDL-SPb (or pGrDL-SPb-5') Agrobacterium suspension. The right half of the leaf was infiltrated with pGrDL-SPb-sensor (or pGrDL-SPb-5'-sensor) Agrobacterium suspension. I ensured a 10 mm space between each infiltration site to avoid cross-contamination. Three leaves from one plant were treated as three biological replicates.

After 72 hr, Agroinfiltrated leaves were medially cut using a razor blade for the control and heat shock-treated samples. A punch of infiltrated leaf (about 5 mg) from SPb and SPb-sensor were harvested individually, snap-frozen in liquid nitrogen, lysed using a tissue-lyser II (Retsch), and briefly stored at -80°C prior to measurement. For heat-treated samples, Agroinfiltrated leaves were placed on ½ MS media plates, treated in a 40°C water bath for 30 min, harvested and snap frozen as described above.

Dual luciferase (Firefly Luciferase and Renilla Luciferase) assay extracts were prepared using the Dual-Luciferase Reporter Assay System kit (Promega). 5 mg of powdered tissue was mixed with 500 µL of the passive lysis buffer (1x PLB) provided in the kit, and the cellular debris was pelleted by using a benchtop microcentrifuge at max speed for 1 min. 10 µL of the supernatant was loaded into a well of a white flat bottom Costar 96 well plate (Corning).

The Dual luciferase assay was performed using a GloMax 96 microplate luminometer (Promega) with two reagent injectors. Luciferase assay reagent and Stop & Glo reagent are provided in the Dual-Luciferase Reporter Assay System kit (Promega). 50 µL of luciferase assay reagent and Stop & Glo reagent

were introduced to each cell with a 2-sec premeasurement delay followed by a 10-sec measurement period for each reporter assay. The dual luciferase result was statistically analysed using GraphPad Prism 6 software (version 6.00).

2.3.8.2. Qualitative Luciferase image assay

N. benthamiana leaves were divided into four parts and infiltrated with: negative control (infiltration buffer), positive control (pGrDL-SPb Agrobacterium suspension) and two sensor vectors. At least 10 mm space between each infiltration site was ensured to avoid cross-contamination. Three leaves from one plant were treated as three biological replicates. Infiltrated plants were incubated for 72 hr under the same growth conditions.

Agroinfiltrated leaves were harvested, placed on a ½ MS media plate and sprayed with 2.5 mM-solution (BioVision) with 0.01% (v/v) Tween 20 (Sun, Zheng & Zhu 2017). Sprayed samples were incubated in the dark for 5-10 min to allow the luciferin solution to react with the luciferase enzymes and quench the chlorophyll auto-fluorescence. The bioluminescence was captured using a CCD camera (Biorad) in Chemi-high sensitivity. The exposure time was adjusted for each leaf as required.

2.4. Results

2.4.1. RNA transcript abundance analysis of *A. thaliana* under control and heat shock treatment conditions

To explore the effect of our heat shock treatment on RNA transcript abundance and transcriptome-wide RNA 5mC levels, RBS-seq was performed on rRNA-depleted RNA from control and heat shock-treated *A. thaliana* seedlings in

triplicate. We obtained 239 million paired-end (150 nt) Illumina reads, about 30 million for each replicate. Sequenced reads were aligned to an *in-silico* bisulfite converted *A. thaliana* transcriptome (TAIR 11) using meRanTK software. We calculated that the bisulfite conversion rate for all samples was over 99.5%, demonstrating that the conversion was very efficient.

We first carried out the Principal Component Analysis (PC analysis) and Correlation analysis on all aligned RBS-seq samples to identify any major differences and correlations between the treatments or replicates. The greatest variance was identified between the control and heat shock treatment (98% variance, Supplementary Figure 2.1A), and a high positive correlation was observed amongst replicates (Supplementary Figure 2.1B), indicating that our heat shock treatment resulted in significant changes at the transcript level. The transcript abundance difference between the control and heat shock-treated samples was further analysed by running an MA plot (Figure 2.1A). RNA transcripts with significantly increased abundance after heat shock treatment are shown as red dots above the X-axis, while those with significantly decreased abundance after heat shock treatment are shown as blue dots below the X-axis (|log2 fold change|>0.5, P value <0.05). We identified 1,385 RNAs with increased abundance and 1,982 with decreased abundance after heat shock treatment (Figure 2.1A). To confirm that the heat shock treatment was successful, we examined the transcript abundance of 56 previously identified genes that increased after heat shock treatment, and these were mainly from the *HSF* or *HSP* families after heat shock treatment. Of these 56, 49 significantly increased in expression after heat shock treatment, while 7 remained at the same level as the control (Supplementary Figure 2.1C & D). This result is consistent with the data on Arabidopsis eFP Browser (Winter et al. 2007). We selected one heat-response gene, *HSP101,* that significantly increased in

abundance after heat shock treatment (log2 fold increase =10.62) and confirmed the upregulation by semi-quantitative rt-PCR (Figure 2.1B, Supplementary Figure 2.1C, primers are shown in Supplementary table 2.1). Together these indicated that our heat shock treatment successfully induced previously identified heat-response genes.

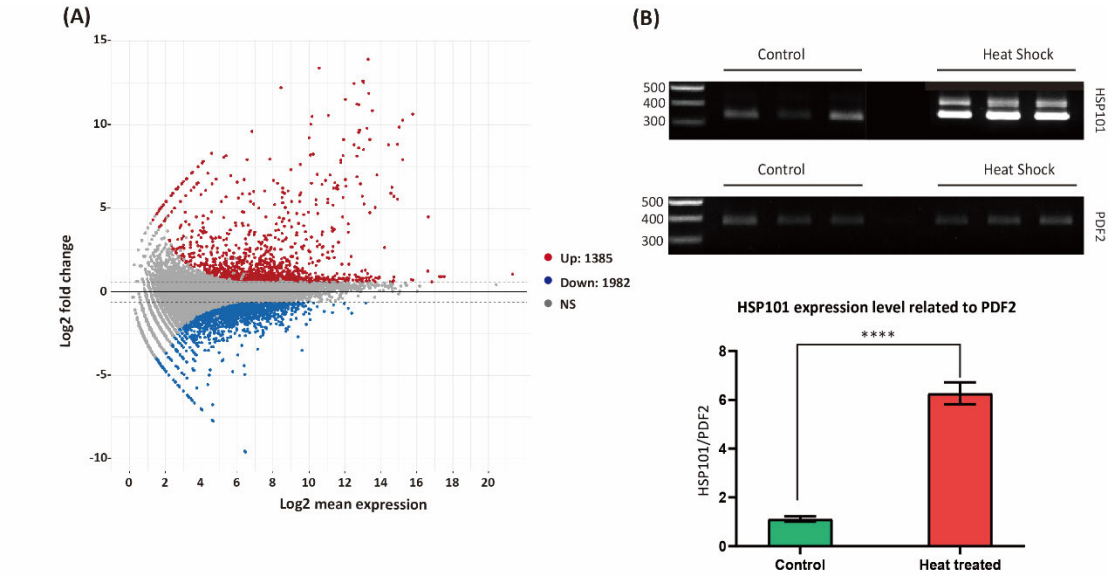


Figure 2.1: Transcriptome-wide identification of RNA transcript abundance after heat shock treatment of *A. thaliana*.

(A) MA-plot analysis of transcriptome-wide RNA abundance of control and heat shock-treated plants using the RBS-seq data. Each dot represents one RNA: M-values (Y-axis) indicate the log2 fold change in expression, and A-values (X-axis) indicates the mean abundance (n=3) of the RNA in the log2 scale. The red dots represent RNAs with significantly increased abundance after heat shock treatment (log2 fold change>0.5, P value <0.05), and the blue dots represent RNAs with significantly reduced abundance after heat shock treatment (log2 fold change< -0.5, P value <0.05), and grey dots represent genes without significant different expression level. After heat shock treatment, 1,385 RNAs were up-regulated, and 1,982 were down-regulated. **(B)** The RNA abundance of one heat-response gene, *HSP101*. The RNA level of the *HSP101* was detected by rt-PCR, normalized with the level of PDF2 (the internal control), and increased 6 times after heat treatment. P value <0.0001.

2.4.2. Discovery of RNA 5mC sites transcriptome-wide responding to heat shock treatment

We determined the heat-responsive RNA 5mC profile by comparing the 5mC distribution after control and heat shock treatment. To reduce the number of false positive 5mC sites, we set a high FDR parameter of 0.01 quality score of >30 and 5mCs sites were required to be present in two or more of the three replicates for the RNA 5mC calling. RNA secondary structure can prevent cytosine sites from bisulfite treatment and result in false positive identification of 5mC sites (Khoddami et al. 2019). I predicted RNA secondary structure transcriptome-wide using RNAfold from the ViennaRNA package (version 2.4.11, Lorenz et al. 2011). The maximum expected accuracy (MEA) of RNA secondary structures was calculated at 75°C (reaction temperature of the bisulfite treatment), and only those candidate 5mC sites not in predicted secondary structures were retained for further analyse.

Heat shock-treated samples had significantly higher 5mC abundance transcriptome-wide than the control samples. We detected a total of 3,141 RNA 5mCs in the control and 5,728 sites in the heat shock-treated samples, with methylation levels ranging from 0.05 to 1 (Figure 2.2A). Of these 5mC sites, 3,657 sites were only detected after heat shock treatment, 1,070 were only detected after control, and 2,071 were present after both control and heat shock treatments. Among 2,071 overlapping sites, 1,748 were up-regulated, 311 were down-regulated after heat shock, and 12 had consistent methylation levels after two different conditions. We then calculated the percentage of clustered RNA 5mCs. Two adjacent 5mCs within 20 bp were defined as clustered 5mCs. Interestingly, 70% of the 5mC sites in control samples appeared as clusters, and this percentage increased to 80% in heat-treated samples. We also analyzed the

distribution of transcripts containing 5mCs. After heat shock treatment, 244 RNA transcripts lost 5mCs, while 293 obtained new 5mCs (Figure 2.2B).

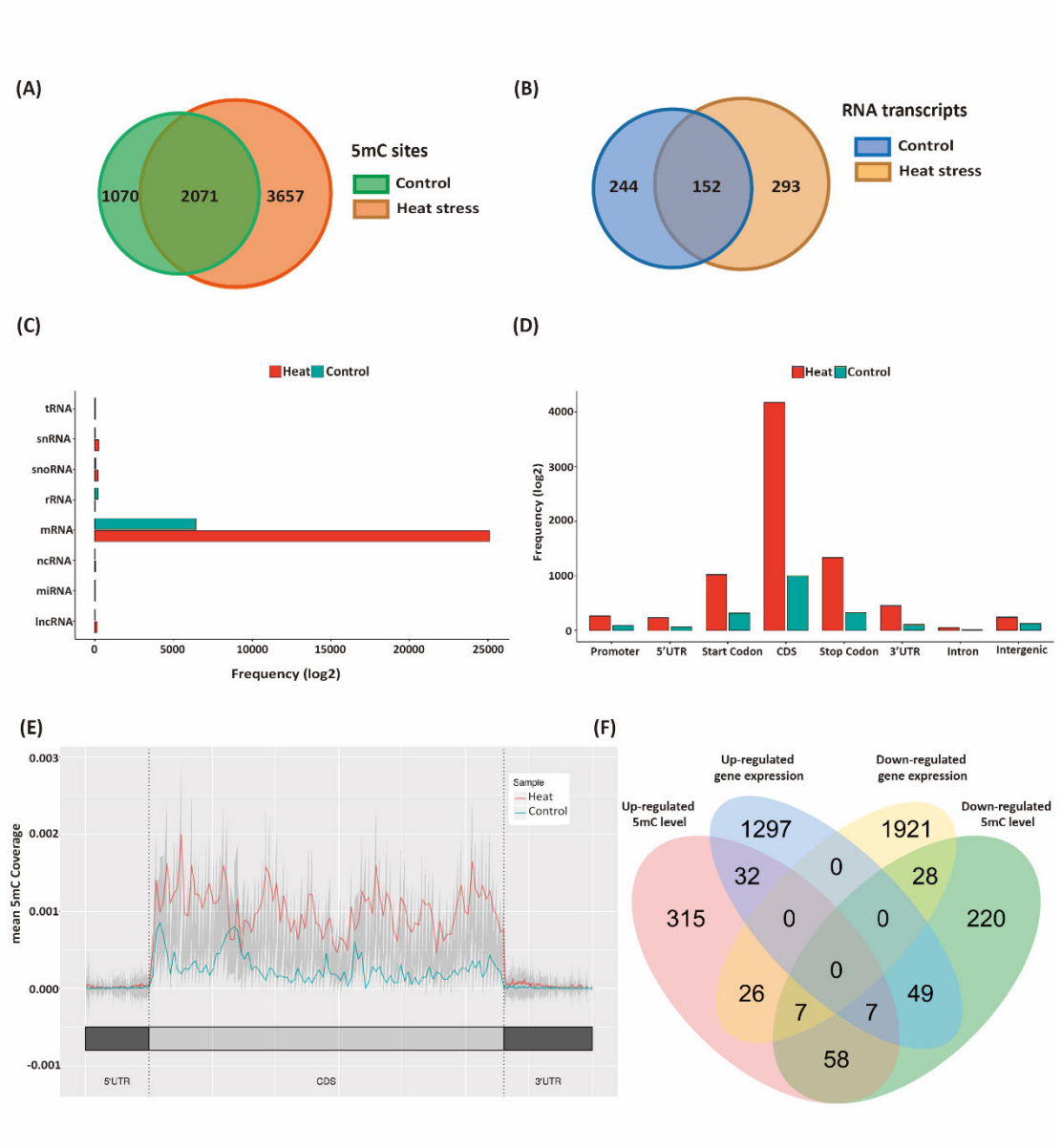


Figure 2.2: Transcription-wide distribution of RNA 5mC sites after control and heat shock treatment in *A. thaliana*.

(A) RNA 5mCs distribution after heat shock and control treatments. After RBS-seq, meRanTK analysis on the control and heat-treated samples, 1,070 5mC sites were only present in the control samples (shown in the green area), 3,657 5mC were only detected in the heat shock-treated samples, 2,071 5mC sites were present after both conditions shown in the overlapped area of the Venn diagram. **(B)** The distribution of transcripts containing 5mCs after control and heat shock treatments. Transcripts with unique 5mCs

to ideal conditions are shown in the blue area (244 in total); Transcripts with unique 5mCs to heat shock treatment are shown in the yellow area (293 in total), and transcripts with 5mCs after both two conditions are shown in the overlapped area (152 in total). **(C)** RNA classes that contained 5mC sites in control or after heat shock treatment. Eight RNA classes are shown on the Y axis, and the X axis shows the frequency of 5mC (Log2 fold change). **(D)** The distribution of 5mCs among different regions of mRNA after heat shock and control treatments. The X-axis shows types of RNAs containing 5mCs, and the Y-axis shows the frequency of 5mC (log2 fold change). **(E)** Spatial distribution of 5mC coverage on mRNAs after heat shock and control treatments. The X-axis shows the mRNA regions, and the Y-axis shows the mean coverage of 5mC (5mC/(5mC+C)). **(F)** Crossover comparison of genes with differences in expression and methylation levels after heat shock treatment. The Venn diagram shows the relationship between genes with increased (1,385) or decreased (1,982) abundance and genes with increased (445) and decreased (369) 5mC levels after heat shock treatment.

The transcriptome-wide distribution of 5mC was similar in both control and heat-treated samples, mainly concentrated on mRNA, followed by a small amount presenting in the snoRNA, ncRNA, snRNA, lncRNA, rRNA, tRNA and miRNA (Figure 2.2C). The heat-treated replicates had higher 5mC frequency on mRNA, snRNA, snoRNA, and lncRNA. Notably, the log2 frequency of 5mC on mRNA of the heat shock-treated samples was about 4 times that of the control samples (Figure 2.2C, Supplementary Figure 2.2A & B). While comparing with the control, the 5mC abundance remained unchanged on tRNA, ncRNA and miRNA and decreased on rRNA after the heat shock treatment (Figure 2.2C).

After heat shock treatment, the striking increased abundance of 5mC sites on mRNAs led us to further analyse the distribution along the transcript. In the control group, 5mC sites were mainly present in the CDS, although many sites were associated with the start and stop codons (Figure 2.2D). Increased numbers of 5mC sites after heat shock treatment were also observed in the CDS, start and stop codons, 3’ and 5’UTRs (Figure 2.2D). RNA transcripts were also identified in the so-called intergenic and promoter regions, and these transcripts also had increased 5mC after heat shock treatment but to a lesser extent. The distribution of mRNA 5mCs after heat shock treatment was largely

similar to that after control, but the frequency was greatly higher than that of the control treatment. We then analysed the 5mC coverage rate (5mC/(5mC+C)) through mRNAs (Figure 2.2E). The control had an about 0.0003 average 5mC coverage on the CDS, and a higher methylation rate near the start codon (0.0008), front of the CDS (0.00075), middle of the CDS (0.0006) and the stop codon (0.0004). After the heat shock treatments, the average 5mC coverage on the CDS increased to about 0.0012, and the highest level was observed near the start codon (0.002). On the 3'UTR and 5'UTR regions, increased 5mC coverage was also found after heat shock treatment (Supplementary Figure 2.2C & D). We observed no specific pattern of 5mC distribution in either the control or heat shock-treated samples. In summary, a significant enrichment of 5mC sites on mRNAs after heat treatment was observed, specifically in the coding region and the 3'UTR (Supplementary Figure 2.2E).

We also tested the correlation of RNAs containing 5mC sites and transcript abundance (Figure 2.2F). After heat shock treatment, among the transcripts with increased abundance, 32 contained up-regulated RNA 5mCs, 49 contained down-regulated 5mCs, and 7 contained up and down-regulated 5mCs. Of those transcripts whose abundance decreased after heat shock treatment, 28 had 5mCs with increased methylation level, 58 had 5mCs with reduced methylation level, and 7 had both increased and decreased 5mCs. While most transcripts with 5mCs did not show any connection with abundance difference after heat shock treatment. RNA 5mCs might have multiple roles in regulating transcript response to heat shock treatment, but these still need to be determined by further studies. We observed no apparent positive or negative correlation between 5mC levels and transcript abundance.

2.4.3. Identification of 5mC sites that vary after heat shock treatment

The dramatic changes in the abundance of mRNA 5mC after heat shock treatment aroused our interest in further exploring the potential role of 5mC in regulating transcription or translation to respond to heat shock. The first step was to screen candidate 5mC sites that were sensitive to heat shock treatment. We targeted 5mC sites that varied by more than 0.1 between the control and heat shock treatments, and these sites were not in an RNA region with a strong RNA secondary structure (P value <0.005, |log2 fold change|>0.5). We identified a total of 643 5mC sites that were only present in the heat shock-treated samples, 431 sites that were only present in the control samples, and 83 sites that showed increased methylation after heat shock treatment in mRNAs. The second step was to screen 5mCs that might regulate mRNA abundance from the candidates above. We analysed the association between the changes in 5mC levels and the changes in transcript abundance and targeted 5mC sites that appeared on mRNAs with increased, decreased or unchanged abundances after heat shock treatment (Supplementary Figure 2.3A).

We selected 47 5mC sites on 11 mRNA fragments that had increased methylation levels after heat shock treatment, 3 were single 5mC sites, and 44 were clustered sites (Figure 2.3, Supplementary Table 2.2). Another 26 clustered 5mC sites on 4 mRNA fragments that showed decreased methylation after heat shock treatment were also selected (Figure 2.3, Supplementary Table 2.3). The abundance of the mRNAs carrying these 5mCs in the control and heat shock treatments is shown in Supplementary Figure 2.3A. Two control fragments were also identified, one containing a single consistent 5mC site and the other without any 5mC sites (Supplementary Table 2.4).

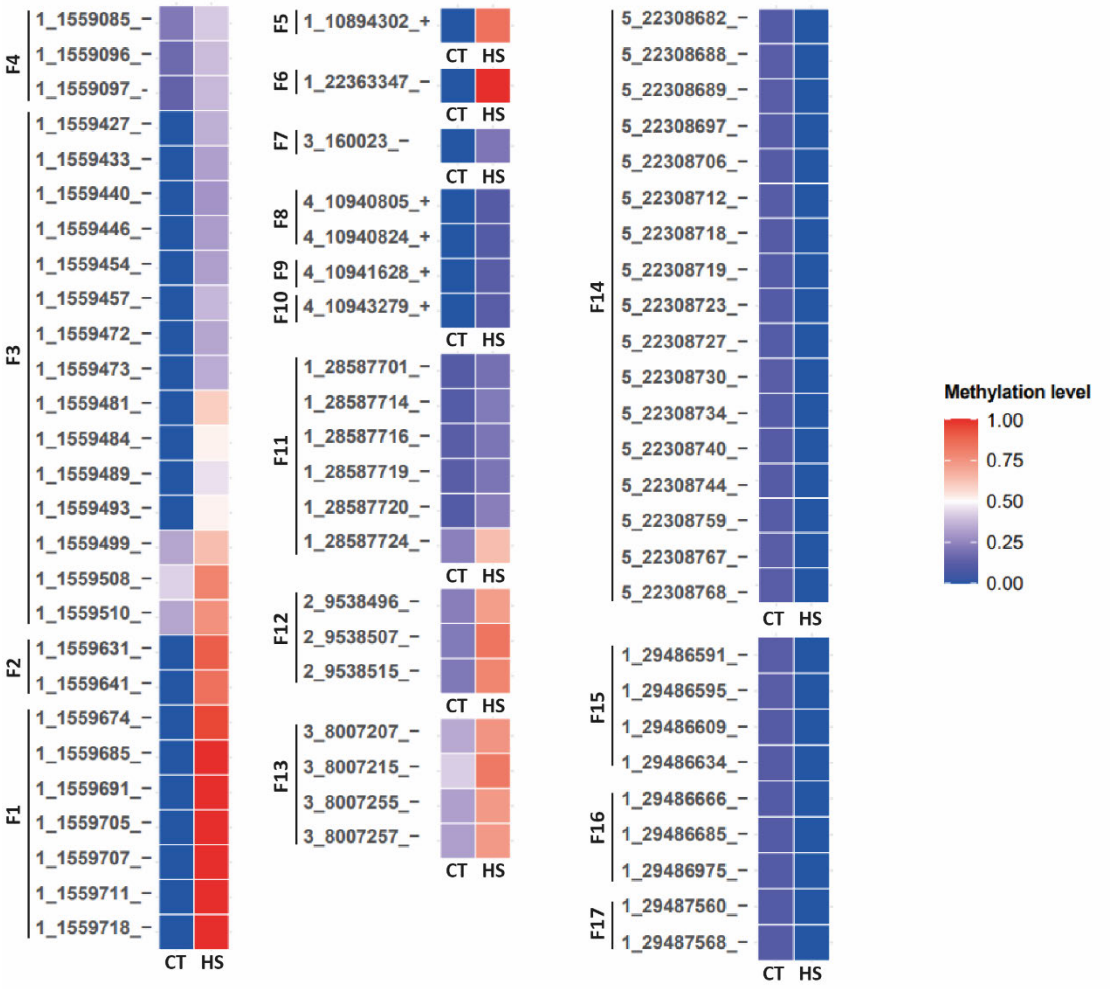


Figure 2.3: The methylation level of candidate 5mC sites in control and heat shock-treated samples

Heatmap analysis of methylation levels at 58 5mC sites on 17 mRNAs: each site was named in the format Chromosome_Position_Strand (Chr_refPos_strand). F1 to F17 are fragments 1 to 17; CT is the control sample; HS is the heat shock treatment. The methylation level was calculated by the formula 5mC/(5mC+C).

2.4.4. Identification of candidate fragments regulating the expression of firefly Luciferase

We then asked the following questions: what are the roles of individual and clustered 5mC sites, and what are the functional differences between 5mC sites in different positions in response to heat shock? To answer these questions, we

tested the effects of candidate 5mC sites on the expression of firefly Luciferase. The pGrDL-SPb plasmid constructed by Moyle et al. (2017) is a quantitative Firefly and Renilla Luciferase (F-Luc, R-Luc) reporter working in plants. In the dual Luc system, the F-Luc cassette reports the actual Firefly expression, while the R-Luc cassette is an internal control to normalize expression between replicates. To investigate the potential roles of different 5mC sites in regulating the expression of F-Luc, 19 candidate fragments, as shown in Supplementary Table 2.2-2.4, were synthesised *in vitro* and inserted into the 3' end of *F-Luc*, respectively (Supplementary Figure 2.3B). The agroinfiltration of constructed plasmids allowed the dual Luc systems carrying candidate fragments (Supplementary Figure 2.3C) to be expressed in *N. benthamiana* leaves. The candidate 5mCs of each fragment could be dynamically regulated by RNA 5mC methyltransferases and other regulators. Accordingly, our previous study found that the methylation of RNA 5mC by methyltransferase requires a minimum sequence length of 51nt (David et al. 2017). Hence, the candidate fragments we designed are not shorter than 51nt, expecting fragments 2 and 4, which are limited by their exon lengths. Infiltrated *N. benthamiana* leaves were incubated after the same growth conditions for 72 hr to maximize the expression of the dual Luc system.

The Luc expression of constructed pGrDL-SPb plasmids was first detected by the Luc image assay to qualitatively and visually measure the expression of F-Luc. Typically, *N. benthamiana* leaves were separated into 4 regions and infiltrated with negative control (infiltration buffer) on the top left, positive control (pGrDL-SPb) on the top right, and two candidate groups on the bottom left and right, respectively. After 72 hr incubation, infiltrated *N. benthamiana* leaves were supplied with 2.5 mM potassium luciferin, and their Luc images were captured by fluorescence. The Luc image results are shown in Figure 2.4. Interestingly, the

expression of F-Luc in 3'-1, 3'-3, 3'-13, 3'-14 and 3'-15 was lower than that in the positive control (pGrDL-SPb). In contrast, 3’-9 showed a higher F-Luc level than the positive control. No significant difference was found in other samples. Two control groups, 3’-18 and 3’-19, also expressed similar F-Luc levels as the positive control.

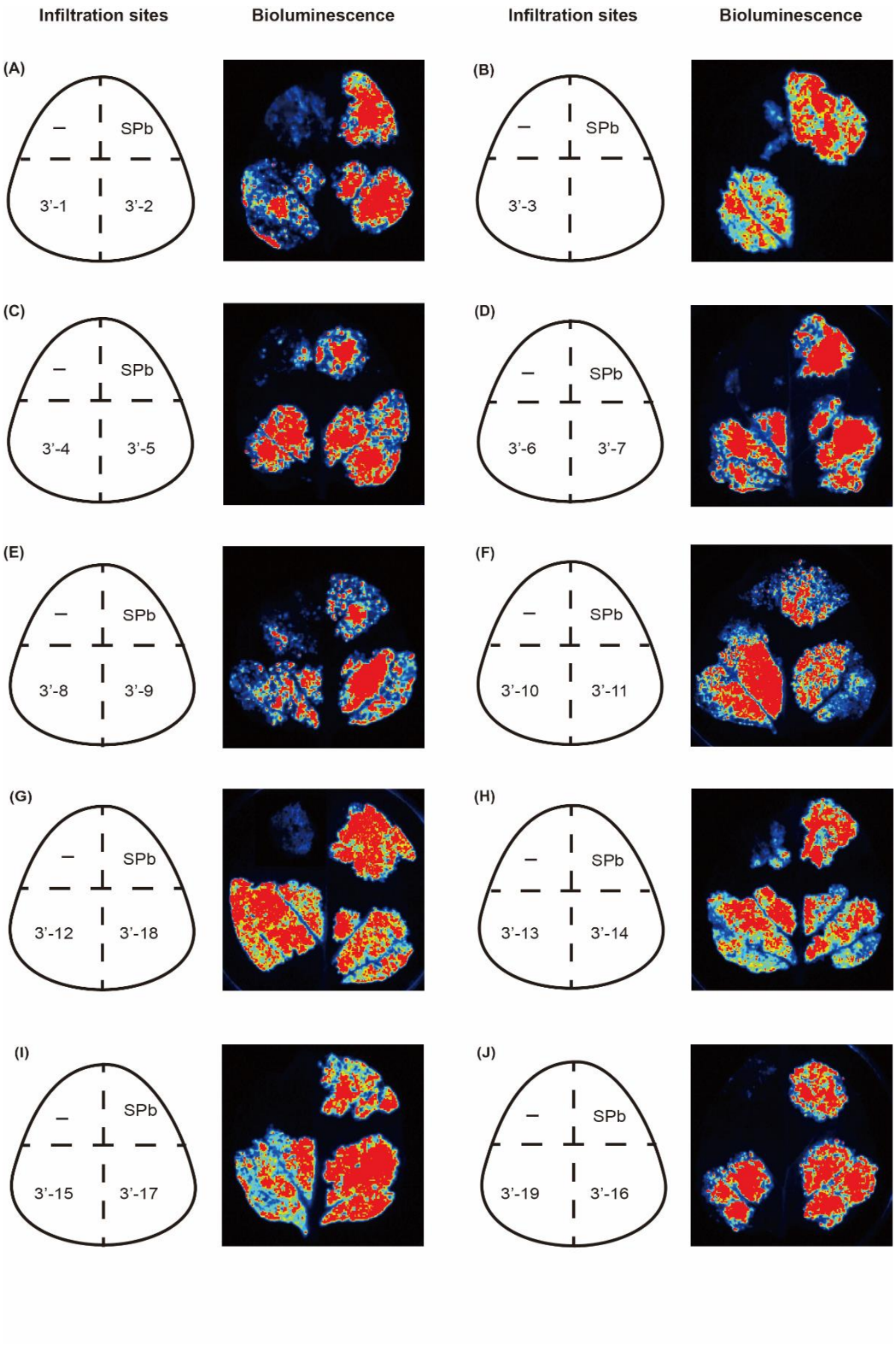


Figure 2.4: The effect of 19 sensor fragments on firefly Luciferase bioluminescence.

Nineteen sensor vectors were constructed by cloning a single sensor fragment at the 3'-end of *firefly* Luciferase in the dual luciferase vector Spb3'. Each gene fragment encoded

an mRNA that contained 5mC sites after heat shock treatment in my early experiment. 3-week-old *N. benthamiana* leaves were divided into 4 parts and infiltrated with either the negative control (-, infiltration buffer), positive control (SPb, dual luciferase vector without an RNA sensor), and two sensor vectors on the bottom left and right, respectively. Infiltrated plants were incubated for 72 hr before imaging using a chemi-high sensitivity program on a CCD camera (Biorad). The exposure time was adjusted for each leaf as required. The bioluminescence false coloured images of infiltrated vectors 3'1-3'19 are shown in **(A)** to **(J)**—left schematic of infiltration positions and right false coloured images after CCD imagining.

To further confirm our findings in the Luc image assay, we performed the quantitative and sensitive Dual-Luciferase® Reporter Assay (DLR assay) on all sample groups, which could detect their luminance activities of F-Luc and R-Luc sequentially and normalize the expression of F-Luc with R-Luc by F-Luc/R-Luc ratio (F/R ratio, Figure 2.5A). After 72 hr of infiltration, the average F/R ratio of the control group (pGRDL-SPb) was 2.03. In comparison, Moyle et al. (2017) found an about 2.8 F/R ratio of pGrDL-SPb at 72 hr of infiltration. The variation between the two experiments may be caused by differences in the growth conditions of *A. thaliana* leaves or Agroinfiltration. Remarkably, consisting of what we observed in the Luc image assay, the average F/R ratios of 3'-1, 3'-3, 3'-13, 3'-14 and 3'-15 were 0.72, 1.00, 0.51, 1.24 and 1.01, which was only 0.36, 0.49, 0.25, 0.6 and 0.49 times of the control, respectively. These results confirmed that adding fragments 1, 3, 13, 14 and 15 at the 3'UTR of *F-Luc* significantly reduces the expression of F-Luc. The decrease in the F/R ratio was also found in those samples indistinguishable from the control group in the Luc image assay results, including 3'-5, 3'-7, 3'-11 and 3'-12. The Luc image assay is not an accurate assessment and could be disturbed by various factors such as the number of infiltrated plasmids, the volume of subjected luciferin and the accuracy of the camera. This might be why we did not observe an obvious reduction of F-Luc in these sample groups in the Luc image assay. Conversely, the F-Luc expression levels of 3'-6 and 3'-9 were 0.4 times higher than the control. In comparison, no noticeable difference was observed in other groups.

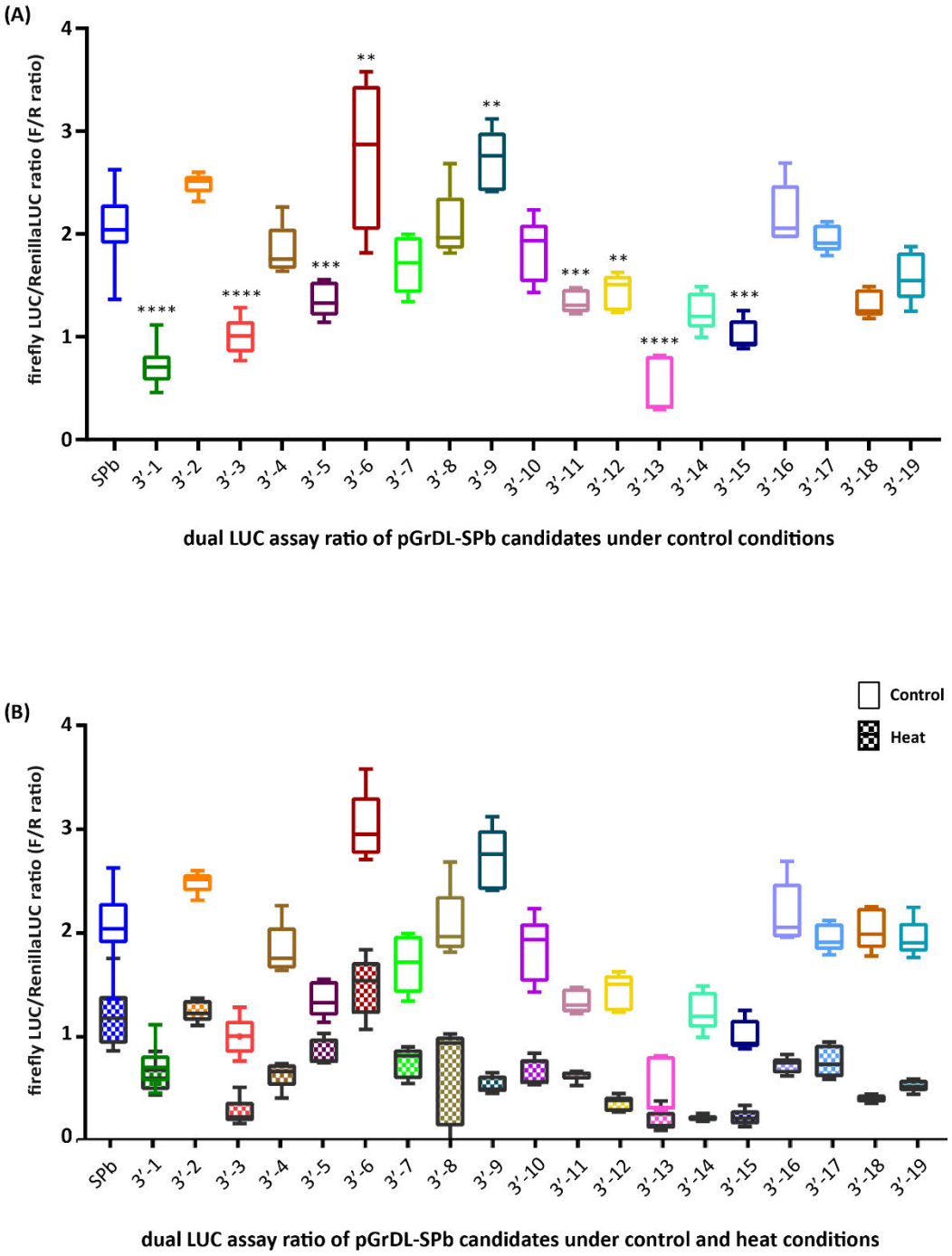


Figure 2.5: Quantitative dual Luciferase assay of sensor vectors 3'1 to 3'20 before and after heat shock treatment.

Constructed plasmids were agroinfiltrated to the *N. benthamiana* leaves and incubated for 72 h under control growth conditions. The infiltrated leaf tissue was evenly divided into two parts and treated either under control conditions or 40°C heat shock treatment for 30 min. About 1mg of leaf tissue and 5 replicates of each construct were used for the dual Luc assay.

The dual Luc assay results are analysed by box and whisker charts. The Y-axis shows the normalised luminance level of the F-Luc to the R-Luc (F/R ratio); The X-axis shows the name of each construct. Outer whiskers are the min and max values of each group. The box is drawn from the 1st to 3rd quartile of sample values; the middle horizontal line is the median value. **(A)** The Dual Luc assay result of 3'-1 to 3'-19 after control conditions. The F/R ratio of each group was compared with that of the control, and their difference was analysed by P value. **: P value < 0.05; ***: P value < 0.005; ****: P value < 0.0001. **(B)** The Dual Luc assay result of 3'-1 to 3'-19 after control and heat shock treatment. The F/R results of each group after different conditions were organised vertically. The colour box: ideal conditions; the black box with the colour dot: heat shock treatment conditions.

We also tested the effects of candidate 5mC sites on the heat tolerance of F-Luc. Specifically, 72 h post-infiltrated *N. benthamiana* leaves containing constructed plasmids were treated with heat shock (40°C for 30 mins) and performed DLR assay immediately as the same way above. The overall result of the DLR assay is shown in Figure 2.5B. After heat shock treatment, the F/R ratio of the control (pGrDL_SPb) decreased by about 50% from 2.03 to 1.18, due to the instability of the F-Luc over 30°C (Ebrahimi et al. 2012). Also, a similar decrease in F/R ratio was found in most groups after heat shock treatment, including 3'-2, 3'-4, 3'-6, 3'-7, 3'-8, 3'-10, 3'-11, 3'-16, 3'-17 and 3'-19. While the F/R ratio of 3'-3, 3'-9, 3'-12, 3'-14, 3'15 and 3'-18 after heat shock treatment rapidly dropped to only about 20% of that after control, indicating that the insertion of these fragments on the 3' end of *F-Luc* might accelerate the degradation of F-Luc or inhibited its expression after heat shock treatment. Interestingly, F-Luc appeared to be more stable after heat shock treatment in some groups, such as 3'-1, 3'-5 and 3'-13, with similar F/R ratios in the control and heat shock treatment, suggesting some potential roles of fragments 1, 5 and 13 in maintaining the expression and preventing further degradation of F-Luc after heat shock treatment.

In summary, among 13 candidate fragments (NO. 1-13) with up-regulated 5mC level after heat shock treatment, 1, 3, 5, 11, 12 and 13 significantly reduced the expression level of F-Luc after inserting to the 3' end of *F-Luc*. While the

insertion of 6 and 9 significantly enhanced the expression of F-Luc. Of 4 fragments (NO. 14-17) with down-regulated 5mC level after heat shock treatment, the insertion of 15 led to a significant reduction of F-Luc level. After introducing heat shock treatment, the F-Luc level of 3’-3, 3’-9, 3’-12, 3’-14 and 3’-15 decreased rapidly, while 3'-1, 3’-5 and 3’-13 remained consistent F-Luc levels as before. These data indicated that these fragments carrying candidate 5mC sites had different regulatory effects on F-Luc expression after different conditions and were worthy of further study. Due to the time limit, we could not test all fragments mentioned above, so 1, 3, 6, 9, 13 and 15, which had more significant influences on F-Luc expression, were selected for further experiments.

2.4.5. Different effects of inserting the candidate fragments at the 3’UTR or 5’UTR of *firefly Luciferase*

We then asked whether the insertion position of candidate fragments impacted firefly luciferase expression. To address this question, we constructed the plasmid pGrDL-SPb-5’F by first mutating the PstI and SalI sites at the 3’UTR of *F-Luc* and then inserting PstI and SalI sites into the 5’UTR of *F-Luc* (Supplementary Figure 2.3D & E). The candidate fragments 1, 3, 6, 9, 13 and 15 were inserted into pGrDL-SPb-5’F to produce six 5’ F-Luc sensor plasmids. I also constructed mutant versions of the six sensor fragments where the methylated cytosines were mutated to guanines. Hence, I mutated the 5mC sites of fragments 1, 3, 6, 9, 13 and 15 to G residues (guanine, Supplementary table 2.3 & 2.4) and inserted these mutated fragments into either the 3’UTR of pGrDL-SPb or 5’UTR of pGrDL-SPb-5’ of *F-Luc*. All 12 constructed plasmids were subjected to both the qualitative Luc imaging and the quantitative DLR assay 72 hrs after infiltration in *N. benthamiana* leaves. For quantitative DLR assays, individual *N.*

benthamiana leaves of a similar size were divided vertically along the midrib, each half infiltrated with the control, pGrDL-SPb, and the same sensor fragment. One infiltrated half was the control, the other infiltrated half was subjected to heat shock, and the bioluminescence was compared. This approach minimised the leaf-to-leaf variation that was observed in our earlier experiments. The overall Luc images and DLR results of fragments 1, 3, 6, 9, 13 and 15 are shown in Figures 2.6 to 2.11. Although some variation was observed amongst replicates within the same leaf, there was no significant difference in the average F-Luc/R-Luc value of pGrDL-SPb between individual experiments. The average F-Luc/R-Luc ratio of the pGrDL-SPb was about 2 in the control conditions and about 1 after heat shock treatment, consistent with our previous results.

Fragment 1 contains seven 5mCs, and the methylation level increased in the endogenous mRNA after heat shock in my previous experiments. The F-Luc/R-Luc ratio of 3'-1 was about half of the control plasmid under control conditions however was similar to the control after heat shock treatment (Figure 2.6 A, B, D & E). Interestingly, when all 5mC sites of fragment 1 in the 3' UTR of *F-Luc* were mutated (named mutated 3'-1), the F-Luc/R-Luc ratio was similar to the control under both before and after heat shock treatment, Figure 2.6 E), suggesting that the 5mC sites caused the reduced bioluminescence level of 3'-1. In contrast, the insertion of fragment 1 at the 5'UTR of *F-Luc* (5'-1) doubled the F-Luc/R-Luc ratio under both control and heat shock conditions compared to the control vector (Figure 2.6 D, E). When the seven 5mCs were mutated to guanines (mutated 5'-1), an extremely low F-Luc bioluminescence was detected in both the qualitative and quantitative Luc assays (Figure 2.6C, D & E). The low bioluminescence could be due to inadvertently creating additional start codon ATG and new open reading frames (ORFs), which prevented the F-Luc from being translated. Similar low F-Luc bioluminescence was also observed in sensor

plasmids 5’-3, mutated 5’-3, 5’-9 and mutated 5’-9 (Figures 2.7C and 2.9C). All these sensor plasmids have additional start codons (ATG) in their inserted fragments. Due to time constraints, we could not determine the cause of the low F-Luc bioluminescence in these plasmids. This does not exclude the possibility that 5mCs of fragments 1, 3 and 9 do not significantly affect the sensors.

Fragments 3 and 13 have four clustered 5mCs that increased in methylation in the endogenous mRNA after heat shock treatment in my previous experiment. Unlike 3’-1, no significant difference in the F-Luc/R-Luc ratio was observed between 3’-3 and mutated 3’-3 under control or heat shock treatments, suggesting that the low F-Luc/R-Luc ratio of 3’-3 was not caused by 5mCs of the fragment 3 (Figure 2.7). Similarly, mutation of the 5mCs in fragment 13 did not change the F-Luc expression, regardless of whether it was inserted into the 3'UTR or 5'UTR, under either control conditions or after heat shock (Figure 2.10).

Previously, fragment 6 was identified within an mRNA to contain a single 5mC site that was up-regulated after heat shock treatment. The F-Luc/R-Luc ratio of sensor 3'-6 was lower than in previous experiments, about 1.4 times that of the control plasmid under control conditions and 1.2 times that of the control plasmid after heat shock treatment (Figure 2.8A, B, D & E). This variation could be caused by differences between individual *N. benthamiana* leaves. When the only 5mC site within sensor 3’-6 was mutated (mutated 3’-6), the F-Luc/R-Luc ratio after control conditions was the same as the control but was slightly reduced compared to the control plasmid after heat shock treatment (Figure 2.8 B, D & E). Interestingly, the insertion of fragment 6 into the 5’ UTR of *F-Luc* reduced F-Luc bioluminescence to 0.7 times the control plasmid under control conditions and 0.9 times after heat shock treatment (Figure 2.8 A, C, D & E). In contrast, the mutated 5’-6 sensor produced an F-Luc/R-Luc ratio of 1.25

compared to the control plasmid under both control and heat shock conditions (Figure 2.8 C, D & E). It seems that fragment 6 and the 5mC site have opposite effects in 3' and 5' UTRs to regulate the expression of F-Luc.

Fragment 9 is another candidate that contained a single 5mC site in the mRNA that increased in methylation after heat shock in my earlier experiment. We did not observe any apparent differences between 3'-9 and the control construct under control or after heat shock treatment (Figure 2.9 A, B, D & E). However, mutation of the 5mC site of 3'-9 (named mutated 3'-9) led to a significant reduction of F-Luc bioluminescence after heat shock, implying a potential role of this 5mC site.

Fragment 15 contained two 5mC sites in the endogenous mRNA and was the only fragment with reduced methylation level after heat shock treatment in my earlier experiment. Consistent with my previous observation, sensor 3'-15 showed 0.7 and 0.6 times the F-Luc/R-Luc ratio compared to the control plasmid under control conditions and after heat shock, respectively. When 5mCs of fragment 15 in 3' UTR of *F-Luc* were mutated, the F-Luc/R-Luc ratio was rescued to a similar level as the control under both control and heat shock conditions (Figure 2.11 B, D & E). Unlike sensor 3'-15, the F-Luc/R-Luc ratio of 5'-15 was 0.3 times and 0.2 times higher than the control plasmid under control controls and after heat shock treatment, respectively. At the same time, the mutated 5'-15 sensor produces a low F-Luc/R-Luc ratio, only 0.65 of the control under control conditions and 0.5 times the control after heat shock treatment. These suggest that the 5mCs of fragment 15 inhibited the expression of F-Luc in the 3' UTR of *F-Luc* but promoted the expression of F-Luc in the 5' UTR of *F-Luc*.

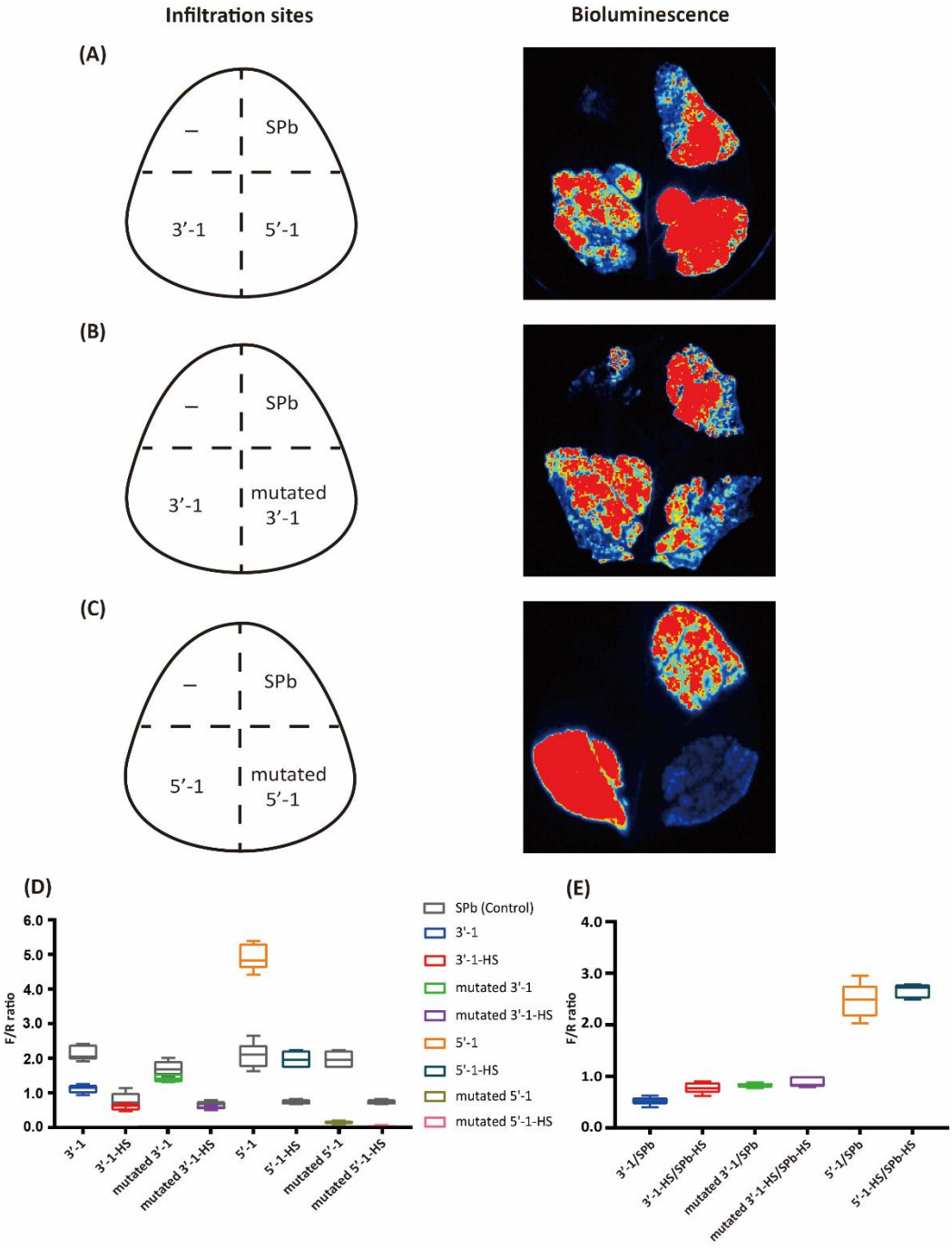


Figure 2.6: Mutational and position analysis of sensor fragment 1 on firefly Luciferase expression.

Fragment 1 and mutated fragment 1 were inserted into either the 3'UTR or 5'UTR of *F-Luc* of plasmids pGrDL_SPb or pGrDL-SPb-5'F, respectively. Constructed plasmids were agroinfiltrated to *N. benthamiana* leaves and incubated for 72 hrs. For the qualitative Luc assay, the *N. benthamiana* leaf was divided into 4 parts and infiltrated with either the negative control (infiltration buffer, top left), positive control (SPb plasmid, top right), and two sensor fragments (bottom left and right). After 72 hrs, the Luc bioluminescence was

captured using a CCD camera. For the dual luciferase (DLR) assay, the *N. benthamiana* leaf was vertically separated into 2 parts and infiltrated with either the positive control (SPb plasmid, left) or sensor constructs (right). After 72 hrs, the infiltrated leaf tissue was horizontally divided into 2 parts, one part was the control, and the other part was heat shock. Five replicates from each group were harvested and then performed for the DLR assay. The DLR results were analysed by box and whisker charts: The Y-axis shows the normalised luminance level of the F-Luc to the R-Luc (F/R ratio), and the X-axis shows the name of each group. Outer whiskers are the min and max values of each group, the box is drawn from the 1st to 3rd quartile of sample values, and the middle horizontal line is the median value.

(A) A qualitative Luciferase false colour image of sensor fragments 3'-1 and 5'-1. The positive control plasmid is SPb. Sensor 3'-1 showed lower F-Luc bioluminescence, while sensor 5'-1 showed higher F-Luc bioluminescence compared to the control. **(B)** A qualitative Luciferase false colour image of sensors 3'-1 and mutated 3'-1. Both 3'-1 and mutated 3'-1 showed lower F-Luc bioluminescence than the positive control. **(C)** A qualitative Luciferase false colour image of sensors 5'-1 and mutated 5'-1. 5'-1 showed higher F-Luc bioluminescence than the positive control. No obvious F-Luc signal was detected in mutated 5'-1. **(D)** The DLR assay of 3'-1, mutated 3'-1, 5'-1 and mutated 5'-1 after control and heat shock treatments. The F/R results of each group and its positive control were organised vertically. The colour box: the experiment group; the grey box: the corresponding positive control. **(E)** The normalized F/R ratio of 3'-1, mutated 3'-1, 5'-1 and mutated 5'-1 to their positive control. The F/R ratio of 5 replicates from each group was arranged from low to high and normalised with that of its positive control.

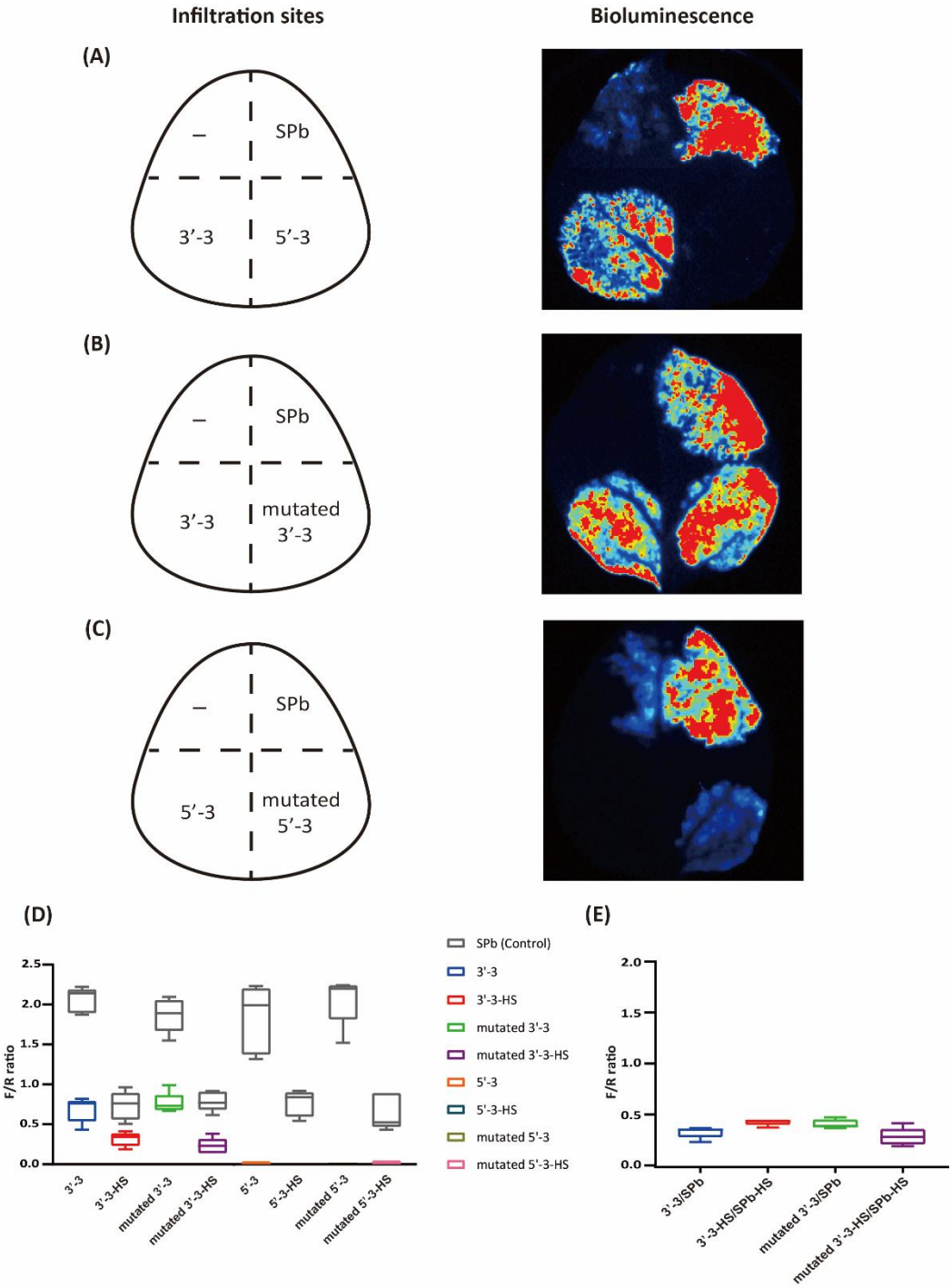


Figure 2.7: Mutational and position analysis of sensor fragment 3 on the bioluminescence of firefly Luciferase.

Fragment 3 and mutated fragment 3 were inserted separately into either the 3'UTR or 5'UTR of *F-Luc*. Qualitative Luc image and the DLR assays were performed as outlined above.

(A) A qualitative Luciferase false colour image of sensors 3'-3 and 5'-3. Sensor 3'-3 showed lower F-Luc bioluminescence than the positive control. No obvious F-Luc

bioluminescence signal was detected in 5’-3. **(B)** A qualitative Luciferase false colour image of 3’-3 and mutated 3’-3. Both 3’-3 and mutated 3’-3 showed lower F-Luc bioluminescence than the positive control. **(C)** A qualitative Luciferase false colour image of sensor 5’-3 and mutated 5’-3. No obvious F-Luc bioluminescence was detected in either sensor 5’-3 or mutated 5’-3. **(D)** Dual Luciferase (DLR) assay of sensors 3’-3, mutated 3’-3, 5’-3 and mutated 5’-3 under control conditions and after heat shock treatments. The F-Luc/R-Luc (F/R) results for each sensor and the corresponding positive control are shown above each other. The colour box: the experiment group; the grey box: the corresponding positive control. **(E)** The normalised F/R ratio of sensors 3’-3 mutated 3’-3, 5’-3 and mutated 5’-3 to the positive control. The F/R ratio of 5 replicates from each group was arranged from low to high and normalised with that of its positive control

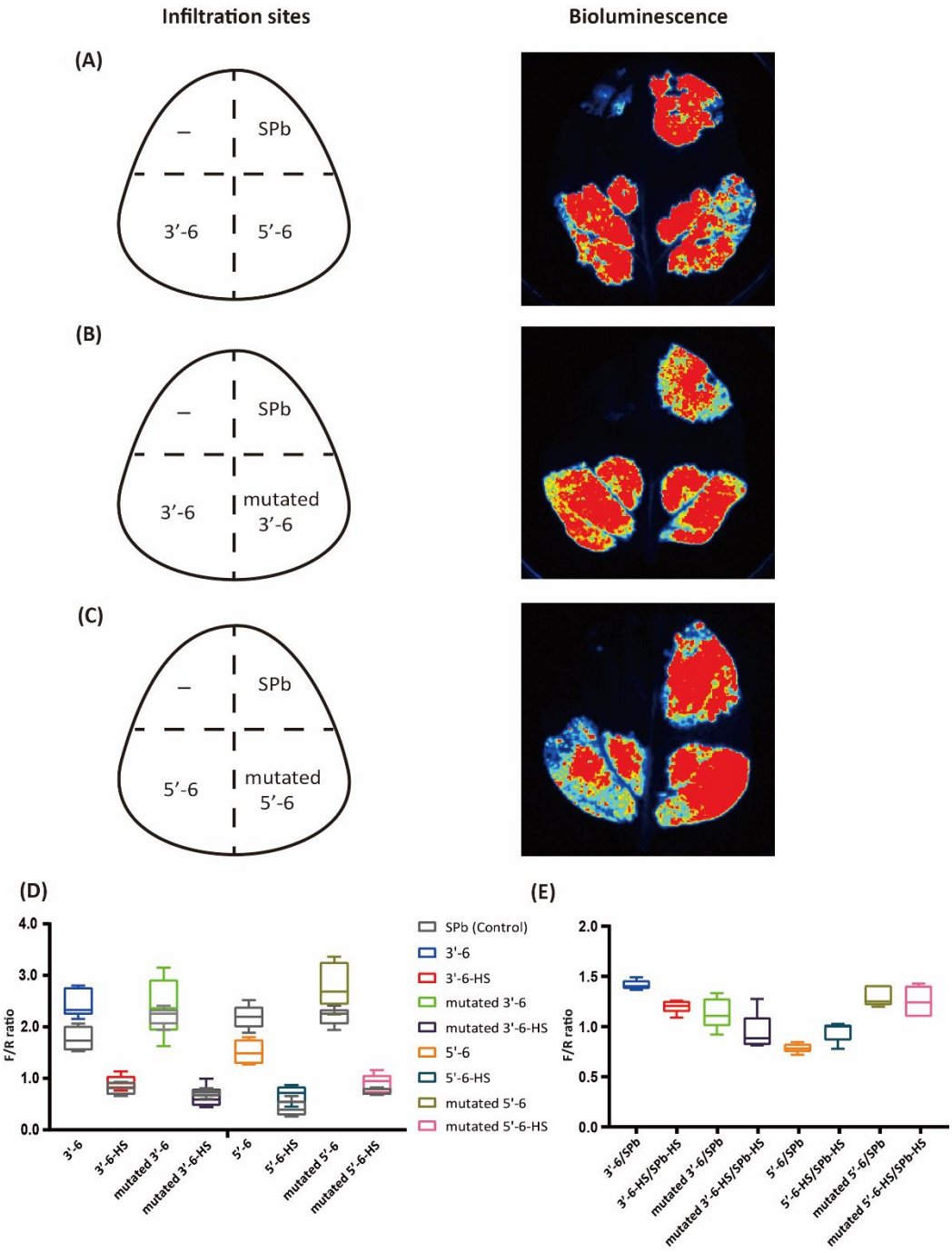


Figure 2.8: Mutational and position analysis of sensor fragment 6 on the bioluminescence of firefly Luciferase.

Fragment 6 and mutated fragment 6 were inserted separately to either the 3'UTR or 5'UTR of *F-Luc,* and the qualitative Luc image assay or DLR assay was performed as described above.

(A) The Luc image assay of 3'-6 and 5'-6. Compared with the positive control, 3'-6 showed a slightly higher F-Luc level, and 5'-6 showed a slightly lower F-Luc level. **(B)** A qualitative Luciferase false colour image of sensors 3'-6 and mutated 3'-6. Both 3'-6 and mutated 3'-6

showed slightly higher F-Luc bioluminescence than the positive control. **(C)** A qualitative Luciferase false colour image of sensors 5'-6 and mutated 5'-6. Compared with the positive control, 5'-6 showed a lower F-Luc level, and mutated 5'-6 showed a similar F-Luc level. **(D)** The DLR assay of sensors 3'-6, mutated 3'-6, 5'-6 and mutated 5'-6 under control conditions and after heat shock treatment. The F/R results of each group and its positive control were organised vertically. The colour box: the experiment group; the grey box: the corresponding positive control. **(E)** The normalized F/R ratio of 3'-6, mutated 3'-6, 5'-6 and mutated 5'-6 to their positive control. The F/R ratio of 5 replicates from each group was arranged from low to high and normalised with that of its positive control.

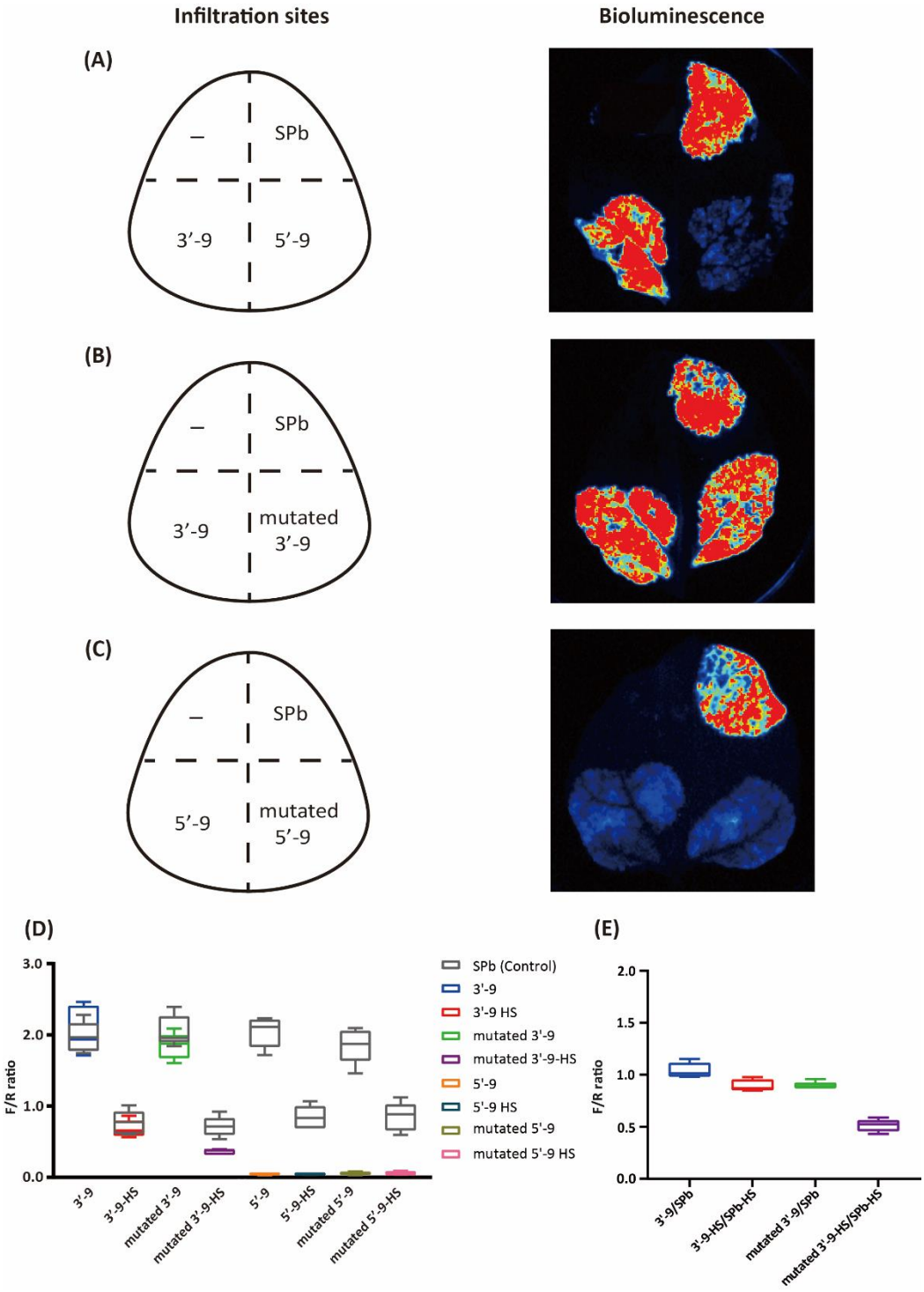


Figure 2.9: Mutational and position analysis of sensor fragment 9 on the expression of firefly Luc.

Fragment 9 and mutated fragment 9 were inserted either into the 3’UTR or 5’UTR of *F-Luc*, and a qualitative Luc image or a DLR assay was performed as described above.

(A) A qualitative Luciferase false colour image of sensors 3’-9 and 5’-9. Sensor 3’-9 showed similar F-Luc bioluminescence as the positive control. No obvious F-Luc

bioluminescence was detected in sensor 5’-9. **(B)** A qualitative Luciferase false colour image of 3’-9 and mutated 3’-9. Both 3’-9 and mutated 3’-6 showed similar F-Luc levels as the positive control. **(C)** A qualitative Luciferase false colour image of 5’-9 and mutated 5’-9. No obvious F-Luc level was detected in 5’-9 and mutated 5’-9. **(D)** The DLR assay of 3’-9, mutated 3’-9, 5’-9 and mutated 5’-9 after control and heat shock treatments. The F/R results of each group and its positive control were organised vertically. The colour box: the experiment group; the grey box: the corresponding positive control. **(E)** The normalized F/R ratio of 3’-9, mutated 3’-9, 5’-9 and mutated 5’-9 to their positive control. The F/R ratio of 5 replicates from each group was arranged from low to high and normalised with that of its positive control.

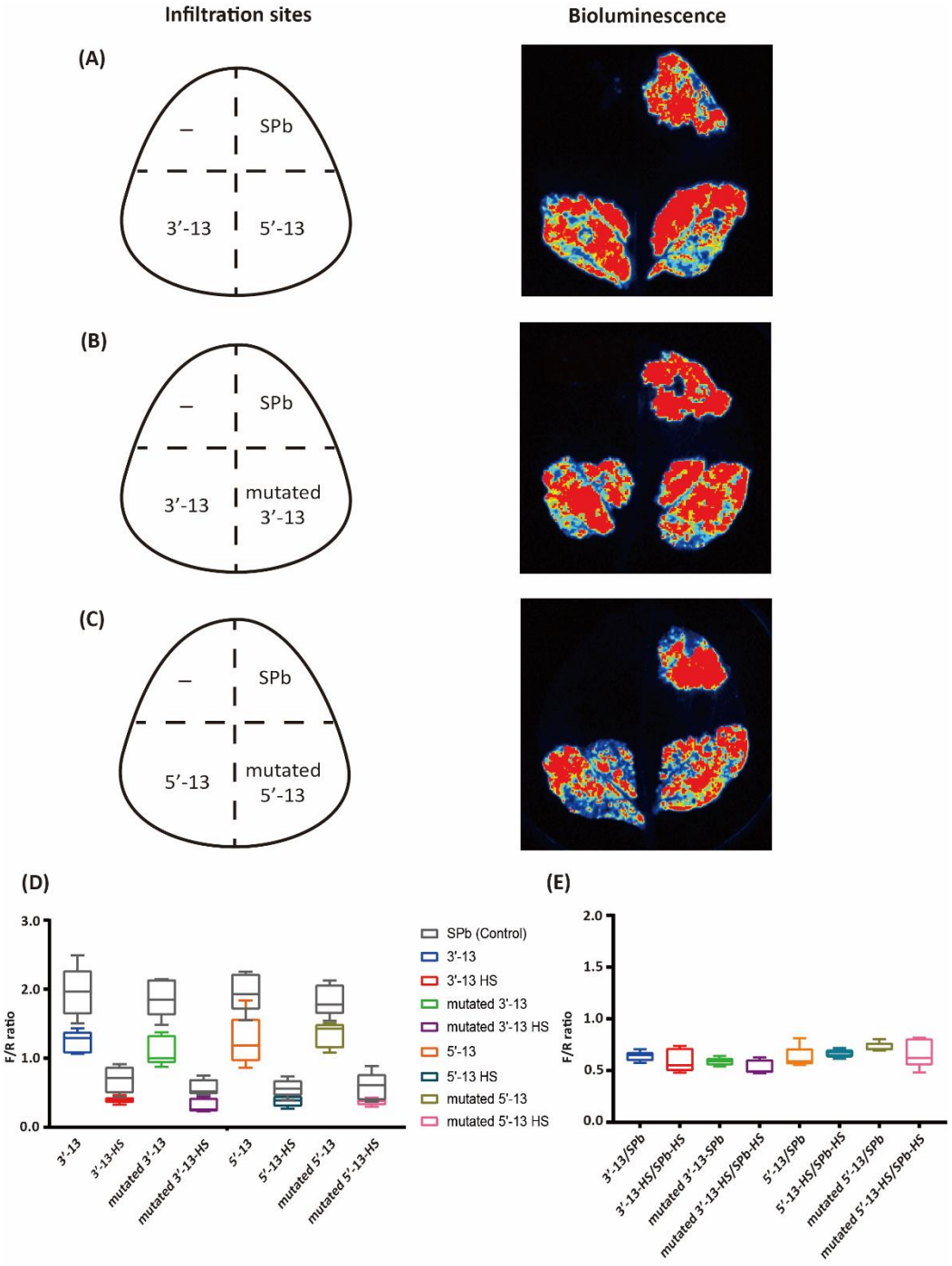


Figure 2.10: Mutational and position analysis of sensor fragment 13 on the expression of firefly Luciferase.

Fragment 13 and mutated fragment 13 were inserted separately to either 3'UTR or 5'UTR of *F-Luc,* and either the qualitative Luc image assay or DLR assay was performed as described above.

(A) A qualitative Luciferase false colour image of sensors 3'-13 and 5'-13. Both 3'-13 and 5'-13 showed slightly lower F-Luciferase bioluminescence than the positive control. **(B)** A

qualitative Luciferase false colour image of sensors 3'-13 and mutated 3'-13. Both 3'-13 and mutated 3'-13 showed slightly lower F-Luc bioluminescence than the control plasmid. **(C)** A qualitative Luciferase false colour image of sensors 5'-13 and mutated 5'-13. Compared to the positive control, 5'-13 and mutated 5'-6 showed lower F-Luc bioluminescence. **(D)** Dual Luciferase (DLR) assays of sensor 3'-13 mutated 3'-13, 5'-13 and mutated 5'-13 under control conditions and after heat shock treatments. The F-Luc/R-Luc (F/R) results of each group and its positive control were positioned above each other. The colour box: the experiment group; the grey box: the corresponding positive control. **(E)** DLR data from D but normalized to their respective positive control. The F/R ratio of 5 replicates from each group was arranged from low to high and normalised with the respective positive control.

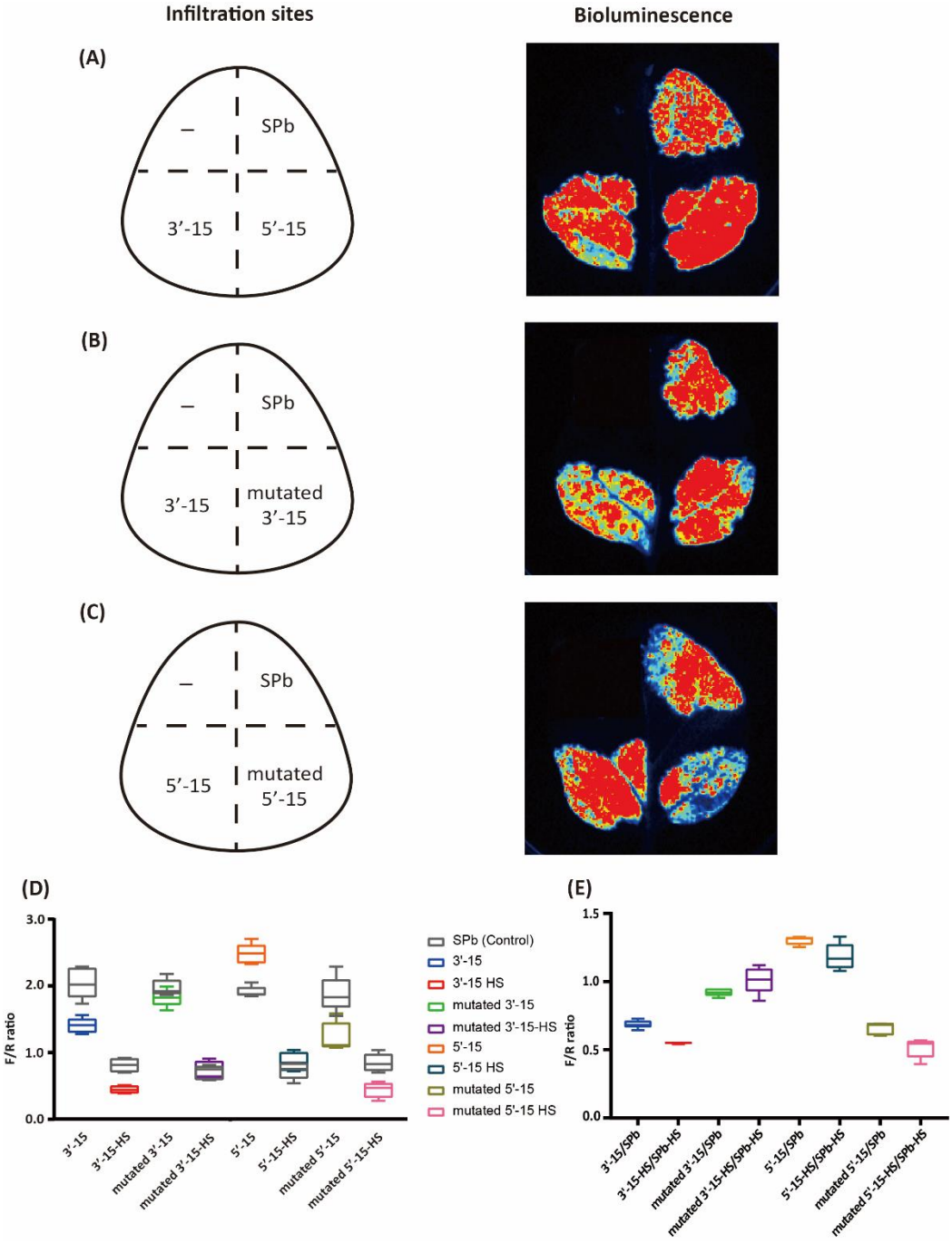


Figure 2.11: Mutational and position analysis of sensor fragment 15 on the expression of firefly Luciferase.

Fragment 15 and mutated fragment 15 were inserted separately into either the 3' or 5'UTR of *F-Luc,* and either the qualitative Luc image assay or DLR assay was performed as described above.

(A) The Luc image assay of 3'-15 and 5'-15. Compared with the positive control, 3'-15 showed a lower F-Luc level, and 5'-15 showed a higher F-Luc level. **(B)** The Luc image assay of 3'-15 and mutated 3'-15. 3'-15 showed a lower F-Luc level than the positive

control, and mutated 3'-15 showed a similar F-Luc level as the positive control. **(C)** The Luc image assay of 5'-15 and mutated 5'-15. Compared with the positive control, 5'-15 showed a higher F-Luc level, and mutated 5'-15 showed a lower F-Luc level. **(D)** The DLR assay of 3'-15, mutated 3'-15, 5'-15 and mutated 5'-15 after control and heat shock treatments. The F/R results of each group and its positive control were organised vertically. The colour box: the experiment group; the grey box: the corresponding positive control. **(E)** The normalized F/R ratio of 3'-15, mutated 3'-15, 5'-15 and mutated 5'-15 to their positive control. The F/R ratio of 5 replicates from each group was arranged from low to high and normalised with that of its positive control.

2.5. Discussion and conclusion

High temperatures are a major and increasingly common abiotic stress in plants, impacting plant growth, development, and crop yields. To improve the survival and yield of crops after higher temperatures, it is necessary to understand the mechanism of heat response, tolerance, and adaption in plants. At the cellular level, plants respond to high-temperature conditions by promptly adjusting membrane structure and function, global gene expression, protein abundance, and changes in primary and secondary metabolite (Sugio et al. 2009; Wang, L et al. 2020; Weston et al. 2011). Although studies of heat stress on plants have been undertaken, a complete understanding of the heat-response mechanisms remains elusive, especially at the molecular level. This chapter explored the potential roles of an RNA modification, 5mC, in responding to heat stress, and steps were undertaken to fill the gap of post-transcriptional regulation in heat stress response.

This experiment set 40°C for 30 min as the heat shock treatment condition. The ideal *laboratory* growth temperature range for *A. thaliana* is 21-23°C (Rivero et al. 2014). When temperatures increase from 28 °C to 37 °C, A. thaliana suffers deleterious effects of heat stress but can still grow. At temperatures of over 40 °C, an overall caseation of cell activity occurs, resulting in following cellular damage or even cell death (Li, B et al. 2018). Another study showed that a heat shock of 40°C for 30 min induces acquired thermotolerance of *A. thaliana* shoot to 45°C (Kaplan et al. 2004). Hence, we selected 40 °C for 30 min to ensure a robust thermal response and acquired acclimation at the cellular and molecular level but to minimize cell death and RNA degradation.

Our RBS-seq result showed that 1,385 transcripts were up-regulated, and 1,982 transcripts were down-regulated after heat shock treatment. Compared to RNA-

seq, it is relatively challenging to undertake robust transcriptome-wide RNA and gene analysis, like Gene Ontology Expression (GEO) analysis, of RBS-seq data as the bisulfite treatment causes RNA fragmentation and degradation (Fleming et al. 2019; Lv et al. 2021; Zhao, L-Y et al. 2020). Nevertheless, we assessed the expression of 56 previously characterized heat-related genes (Agarwal et al. 2001; Krishna & Gloor 2001; Lin et al. 2001; Nover et al. 2001; Scharf, Siddique & Vierling 2001; Swindell, Huebner & Weber 2007) and found the transcript abundance of 49 genes were up-regulated after heat stress treatment except *HSFA1D*, *HSFA6B*, *HSFB4*, *HSP70-9*, *HSP90-7*, *HSP100-5*, and *HSP100-7*. These 7 genes have no evident function to activate their transcription to respond to heat shock (Begum, Reuter & Schöffl 2013; Berka et al. 2022; Chong et al. 2015; Huang, Y-C et al. 2016; Nishimura et al. 2015; Ohama, Naohiko et al. 2016; Zhang, S et al. 2018). Our results demonstrated that our heat shock treatment successfully induced changes at the transcriptional level.

A total of 3,141 RNA 5mC sites were identified in the control. Compared with our laboratory's previous RBS-seq dataset that only detected 1,060 5mC sites in shoot and root transcriptomes (David et al. 2017), we identified 2,081 more 5mC sites due to a combination of greater, 3 times, sequencing depth, and using an improved mapping and 5mC calling method (meRanTK, Rieder et al. 2016). In addition to RBS-seq, other methods to detect RNA 5mC sites may be more sensitive. Cui et al. (2017) identified a total of 6,045 RNA 5mC sites by 5mC-RIP-seq. In comparison, another group used a machine learning-based 5mC predictor (PEA-5mC) to predict an overall 303,421 5mC site transcriptome-wide from the Araport 11 database (Song et al. 2018). Although these methods could detect more RNA 5mC sites, RBS-seq has the advantage of detection accuracy and precise positioning of 5mCs. Similar to other datasets (Cui et al. 2017; David et al. 2017), most of the RNA 5mCs we detected were located in the mRNA, and

only a small proportion was identified in other kinds of RNAs, such as snoRNA, snRNA and rRNAs. We then analysed the distribution of 5mCs on mRNA and showed that 5mC sites were mainly concentrated in the CDS (mean coverage of 0.005), especially near the stop and the start codons, followed by the 3'UTR and the 5'UTR. Despite some subtle differences in some specific regions, our result largely agreed with our previous RBS-seq data (David et al. 2017) and Cui's data (Cui et al. 2017), indicating that the 5mC distribution is similar to either RBS-seq or 5mC-RIP-seq.

The total number of 5mC sites increased to 5,728 after heat shock treatment, 2,587 more than under control conditions. Compared with the control, 1,070 5mC sites were lost, and 3,657 new sites were detected after heat shock treatment. The distribution of the RNA 5mCs after heat shock was like the control, that is, mRNA contained most 5mC sites, especially the CDS region. In addition, the overall level of 5mC in CDS was also significantly increased after thermal stress, especially near the stop codon and start codon. The increased mRNA 5mC amount and abundance after heat shock treatment suggests a heat-response mechanism that is currently unclear.

To explore the potential role of 5mC on transcript abundance, we analysed the methylation level changes of 5mC sites on transcripts with different abundance in the control and after heat shock treatment. We did not find a positive correlation between RNA 5mC levels and transcript abundance, which is similar to observations by David et al. (2017) and Cui et al. (2017).

Interestingly, we found a significant number of 5mC sites present as clusters in the control group and the heat-treated samples. Although clustered 5mCs were recently reported in mammals (Fang et al. 2020; Huang, T et al. 2019; Zhang, Q et al. 2020), the existence of the clustered 5mC sites has long been controversial

due to possible artefacts. During bs-RNA-seq, bisulfite conversion of RNA cytosine bases can be inhibited by secondary or tertiary structures leading to the protection of local cytosines and hence unmodified cytosine sites called false positive 5mC sites (Amort et al. 2017; Haag, S. et al. 2015; Huang, T et al. 2019; Shapiro et al. 1973; Trixl, Lukas & Lusser 2018). To reduce the disruption of RNA structure on BS conversion, a denaturation step was performed before the BS treatment to stretch the folded regions of RNA. Any 5mC sites that might exist in regions with RNA secondary structures were excluded via an RNA secondary structure prediction analysis. While we still observed a large number of 5mC clusters, especially in heat-treated samples. It cannot be excluded that emerging 5mC cluster sites after heat shock treatment have some special thermal response mechanisms rather than incomplete conversion of cytosine sites. Until now, no research reported the role of clustered 5mCs in plants and environmental stress response.

The potential function of single or clustered RNA 5mC sites in regulating RNA transcripts was explored after both controls and heat shock treatments. 17 candidate fragments from 10 mRNAs were selected, and their effects on the expression of F-Luc bioluminescence were tested in *N. benthamiana*. Of these 17 fragments, 13 contained cytosines with up-regulated 5mC levels after heat shock treatment, and 4 contained cytosines with down-regulated 5mC levels after heat shock treatment in a single or cluster. The gene *At1g05340* encodes the Cysteine-Rich Transmembrane Module 1 protein, which is important in protecting plants against environmental stresses, especially heat shock (Joshi, Singh & Friedman 2020; Vandenabeele et al. 2004; Xu, Y et al. 2018). Joshi, Singh and Friedman (2020) found that a heat shock treatment of 50°C for 30 min significantly induces the transcription of *At1g05340*. While we did not observe an obvious increase of *At1g05340* transcript abundance after our heat

shock treatment (40°C for 30 min). Our heat shock temperature might be too low to increase the transcription of At1g05340 significantly. However, we found that the overall 5mC abundance on the *At1g05340* transcript was largely increased after heat shock treatment, and 36 heat-unique 5mC sites appeared as clusters. To investigate the role of these hypermethylated cytosines, we synthesized four fragments of different regions on the *At1g05340* transcript *in vitro* and tested the effect of their heat-sensitive hypermethylated cytosines on F-Luc bioluminescence. The insertion of fragment 1 derived from the 5'UTR of *At1g05340* carrying 7 5mC sites in the 5' UTR of *F-Luc* significantly enhanced the expression of F-Luc, 2.5 times more than the control under both control and heat shock treatments. When we inserted fragment 1 into the 3'UTR of *F-Luc*, the expression level of F-Luc after control conditions decreased to only 0.5 times that of the control, while after thermal conditions, the expression level of F-Luc remained unchanged compared to the control.

Similarly, inserting a segment of the exon 2 with 18 up-regulated 5mC sites after heat shock treatment (Fragment 3) on the 3'UTR of *F-Luc* leaded to a 50% reduction of F-Luc bioluminescent under control conditions and after heat shock treatment. However, when we mutated hypermethylated cytosines of fragments 1 and 3, the level of F-Luc bioluminescent recovered to a certain extent, regardless of the insertion positions and treatment conditions. These findings suggest that the 5mC might involve in the post-transcriptional regulation of *At1g05340*. The 5mCs at the 5'UTR might be able to enhance and maintain gene expression. The 5mCs at the 3'UTR might have a role in reducing gene expression and maintaining the transcript abundance after heat shock treatment. The functions 5mCs on its CDS region seem more complicated than that on other regions, and the 5mCs on exon 2 might be required to retain its low transcription. Although no significant regulation of 5mC sites on exons 1

and 3 (Fragment 2 and Fragment 4) was observed on the expression of F-Luc, I cannot rule out their potential roles in other aspects of RNA, such as RNA structural stabilization or RNA transportation.

The gene *At3g22260* encodes a liquid transfer protein, namely LTP5 (Glycosylphosphatidylinositol-Anchored Lipid Protein Transfer 5), which was shown for resistance against a diverse range of pathogens and non-host mildews (Fahlberg et al. 2019; Gao et al. 2021). *At3g22260* is not a heat-responsive gene, and its expression after heat shock treatment was low before and after heat shock. We identified four cytosines in exon 1, and the average methylation levels increased by 0.4 after heat shock. The insertion of the exon 1 of *At3g22260* (Fragment 13) resulted in a 50% reduction of F-Luc bioluminescence, either in the 3' or 5'UTR, with or without 5mC sites, under control or heat shock stress. The low but consistent F-Luc bioluminescence suggested that the expression of *At3g22260* might be inhibited by other components of exon 1 instead of the 5mC sites. The exact role of 5mCs on exon 1 of *At3g22260* remains unclear.

We also investigated the role of a single 5mC site on F-Luc bioluminescence. One heat-unique 5mC site in the exon 2 of *At1g60750* (NAD-P linked oxidoreductase superfamily protein) was presented hypermethylated (0.91) after a heat shock treatment and seemed to have multiple roles on gene expression. Adding the fragment with this single 5mC (fragment 6) to the 3' UTR of *F-Luc* enhanced the bioluminescence by 1.4 times the control after control conditions and about 1.2 times the control after heat shock treatment. In comparison, the insertion of fragment 6 in the 5'UTR of *F-Luc* reduced bioluminescence to about 0.7 times the control under control conditions and 0.8 times after heat shock treatment. Another 5mC on the exon 2 of *At4g20260* (plasma-membrane associated cation-binding protein 1, ATPCAP1) with low

methylation level (0.0 under control conditions, 0.09 after heat shock treatment) did not significantly affect F-Luc bioluminescence. Interestingly, the mutation of the cytosine caused a 40% reduction of F-Luc bioluminescence after heat shock treatment. It is possible that this site had a function to prevent transcript degradation after heat shock. The single 5mC site seems to have a relatively limited effect on gene expression compared with clustered sites. However, we could not test more 5mC sites on gene expression due to time constraints. The exact roles of single 5mC on transcriptional regulation still need further studies.

We observed that mRNAs with increased abundance were more likely to carry cytosines with up-regulated methylation levels after heat shock treatment. We identified two up-regulated mRNAs, *At5g54940* (Translation Initiation Factor SUI1) and *At1g78380* (Glutathione S-Transferase TAU 19), in which methylation levels of all 5mC sites reduced to 0 after heat shock treatment. This may suggest that the mRNAs having 5mC were rapidly degraded upon heat shock, for example, by a single-stranded exonuclease. Only the fragment of *At1g78380* exon 1 containing two 5mC sites (fragment 15) increased F-Luc bioluminescence. We mimicked the position of these two sites by inserting fragment 15 in the 5'UTR of *F-Luc* and observed an increase of F-Luc both under control conditions and after heat shock treatment. While insertion of fragment 15 without the 5mC sites caused a significant reduction of F-Luc bioluminescence both under control conditions and after heat shock treatment, suggesting that these two sites might have a role in maintaining the gene expression.

Taken together, our study is the first to reveal the potential role of RNA 5mC in the thermal response of plants. Heat shock-induced transcriptome-wide alteration of 5mC sites and abundance, especially on mRNAs. In our transient assay, only a small proportion of heat-response 5mCs had a clear role in regulating transcript abundance. Especially two hypomethylation regions of

AT1G05340, 5’UTR and exon 2, maintained bioluminescence after heat shock. Single 5mC sites seem to have less effects than clustered 5mCs on gene expression. However, the exact regulation pathways of these 5mCs on gene expression and the roles of other heat-response RNA 5mCs are still unclear. Further studies are encouraged to focus on them.

3. Chapter 3:

Towards identification of RNA 5mC demethylases in *A. thaliana*

3.1. Abstract

RNA methylation and demethylation are opposite regulatory pathways that dynamically regulate methylation levels. In plants, RNA cytosine is methylated to 5mC by methyltransferases from the RCMT family, TRM4, NOP2, RCMT9 and TRDMT1. Our understanding of 5mC demethylation lags behind other modifications like *N6*-mA, and no RNA 5mC demethylase has been identified in plants. While a putative oxidative product of RNA 5mC, 5-hydroxymethylation (5hmC) was previously identified in *A. thaliana*, no oxidative enzymes were identified. In other organisms, oxidative demethylation is generally catalysed by AlkB-like and AlkB homolog (ALKBH) proteins.

In this chapter, I predicted 14 AlkB-like candidates in *A. thaliana* with high protein similarity to AlkB and ALKBH1-8. I then successfully screened T-DNA-inserted mutant alleles for 11 of the 14 candidates. As a first step, the demethylation activities of these 11 AlkB-like proteins on RNA 5mC were detected by measuring the RNA 5mC abundance of *alkb-like* mutants at *MAG5* C_{3349} under control conditions and heat shock treatment by BS-amplicon-seq. Although no apparent alteration of 5mC level was identified in any single *alkb-like* mutant under control conditions, *alkbh8 (at1g31600)* and *at5g01780* mutants notably increased 5mC level at C_{3349} of *MAG5* after a heat shock. Due to limited time, the demethylation activity of ALKBH8 and AT5G01780 on 5mC was not further investigated. However, my results provided an initial hint that ALKBH8 and AT5G01780 might directly or indirectly affect 5mC, for example, alter RNA turnover or other aspects of RNA metabolism, and is worth further investigation.

3.2. Introduction

5-methylcytosine (5mC) is an abundant RNA modification that extensively occurs on various RNAs, including mRNAs, tRNAs, rRNAs and other ncRNAs. Until now, over 10,000 RNA 5mC sites have been found in humans and over 1,000 in plants (Cui et al. 2017; David et al. 2017; Squires, Jeffrey E et al. 2012). The methylation process of RNA 5mC is mainly catalysed by two classes of methyltransferases, RCMT subfamilies (RCMT1-RMCT9) and RNA aspartic acid methyltransferase (TRDMT1), which transfer a methyl group from AdoMet to the fifth carbon of cytosine (King & Redman 2002; Squires, Jeffrey E et al. 2012; Walbott et al. 2007). While the understanding of RNA 5mC demethylase and demethylation pathway is still mainly in the infant stage.

In mammals, DNA and RNA 5mC can be further converted to oxidative derivatives by TET family proteins and then probably reversed to unmodified cytosine (He, Y-F et al. 2011; Ito et al. 2011; Kohli & Zhang 2013). TETs were first known as DNA 5mC dioxygenases that can repeatedly oxidize 5mC first to 5-hydroxymethylcytidine (5hmC), then to 5-formylcytosine (5fC) and finally to 5-carboxylcytosine (5caC), so-called oxidative demethylation pathway (Ito et al. 2011). Both 5fC and 5caC can be excised by TDG, and the base pair gap is subsequently repaired via the BER pathway (Kohli & Zhang 2013; Shen et al. 2014). Subsequently, TET1 was demonstrated as the first enzyme to oxidise RNA 5mC to generate 5hmC and 5fC in mammals (Fu, L et al. 2014). While TET1's catalytic activity on an RNA substrate was much lower than on a DNA substrate. Furthermore, the deletion of all three TET proteins causes a decrease of RNA 5hmC in mouse embryonic stem (ES) cells instead of the lack of 5hmC (Fu, L et al. 2014). These findings suggest that the TET1 might not be the only demethylase of RNA 5mC, and other unknown enzymes are probably involved in the

demethylation process. Up to date, no other RNA 5mC demethylase has been found in any organism.

The exploration of RNA 5mC metabolism has not stalled. In 2015, Balasubramanian's group identified RNA 5hmC, an oxidation oxidative of 5mC, present in RNA isolated from mammals, a plant (*A. thaliana*) and bacteria (Huber et al. 2015). This finding suggests a possibility that RNA 5mC can be further oxidised to 5hmC and even demethylated to unmodified cytosine in different organisms. As no TET-like proteins were identified in the *A. thaliana* genome (Searle *et al.,* unpublished data), how RNA 5hmC is produced is unclear and whether it is the oxidative product of 5mC is also currently unclear. It is possible that some unknown dioxygenases in *A. thaliana* oxidatively convert RNA 5mC to 5hmC.

Cellular oxidative demethylation processes are generally catalysed by *Escherichia coli* AlkB, which is an iron (II)/αKG-dependent dioxygenase and AlkB-like proteins, such as AlkB human homologues ALKBH1-8, FTO and TET proteins (Chen et al. 2014; Shen et al. 2014; Ye et al. 2014). AlkB and AlkB-like dioxygenases share conserved core residues and structure in their catalytic domains, in which the iron (II) is surrounded by a water molecule αKG and an HXD…H motif (Shen et al. 2014; Stephanie & Timothy 2013; Ye et al. 2014). The HXD…H motif is present in all AlkB-like dioxygenases, but the location and distance between the HXD and H residues vary among different dioxygenases (Kurowski et al. 2003; Yu et al. 2006). The main role of the HXD…H motif is for αKG binding.

If an RNA 5mC oxidative demethylase is present in *A. thaliana*, it is reasonable to speculate it may be an AlkB or AlkB-like protein and contains the HXD…H motif. Based on this speculation, in this chapter, I identified AlkB-like genes in

the *A. thaliana* genome, characterized loss-of-function mutants, and started to test their effect on RNA 5mC abundance.

3.3. Materials and Methods

3.3.1. Plant materials

The *A. thaliana* ecotype used throughout this project was Columbia-0 (Col0). T-DNA inserted alleles described were *atalkb-1*: FLAG_143G11; *AtAlkB-2*: SALK_062071; *atalkb-b:* SK15155; *AtAlkB-B-2*: SALK_052340; *atalkb-c-1*: SAIL_193_C03; *atalkb-c-2*: SALK_021477C; *AtAlkB-C-3*: SALK_035671; *atalkbh1*: SALK_007964; *AtAlkBH2-1*: SAIL_1239_H02; *AtAlkBH2-2*: WISCDSLOXHS085_11E; *AtAlkBH2-3*: WISCDSLOXHS038_10; *atalkbh2b*: SAIL_68_G03; *AtALKBH2B-2*: SAIL_223_C05; *AtALKBH5-1*: GABI_423C08; *AtALKBH5-2*: FLAG_156_D06, *atalkbh6*: SALK_138864; *AtALKBH6-2*: FALG_564H09; *atalkbh8-1*: SALK_094502C; *atalkbh8-2*: SALK_083838C; *atalkbh8b*: SALK_024872C; *AtALKBH8B-2*: SAIL_756_A08; *attrm9*: SALK_135308C; *AtTRM9-2*: SALK_003181; *atalkbh9a*: SALK_204823; *AtALKBH9A-2*: SAIL_402_E12; *AtALKBH9B*-1: SALK_111811; *AtALKBH9B*-2: WISCDSLOX501_D02; *atalkbh9c*: SALK_021775; *AtALKBH9C-2*: WiscDsLoxHs118_10A.

All *A. thaliana* seeds were desiccated after harvesting and stored at 4°C in sealed Eppendorf tubes for more than two weeks to break dormancy and improve germination.

3.3.2. Plant growth conditions

3.3.2.1. Soil-based growth conditions

A. thaliana seeds were planted on a soil mixture (2 portions of Debco Seed Raising Soil and 1 portion of Debco® Seed & Cutting Mix), and the pots were placed in a Phoenix Biosystem temperature-controlled room at 21°C with metal

halide lights (photosynthetic active radiation: 100 µmol/m^2/s). The plants were grown after long day photoperiod conditions (16 hr light, 8 hr darkness).

3.3.2.2. Media-based growth conditions:

A. thaliana seeds were sterilized by treating the seeds with chlorine gas for 16 hrs. The seeds were placed in an enclosed glass chamber with a beaker containing 2 mL of concentrated HCl (25%) and 100 mL of NaOCl (5%). Sterilized seeds were plated on ½ MS media plates supplemented with 1% sucrose, stratified at 4°C in the dark for 3 days, and then grown at 21°C after long day conditions (16 hr light and 8 hr darkness) for 10 days in a Phoenix Biosystem controlled environment room with metal halide lights (100 µmol/m^2/s).

3.3.2.3. Heat shock treatment

To induce heat shock, 10-day-old *A. thaliana* seedlings on ½ MS media plates were placed in a 40°C water bath for 30 min after the same light conditions. The water temperature was measured by using a thermometer. After heat shock treatment, 10 seedlings (about 100 mg) were harvested immediately into an RNase-free Eppendorf tube by snap frozen in liquid nitrogen and stored at -80°C ultra-freezer. For the control group, 10-day-old seedlings without heat treatment were harvested directly into an RNase-free Eppendorf tube, snap-frozen in liquid nitrogen and stored at -80°C.

3.3.3. Identification of T-DNA insertion mutants

3.3.3.1. Identification of homozygous genotype

Homozygous T-DNA lines were genotyped by PCR amplification using the

primers shown in Supplementary Table 3.1. Genomic DNA was isolated from true leaves using the Quick Miniprep Plant DNA Isolation protocol (Weigel & Glazebrook 2009), and PCR was amplified using a genomic left border primer (LP) and a genomic right border primer (RP), and a T-DNA left border primer (LB), according to the T-DNA Primer Design Tool (http://signal.salk.edu/tdnaprimers.2.html).

3.3.3.2. Identification of the expression of T-DNA alleles

The expression of T-DNA alleles was characterized by rt-PCR, and the primers are shown in Supplementary Table 3.1. RNA was first isolated from 100mg liquid nitrogen frozen tissue using a TissueLyser II (Retsch) and subsequently purified using the TRIzol reagent (Invitrogen ™), following the manufacturer's instructions. Purified RNA was then transcribed to cDNA using Superscript II Reverse Transcriptase (Invitrogen) and a poly dT primer (Invitrogen), following the manufacturer's instructions. The expression of the region downstream or upstream of the T-DNA insertion site in the homozygous mutant was specifically amplified using the RT-F and RT-R primers, as shown in Supplementary Table 3.1. The knockout T-DNA mutants have a null expression of the amplified region.

3.3.4. RNA extraction, quantification and quality assessment

About 100 mg of leaf tissues were snap-frozen in liquid nitrogen and ground with two 3mm steel beads using a TissueLyser II (Retsch) at a shaking frequency of 25Hz for 1 min. Total RNA was purified, and DNA was digested using a SpectrumTM Plant Total RNA Kit, following the manufacturer's instructions. All RNA samples were stored in a -80°C ultra-freezer to prevent degradation.

RNA concentrations were measured using a Qubit™ RNA HS Assay Kit on a Qubit® 2.0 Fluorometer, following the manufacturer's instructions. RNA quality was assessed after RNase-free agarose gel electrophoresis. 500 ng of total RNA was mixed with equal volumes of 2x RNA loading buffer, incubated at 65°C for 5 min to denature the RNA and remove secondary RNA structure, and the denatured samples were loaded on a 2% agarose gel. Samples were separated at 100 V for 90 min and stained with Red Safe.

3.3.5. cDNA synthesis of RNA samples

500 ng of total RNA was reverse-transcribed to first-strand cDNA using the SuperScript™ III Reverse Transcriptase (ThermoFisher Scientific) using OligodT, following the manufacturer's instructions.

3.3.6. RNA Bisulfite-Amplicon Sequencing (RBS-amplicon seq)

3.3.6.1. rRNA depletion

Total RNA samples were subjected to ribosome RNA depletion by using the Ribo-Zero rRNA Removal Kit (Plant leaf, Illumina). Successful rRNA depletion was verified using a Bioanalyzer 2001 (Agilent Technologies).

3.3.6.2. RNA sodium bisulfite treatment

2.5 µg of rRNA-depleted RNA samples were denatured at 75°C for 5 min and then reacted with pre-heated sodium bisulfite solution (40% sodium metabisulfite and 0.6 mM hydroquinone) in the dark for 4 hr. Sodium bisulfite treated RNA samples were passed through BioRad™ Micro Bio-Spin Columns

(Bio-Gel P-6 in Tris Buffer) twice to remove the remaining salts, following the manufacturer's instructions. An equal volume of Tris buffer (pH 9.0) was added to the desalted RNA solutions and incubated at 75°C for 1 hr to allow the generation of uridine. 1/10 volume of 3 M NaAc, 3 volumes of 100% ethanol and 2 µL of 10 mg/ml glycogen were added to the RNA samples and incubated at -20°C overnight to precipitate the RNA. On the following day, RNA samples were precipitated by centrifugation in a benchtop centrifuge at 13,000 rpm, 4°C for 25 min and then washed with 70% ethanol twice. Cleaned RNA samples were resuspended in 30-40 µL of ultra-pure water and stored at -80°C.

3.3.6.3. Library construction for RBS-amplicon seq

The BS-treated RNA is fragmental and sheared, making it relatively difficult to synthesise cDNAs using Oligo-dT. Hence, targeted cDNA sequences were efficiently synthesised using gene-specific primers (GSPs) at about 100 nt 3' of the potential 5mC sites.

RBS-amplicon library is constructed according to Li, J et al. (2021). Typically, 500 ng of BS-treated RNA samples were synthesised to first-strand cDNAs using SuperScript™ III Reverse Transcriptase and GSPs mix, following the manufacturer's instructions. BS-amplicons were PCR amplified using the KAPA HiFi™ high fidelity polymerase and BS-altered reverse transcript (rt) primers, following the manufacturer's instructions (Kapa Biosystems Inc, RT-primers are shown in Supplementary Table 3.2. Contaminants (dNTPs, salts, primers, primer dimers) were then removed from BS-amplicons by twice AMPure XP beads clean up. Purified BS-amplicons from the same biological replicate were bulked and indexed in the second round PCR, followed by another twice AMPure XP beads clean up. Three replicates for each mutant and condition were performed for Illumina sequencing on a MiSeq platform (2x 150 nt paired-end) at the ACRF,

Adelaide.

3.3.7. Bioinformatic analysis of BS-amplicon sequencing

3.3.7.1. Quality control

The raw Illumina sequence was trimmed by TrimGalore! (version 0.4.2) to remove adaptor contamination and low-quality reads using a Phred score threshold of 20 (Krueger 2012).

3.3.7.2. Sequencing read alignment

Trimmed reads were then aligned to the index C-T converted references (Supplementary Table 3.3) by performing the meRanCall of the meRanTK (version 1.2.1, Rieder et al. 2016), following the manual.

3.3.7.3. 5mC level calculation

The methylation level of each candidate 5mC site was calculated by: C/(C+T), which infers 5mC/(5mC+C).

3.4. Results:

3.4.1. Bioinformatic prediction of candidate AlkB-like proteins in A. *thaliana*

The main purpose of this chapter was to investigate the enzymes that catalyse the oxidation of RNA 5mC to 5hmC. AlkB, ALKBH-like proteins ALKHB1-8, FTO and TET1-3 are known to catalyse the oxidative demethylation of different RNA substrates (Kohli & Zhang 2013; Kurowski et al. 2003; Marcinkowski et al. 2020; Toh et al. 2020; Yu et al. 2006). For example, FTO and ALKBH5 oxidatively demethylate RNA *N6*-mA to *N6*-hydroxymethyladenosine (*N6*-hmA) and *N6*-formyladenosine (*N6*-fA) and subsequently reverse to adenosine (Aik et al. 2014; Jia et al. 2011; Toh et al. 2020). Little is known about AlkB-like proteins in plants. To date, only a few AtAlkB proteins have been identified and characterized in plants, including AtALKBH6, AtALKBH8, AtALKBH9 and AtALKBH10 (Duan et al. 2017; Huong, Ngoc & Kang 2020; Leihne et al. 2011; Masiello & Biggiogera 2017), and none of them target RNA 5mC. In this chapter, I explored the potential demethylase of RNA 5mC. I first identified potential AtAlkB candidates by bioinformatic analysis. Then, I screened and isolated T-DNA mutant lines for candidate *AtAlkB* genes. After that, the demethylation activity of AlkB-like candidates towards RNA 5mC was determined by measuring the RNA 5mC/5-hmC abundance of their T-DNA mutant alleles after control and heat shock treatments by BS-amplicon-seq. Due to time constraints, assays to measure 5mC and 5-hmC were not undertaken.

The first step was to predict AtAlkB candidates in *A. thaliana*. The protein sequences of *E. coli* AlkB, *Homo sapiens* ALKBH1-8, *H. sapiens* FTO and *H. sapiens* TETs were aligned to the Araport11 protein database (TAIR 11) by using

BLASTP. At least 20% identity within the functional domain was required as a cut-off for candidates. In total, 14 AtAlkB candidates were identified with protein similarity to *E. coli* AlkB and *H. sapiens* ALKBH1-8, and their gene IDs are listed in Table 3.1. Unfortunately, no clear TET-like or FTO-like proteins were identified. To gain further insight into the relationship between the *A. thaliana* candidates and the homologues, I performed a phylogenetic analysis with the *Arabidopsis* candidates and the AlkB-like proteins using Clustal Omega (Figure 3.1). As expected, the *A. thaliana* candidates with high similarity to the AlkB homologue are within the same clade. Using the protein similarity and phylogenetic relationship, I determined AT3G14160, AT3G14140 and AT5G01780 as AtAlkB, AtAlkB-B and AtAlkB-C, respectively; AT1G11780 as AtALKBH1; AT2G22270 as AtALKBH2B; AT2G48080 as AtALKBH5; AT4G02485 as AtALKBH8B (Figure 3.2). Other candidate proteins that were described in previous publications were named according to those publications, and these included AT2G22260 (AtALKBH2, Meza et al. 2012), AT4G20305 (AtALKBH6, Huong, Ngoc & Kang 2020), AT1G31600 and AT1G36310 , and AT1G48980, AT2G17910 and AT4G36090 (AtALKBH9A, AtALKBH9B and AtALKBH9C, Leihne et al. 2011; Marcinkowski et al. 2020).

Table 3.1：AlkB-like genes and putative *A. thaliana* orthologues

AlkB family dioxygenase	Alkylated substrates	Candidates with similarity in *A. thaliana* (amino acid identity %)
E. coli ALKB	3mC or *N1*-mA of DNA and RNA	AT5G01780 (30%); AT3G14160 (28%); AT1G11780 (28%); AT3G14140 (25%);
H. sapiens ALKBH1	*N1*-mA of tRNA; 3mC of mRNA *N1*-mG of RNA; *N6*-mA of ssDNA; methylated lysine of histone H2A	AT1G11780 (32%); AT5G01780 (39%); AT3G14160 (34%); AT3G14140 (28%)
H. sapiens ALKBH2	3mC or *N1*-mA of dsDNA and RNA	AT2G22260 (37%); AT2G22270 (25%)
H. sapiens ALKBH3	3mC or *N1*-mA of ssDNA and RNA; *N1*-mG of RNA;	AT2G22260 (34%); AT4G02485 (22%)
H. sapiens ALKBH4	monomethylated Lys-84 (K84me1); *N6*-mA of DNA	AT1G31600 (28%)
H. sapiens ALKBH5	*N1*-mA of RNA and ssDNA	AT4G36090 (31%); AT2G17970 (30%); AT1G48980 (29%); AT2G48080 (28%); AT4G02485 (28%)
H. sapiens ALKBH6	Unknown	AT4G20350 (47%); AT1G31600 (24%); AT4G02485 (26%)
H. sapiens ALKBH7	Leu-110	AT4G02485 (37%)
H. sapiens ALKBH8	C5mU of tRNA	AT1G36310 (35%); AT1G31600 (38%); AT4G02485 (40%)
H. sapiens FTO	*N6*-mA of RNA	-
H. sapiens TET	5mC of RNA and DNA	-

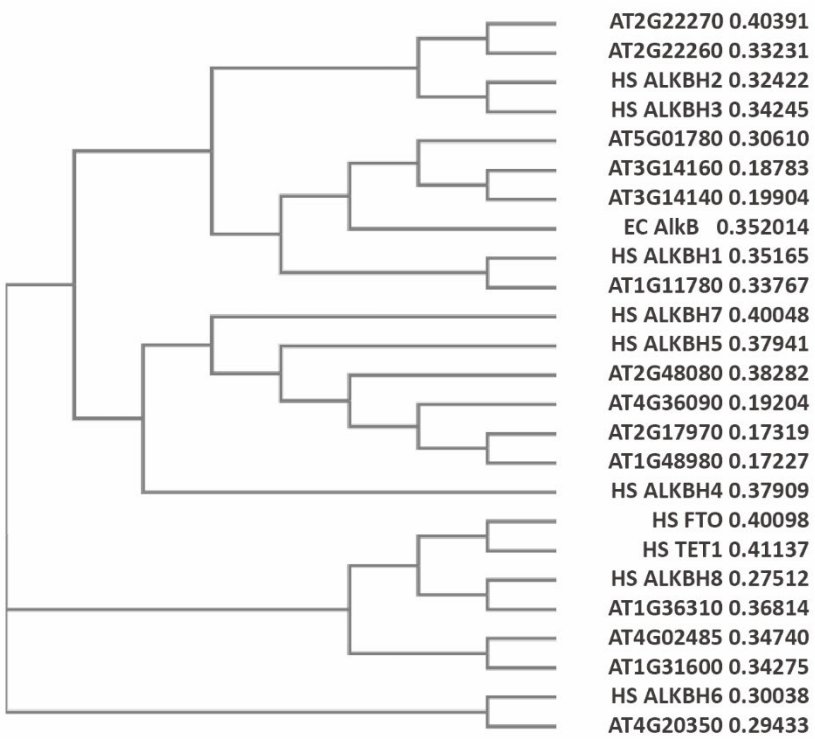


Figure 3.1: Phylogenetic analysis of AtAlkB-like proteins.

The phylogenetic tree was generated by aligning the protein sequences of *A. thaliana* AlkB-like proteins with BA (*B. abortus*) AlkB, HS (*H. sapiens*) ALKBH1-8, HS FTO and HS TET1 using Clustal Omega (ClustalW and ClustalX version 2, Goujon et al. 2010; Larkin et al. 2007). The phylogenetic tree is a Neighbour-joining tree without distance corrections. The distance values shown are the number of substitutions as a proportion of the length of the alignment (excluding gaps). The distance between two sequences represents the identity rate of the two sequences (1-No. identical residues/NO. aligned residues).

Next, I then analysed the functional motif of the AtAlkB candidates. ALKB-like proteins have a well-characterized HXD…H motif that is essential for binding the cofactor α-ketoglutarate (Kurowski et al. 2003; Yu et al. 2006). Most AtAlkB candidates contained the HXD…H motif, except AT3G14140, AT1G36310 and AT2G22270 (Table 3.2). This suggests that all HXD…H containing proteins can bind α-ketoglutarate and potentially have dioxygenase activity.

Table 3.2: Identification of AtAlkB-like genes in *A. thaliana*

Gene ID	Name	HXD…H motif (amino acid position)
AT3G14160	AtAlkB[1]	H (363), D (365), H (423)
AT3G14140	AtAlkB-B[1]	-
AT5G01780	AtAlkB-C[1]	H (350), D (352), H (410)
AT1G11780	AtALKBH1[1]	H (243), D (245), H (299)
AT2G22260	AtALKBH2[2]	H (213), D (215), H (293)
AT2G22270	AtALKBH2B[1]	H (133), D (135), H (193)
AT2G48080	AtALKBH5[1]	H (283), D (285), H (238)
AT4G20350	AtALKBH6[2]	H (97), D (99), H (131)
AT1G31600	AtALKBH8[2]	H (226), D (228), H (297)
AT4G02485	AtALKBH8B[1]	H (137), D (139), H (199)
AT1G36310	AtTRM9[2]	-
AT1G48980	AtALKBH9A[2]	H (220), D (222), H (279)
AT2G17910	AtALKBH9B[2]	H (239), D (241), H (285)
AT4G36090	AtALKBH9C[2]	H (339), D (341), H (399)

* [1] predicted AtAlkB-like candidates; [2] published AtAlkB-like candidates: AtALKBH2 (Meza et al. 2012); AtALKBH8 (Leihne et al. 2011); AtALKBH9C, AtALKBH9B & AtALKBH9A (Martínez-Pérez et al. 2017), AtALKBH6 (Huong, Ngoc & Kang 2020), AtTRM9 (Leihne et al. 2011),

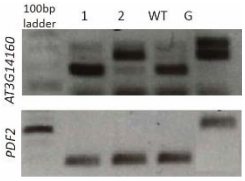
Taking together, I identified 14 AtAlkB-like proteins in *A. thaliana*, and 11 of them have the HXD…H motif and potential dioxygenase activity. The following sections will take steps towards testing these candidates' capacity to demethylate RNA 5mC.

3.4.2. Identification and characterization of AtAlkB T-DNA insertion mutants

A first step towards understanding the function of *A. thaliana* AlkB candidates was identifying T-DNA insertion mutants of these genes. I searched for T-DNA lines for each gene, obtained seeds and identified homozygous T-DNA mutants via a DNA PCR genotyping assay. To further characterize the insertion mutants, the mRNA abundance of each gene was measured by rt-PCR (primers used are shown in Supplementary Table 3.1), and *PDF2* (*Prefoldin 2, AT3G22480*) was also amplified as an internal control (Figure 3.2). We found four different outcomes for each T-DNA mutant: null mRNA expression (a so-called knock-out), decreased mRNA expression (a so-called knock-down), unchanged mRNA expression and increased expression (Table 3.3). The T-DNA lines that lead to null expression were *atalkb* (FLAG_143G11), *atalkb-b* (SK15155), *atalkb-c-1* (SALK_021477C), *atalkb-c-2* (SAIL_193_C03), *atalkbh1* (SALK_007964), *atalkbh2b* (SAIL_68_G03), *atalkbh6* (SALK_138864), *atalkbh8-1* (SALK_083838C; *atalkbh8-2* (SALK_094520C), *atalkbh8b* (SALK_024872), *attrm9* (SALK_135308C), *atalkbh9a-1* (SALK_204823), *atalkbh9c* (SALK_021775). While no knockout T-DNA line was found in *AtAlkBH2*, *AtAlkBH5* or *AtAlkBH9B*. Our finding is consistent with O'Malley, Barragan and Ecker (2015), that knock-out T-DNA

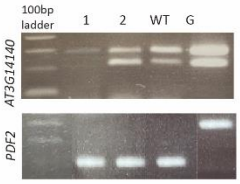
alleles were mainly identified within exons.

(A): *AtAlkB*



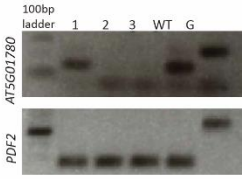
1: FLAG_143_G11
2: SALK_062071
WT: Col2n RNA
G: Col2n DNA

(B): *AtAlkB-B*



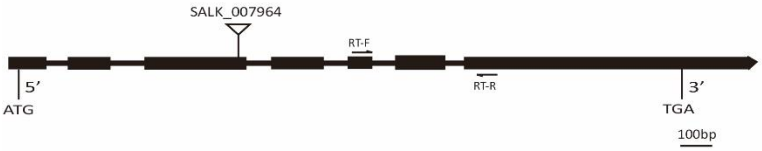
1:SK15155
2: SALK_052340
WT: Col2n RNA
G: Col2n DNA

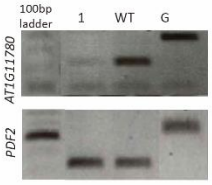
(C): *AtAlkB-C*



1: SALK_035671
2: SALK_021477C
3: SAIL_193_C03
WT: Col2n RNA
G: Col2n DNA

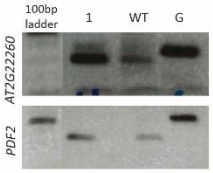
(D): *AtALKBH1*





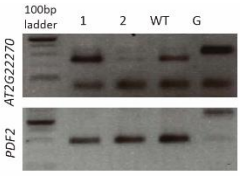
1: SALK_007964
WT: Col2n RNA
G: Col2n DNA

(E): *AtALKBH2*



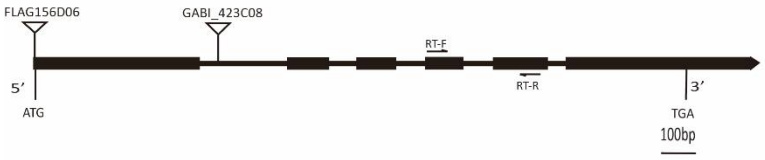
1: WiscDsLoxHs085_11E
WT: Col2n RNA
G: Col2n DNA

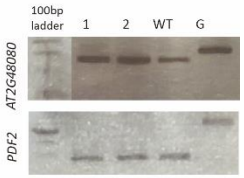
(F): *AtALKBH2B*



1: SAIL_223_C05
2: SAIL_68_G03
WT: Col2n RNA
G: Col2n DNA

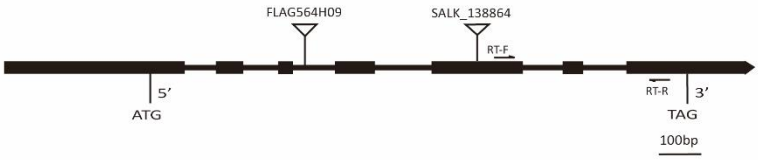
(G): *AtALKBH5*

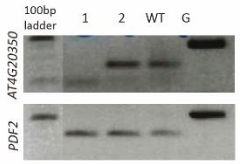




1: FLAG156D06
2: GABI_423C08
WT: Col2n RNA
G: Col2n DNA

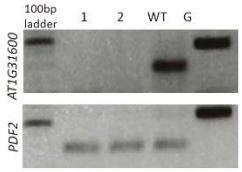
(H): *AtALKBH6*





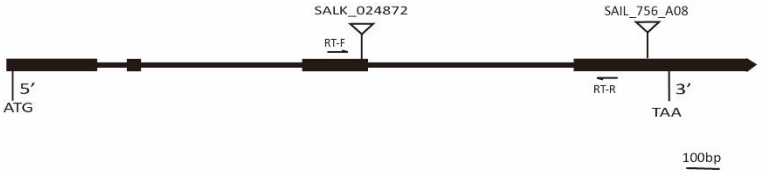
1: SALK_138864
2: FLAG564H09
WT: Col2n RNA
G: Col2n DNA

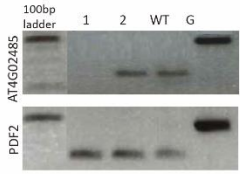
(I): *AtALKBH8*



1: SALK_083838
2: SALK_094520
WT: Col2n RNA
G: Col2n DNA

(J):*AtALKBH8B*





1:SALK_024872
2: SALK_756_A08
WT: Col2n RNA
G: Col2n DNA

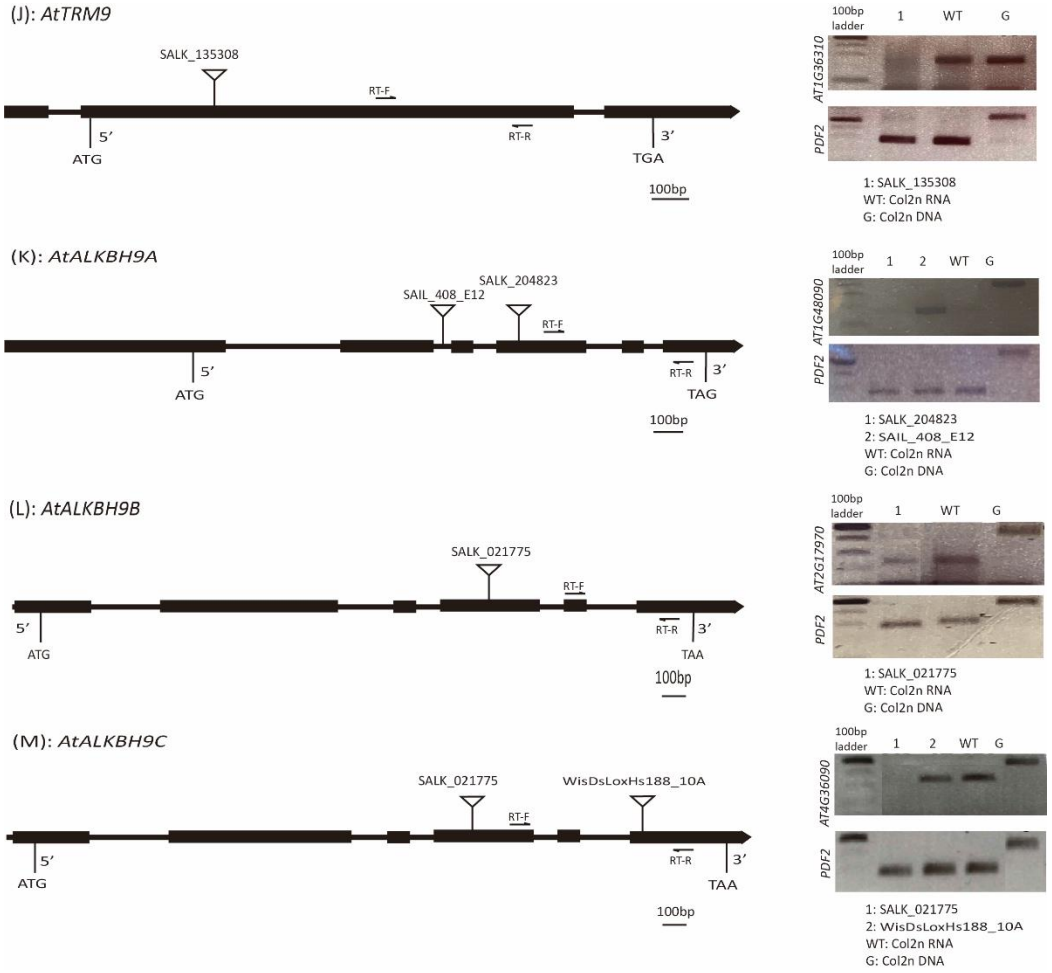


Figure 3.2: Identification of T-DNA insertion mutant alleles in *AtAlkB-like* genes.

Mutants of *AtAlkB-like* genes were identified by identifying homozygous T-DNA insertions (A-N). Left, schematic representation of the *AtAlkB-like* genes whereby black boxes represent exons and connecting thin lines are introns. Triangles represent the T-DNA insertion. rt-F and rt-R are the primers used for rt-PCR. Right, assessment of the mRNA abundance by end-point rt-PCR in mutants and controls. Bottom, PDF2 abundance was measured as an internal control. WT= wildtype *A. thaliana*; G: gDNA control, 1-3 are mutants.

Table 3.3: Identification and characterization of T-DNA mutant alleles in *AtAlkB* genes.

Gene Name	Mutant allele	T-DNA allele	Genotype	mRNA abundance
AtAlkB	*atalkb*	FLAG_143G11	Homozygous	not detected
	AtALKB-2	SALK_062071	Homozygous	same as WT
AtAlkB-B	*atalkb-b*	SK15155	Homozygous	not detected
	AtAlkB-B-2	SALK_0523400	Homozygous	same as WT
AtAlkB-C	*atalkb-c-1*	SAIL_193_C03	Homozygous	not detected*
	atalkb-c-2	SALK_021477C	Homozygous	not detected
	AtALKB-C-3	SALK_035671	Homozygous	reduced
AtALKBH1	*atalkbh1*	SALK_007964	Homozygous	not detected
	AtALKBH1-2	SALK_102868	Heterozygous	-
AtALKBH2B	*atalkbh2b*	SAIL_68_G03	Homozygous	not detected
	AtALKBH2B-2	SAIL_223_C05	Homozygous	same as WT
AtALKBH2	*AtALKBH2-1*	WHS085_11E	Homozygous	Increase
	-	SAIL_1239_H02	Heterozygous	-
AtALKBH5	*AtALKBH5-1*	Gabi_423C08	Homozygous	same as WT
	AtALKBH5-2	FLAG_156D06	Homozygous	same as WT
AtALKBH6	*atalkbh6*	SALK_138864	Homozygous	not detected
	AtALKBH6-2	FLAG_564H09	Homozygous	same as WT
AtALKBH8	*atalkbh8-1*	SALK_083838C	Homozygous	not detected
	atalkbh8-2	SALK_094502C	Homozygous	not detected
AtALKBH8B	*atalkbh8b*	SALK_024872	Homozygous	not detected
	AtALKBH8B-2	SAIL_756_A08	Homozygous	same as WT
AtTRM9	*attrm9*	SALK_135308C	Homozygous	not detected
	-	SALK003180	Heterozygous	-
AtALKBH9A	*atalkbh9a*	SALK_204823	Homozygous	not detected
	AtALKBH9A-2	SAIL_402E12	Homozygous	same as WT
AtALKBH9B	*AtALKBH9B-1*	SALK_111811	Homozygous	same as WT
	-	WisD501D02	Heterozygous	-
AtALKBH9C	*atalkbh9c*	SALK_021775	Homozygous	not detected
	AtALKBH9C-2	WHS118_10A	Homozygous	Decrease

* not detected- the mRNA was absent after rt-PCR, and therefore the mutant was considered a complete loss of function.

Due to time constraints, we could not test more T-DNA lines and obtain knock-out mutants of *AtAlkBH2*, *AtAlkBH5* and *AtAlkBH9B*. In this study, I characterized *AtAlkB*, *AtAlkB-B*, *AtAlkB-C*, *AtAlkBH1*, *AtAlkBH2B*, *AtAlkBH6*, *AtAlkBH8*, *AtAlkBH8B*, *AtTRM9*, *AtALKBH9A*, *AtALKBH9C* in regulating RNA 5mC/5-hmC in *A. thaliana*.

3.4.3. Induction of heat shock treatment to *atalkb-like* mutants

The next step was to induce an increase in 5mC on a target mRNA. Our data in Chapter 2, the overall mRNA 5mC level in *A. thaliana* is very low (about 0.03%), making it relatively difficult to observe the change of *atalkb*-like mutants on RNA 5mC. However, a heat shock treatment of 40°C for 30 min induced a 3-times increase of 5mC level on many transcripts (Chapter 2). The rapid increase of methylation after heat shock treatment suggested a regulatory mechanism on RNA 5mC in response to heat shock. These may involve the participation of a demethylase, and further experiments were planned but not conducted. In this chapter, I induced the same heat shock treatment as in Chapter 2 to examine the effect of *atalkb-like* mutants on the methylation level of specific mRNA 5mC sites.

To confirm the effect of heat shock treatment on the transcriptional level, I measured the mRNA abundance of a heat-sensitive gene *HSP20* (*Heat Shock Protein 20*, *AT1G07400*) relative to an internal control *PDF2* by semi-quantitative rt-PCR. The abundance of *HSP20* in heat-treated samples was five times higher than in the control samples, demonstrating that the heat shock treatment successfully induced transcriptional changes (Figure 3.3).

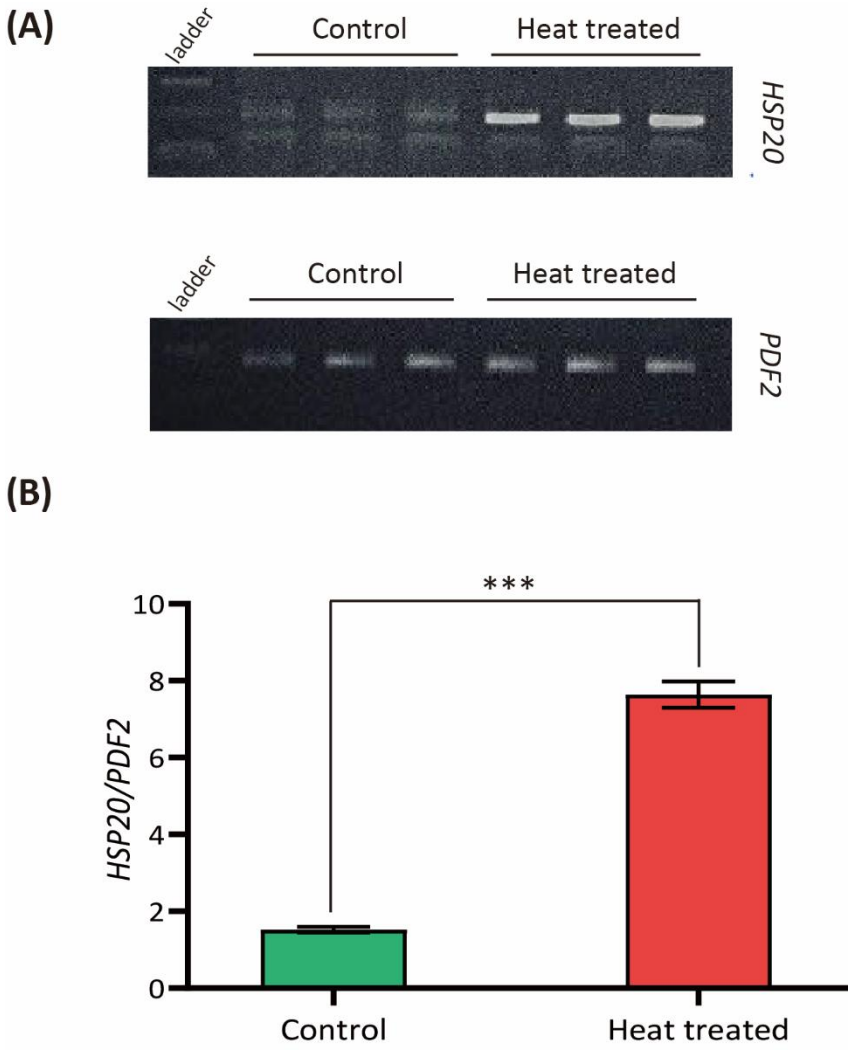


Figure 3.3: Induction of heat-responsive gene *HSP20* after heat shock treatment.

To validate the effect of the heat shock treatment on 10-day-old *A. thaliana* seedlings, the mRNA abundance of one well-characterized heat-response gene, *HSP20,* was measured after both control conditions and after heat shock treatment. RNA was purified, and semi-quantitative rt-PCR was performed in triplicate.

(A) Agarose gel detection of semi-quantitative rt-PCR products of heat-responsive gene *HSP20* and the control gene *PDF2*. The DNA ladder is the 100 bp ladder and shows the *HSP20* and *PDF2* fragments. **(B)** Quantifying *HSP20* mRNA abundance relative to *PDF2* as shown in (A). A PCR band density test by ImageJ was performed to compare the abundances of each PCR-amplified band in (A). *HSP20* mRNA abundance was significantly increased by about 4 times compared to the control. *** P value <0.001. The error bars shown are the standard deviation of the mean of 3 replicates.

3.4.4. Measuring RNA 5mC abundance in *alkb-like* mutants

During my Master's research, I identified a heat-responsive RNA 5mC site, C_{3349} of *MAG5* (*AT5G47480*), in total RNA isolated from whole seedlings. The *MAG5* C_{3349} steady-state, 5mC level increased from 48 to 71% after heat shock treatment using RNA Bisulfite amplicon sequencing (RBS-amplicon-seq, unpublished data). The steady-state *MAG5* mRNA abundance did not change after heat shock compared (unpublished data) and previously was shown to have the same mRNA abundance in WT and the methyltransferase mutant *trm4b* (David et al. 2017). It is worthwhile to reintegrate that BS-amplicon-seq cannot discriminate between 5mC, 5hmC or another cytosine modification that prevents the bisulfite-mediated conversion of cytosine. However, it is possible to differentiate the various modifications with other approaches, like antibody-specific RNA-IP. The apparent increase of *MAG5* C_{3349} methylation after heat shock could be interpreted in several ways, including (1) degradation of unmodified *MAG5* mRNAs, (2) increased transcription of *MAG5* mRNAs that were subsequently methylated, (3) attenuation of a 5mC mRNA degradation mechanism, (4) increased cytosine methyltransferase activity or (6) induction of a mechanism converting m5C to another modification like 5hmC, at the single cell or tissue level during the heat shock period of 30 mins. A combination of these plausible mechanisms cannot be ruled out. As the steady-state *MAG5* mRNA abundance before and after heat shock was the same, I ruled out mechanisms 1, 2 and 3.

Dynamic alteration of *MAG5* C_{3349} 5mC level after heat shock encouraged us to explore this site further. In this Chapter, the methylation level of *MAG5* C_{3349} before and after heat shock treatment in WT and *atalkb* mutants was tested as

a step forward to investigate potential RNA 5mC demethylase in plants. A second RNA cytosine site, C_{328} of *AT2G36120,* showed a constant 5mC level before, and heat shock treatment was used as a control. RBS-amplicon-seq was performed on RNA isolated from WT and *atalkb-like* mutants before and after heat shock. Briefly, purified RNA samples were treated with sodium metabisulfite to convert unmodified cytosines to uracils, while modified cytosines were not converted. During BS-treated, RNA is fragged, making the RNA difficult to synthesize first-strand cDNA using an oligo dT primer. Hence, we used a gene-specific primer (GSP, Supplementary Table 3.2) 3' of *MAG5* C_{3349} and *AT2G36120* C_{328} to synthesise the first strand cDNAs of BS-treated samples. Then, two rounds of PCR were performed to amplify target amplicons and add Fluidigm universal tags CS1 and CS2, Illumina adapters and Fluidigm barcodes. RBS-amplicons were sequenced on an Illumina 2x150 MiSeq system. Sequencing reads were trimmed using Trim Galore! and then mapped to in silico BS-converted reference sequences using meRanTK. The coverage of each sample ranged from 1,000-10,000 reads, and the overall bisulfite conversion rate was over 99.5%.

The overall sequencing results are shown in Figures 3.4 & 3.5 (see Supplementary Figures 3.1 & 3.2 and Supplementary Table 3.4). As expected, the methylation level at *AT2G36120* C_{328} before and after heat shock treatment was similar in WT and the mutants*,* ranging from 20 to 30% (Figure 3.4). No significant difference was found amongst WT, mutants, control or the heat shock treatment. This confirmed *that AT2G36120 C328 was a heat-insensitive RNA 5mC site and showed that any AtAlkB-like genes do* not regulate the site. In contrast, *MAG5* C_{3349} is a heat-response 5mC site. Under control conditions, the methylation level of *MAG5* C_{3349} in most mutants was between 25 to 35%, except for *atalkb, atalkbh1, ataklkbh2b* and *atalkbh6,* which showed about 43%

(Figure 3.5). Although the 5mC level was higher in these mutants, further replication is required due to the high variation amongst replicates.

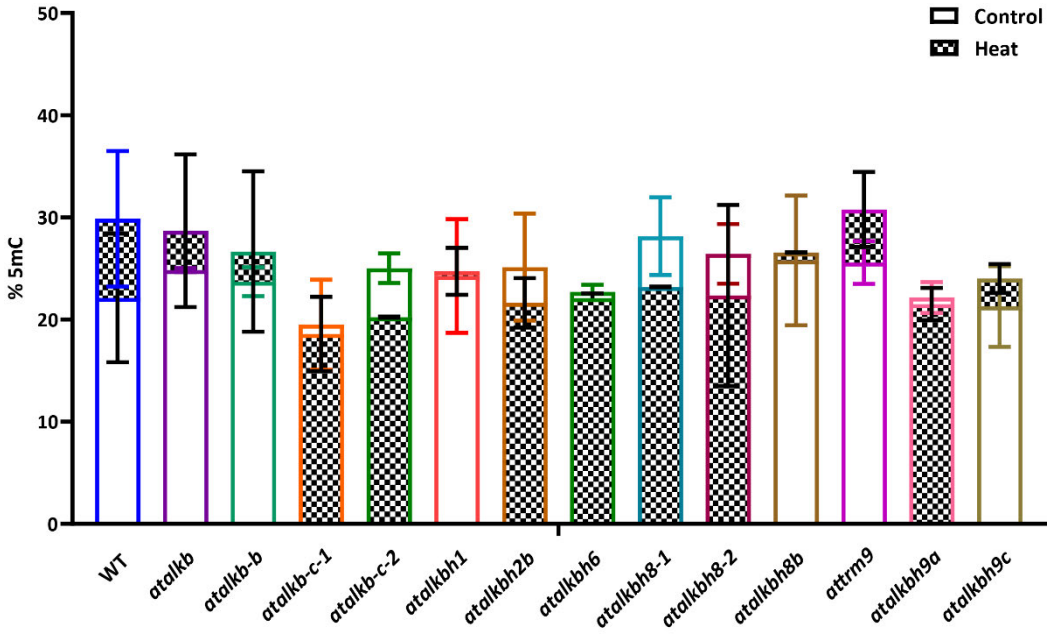


Figure 3.4: RNA 5mC level of *atalkb-like* mutants at control site C_{328} of *AT2G36120*.

Wild type (WT) and 11 *atalkb-like* mutants, *atalkb* (FLAG_143G11), *atalkb-b* (SK15155), *atalkb-c* (*atalkb-c-1*: SALK_021477C; *atalkb-c-2*: SAIL_193_C03), *atalkbh1* (SALK_007964), *atalkbh2b* (SAIL_68_G03), *atalkbh6* (SALK_138864), *atalkbh*8 (*atalkbh8-1*: SALK_083838C; *atalkbh8-2*: SALK_094520C), *atalkbh8b* (SALK_024872), *attrm9* (SALK_135308C), *atalkbh9a* (SALK_204823), *atalkbh9c* (SALK_021775). *A. thaliana* seedlings were grown under control conditions for 10 days and harvested (shown as no filling in the graph) or harvested after applying a 40°C, 30 mins heat shock treatment (shown as filling in the graph). RNA was purified from the seedlings, and RBS-amplicon-seq was performed on triplicates to detect the 5mC level at C_{328} of *AT2G36120*. The Y-axis shows the 5mC level as a percentage; the X-axis shows the sample name. Control conditions are shown in white, and after heat treatment is shown in black dots. The error bars show the standard deviation of the mean of three replicates.

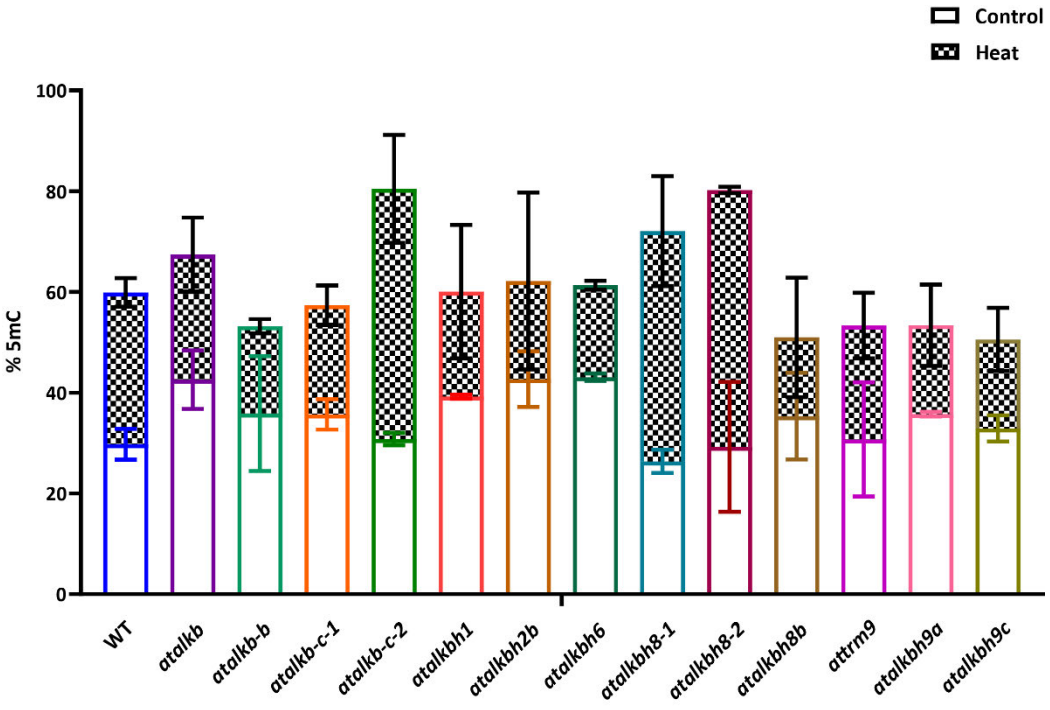


Figure 3.5: RNA 5mC level of *atalkb-like* mutants at C_{3349} of *MAG5*

The RNA cytosine methylation levels of Wild type (WT) and 11 *atalkb-like* mutants at *MAG5* C_{3349} were measured as described in Figure 3.4. Control conditions are shown in white, and after heat treatment is shown in black dots. The error bars show the standard deviation of the mean of two replicates.

After heat shock treatment, WT and all the mutants showed increased 5mC at *MAG5* C_{3349} (Figure 3.5). In WT, the methylation level of *MAG5* C_{3349} significantly increased from about 30 to 60% after heat shock treatment as expected ($P < 0.05$). Similarly, a 15-25% increase in 5mC after heat treatment was observed in *atalkb, atalkb-b, atalkbh1, ataklkbh2b, atalkbh6, atalkbh8b, attrm9, alkbh9a* and *atalkbh9c* mutants. This suggested that these genes do not play a role in the heat shock response at *MAG5* C_{3349}. In contrast, mutant *AT5G01780*/*atalkb-c-2* showed 80% methylation, 20% more than WT, after heat shock treatment, although the second allele *atalkb-c-2* showed a similar 5mC heat response as WT. Due to time constraints, I did not have the opportunity to repeat the *atalkb-c-2* result. In the future, I suggest confirming these *AT5G01780*/*atalkb-c* results.

Both mutant alleles of *AtALKBH8*, *atalkbh8-1* and *atalkbh8-2,* showed similar methylation levels at *MAG5* C_{3349} as WT under control conditions but remarkedly showed over 80% methylation, 20% more than WT, after heat shock treatment (Figure 3.5). The significantly higher increase of 5mC levels in both *AtALKBH8* mutants shows that *AtALKBH8* plays either a direct or indirect role. It is likely that *AtALKBH8* indirectly affects *MAG5* C_{3349} methylation as previously, Leihne et al. (2011) showed AtALKBH8, in a complex with AtTRM12, is a hydroxylase that generates 5-(S)-[methoxycarbonyl-(hydroxy) methyluridine, from 5-methoxycarbonyl-methyuridine on $tRNA^{GLY}_{UCC}$.

3.4.5. Identifying the potential roles of AtALKBH8-like genes in demethylating RNA 5mC

The increased methylation level detected at *MAG5* C_{3349} of two *atalkbh8*/ *at1g31600* mutant alleles (*atalkbh8-1* and *-2*) after heat shock treatment caught our attention. AtALKBH8/AT1G31600 has high similarity to the N-terminal region of human ALKBH8, including the RRM and AlkB domains (Leihne et al. 2011). The currently described function of AtALKBH8 is to hydroxylate 5-methoxycarbonylmethyluridine (mcm^5U) to (S)-$mchm^5U$ in $tRNA^{Gly}$ (Leihne et al. 2011). The C-terminal part of human ALKBH8 is encoded by another gene, *AT1G36310* (*AtTRM9*), in *A. thaliana,* which can catalyse the final step in mcm^5U formation (Leihne et al. 2011). However, the current understanding of these two genes in *A. thaliana* is still limited. Whether *AtALKBH8* homologous (*AT1G36310*, *AT1G31600*) can directly or indirectly alter RNA 5mC remains to be studied.

We selected two additional RNA 5mC sites, C_{1353} of *AT5G17920* and C_{389} of *AT1G71350*, together with C_{3349} of *MAG5* and C_{328} of *AT2G36120*, to further characterise *AtALKBH8* and *AtTRM9's* roles in regulating RNA 5mC. Due to time

and funding limitations, we could only test the methylation level in *AtALKBH8* and *AtTRM9* mutants, *atalkbh8-2* and *attrm9-1* at these four sites via RBS-amplicon-seq under control conditions. The RBS-amplicon-seq results are shown in Figure 3.6 (Supplementary Figure 3.3 and Supplementary Table 3.5). The average methylation level of C_{328} of *AT2G36120*, C_{3349} of *MAG5*, C_{1353} of *AT5G17920* and C_{389} of *AT1G71350* in WT was 29%, 44%, 4% and 19%, respectively. Both two mutants, *atalkbh8-2* and *attrm9-1*, had similar methylation levels as WT at C_{328} of *AT2G36120*, C_{3349} of *MAG5* and C_{1353} of *AT5G17920*. We observed slightly higher methylation levels at C_{328} of *AT2G36120* and C_{3349} of *MAG5* in WT and mutant samples compared to the previous experiments (Figure 3.4 & 3.5). Unexpectedly, *atalkbh8-2* and *attrm9-1* mutants only had 8% 5mC at C_{389} of *AT1G71350*, compared with 19% in WT. The decreased 5mC level of *atalkbh8-2* and *attrm9-1* at C_{1353} of *AT5G17920* may be caused by PCR or sequencing bias or some unknown coactions between *AtALKBH8* and *AtTRM9* to maintain methylation level at this site. We did not observe any increased methylation in a single mutant in this dataset, suggesting that *AtALKBH8* and *AtTRM9* do not have a broad role in 5mC. However, I did not explore the role of these two genes in heat stress response on RNA 5mC.

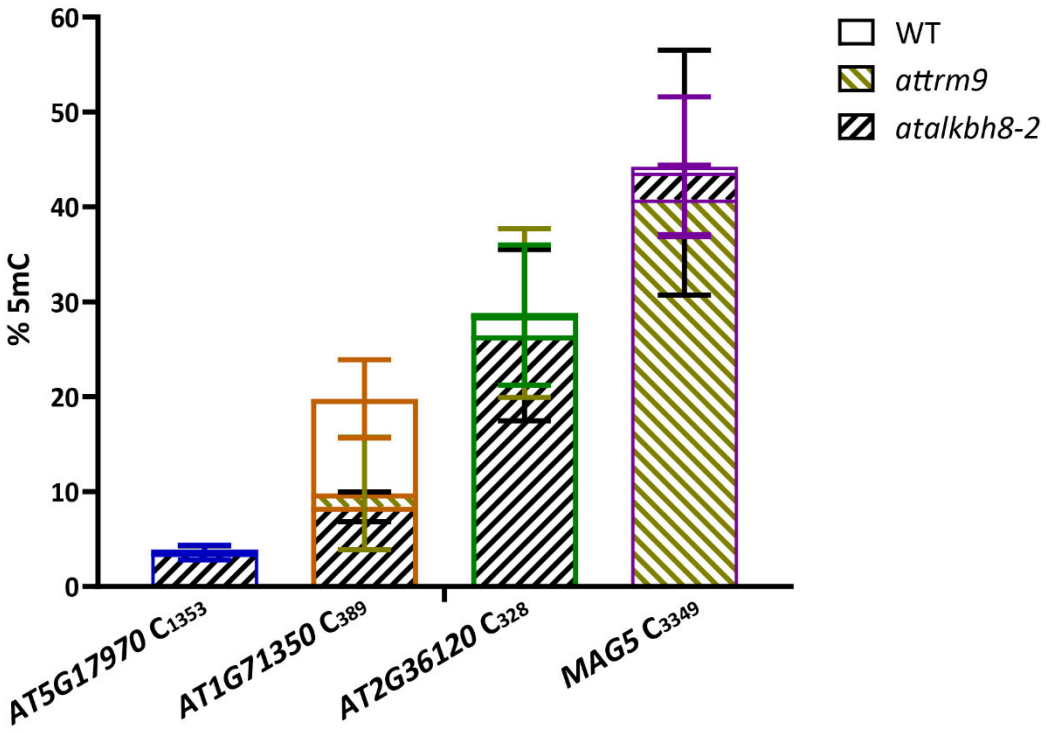


Figure 3.6: 5mC levels at 4 RNA 5mC sites in WT, *atalkbh8* and *attrm9*.

The RNA cytosine methylation levels at 4 RNA 5mC sites: C_{1353} of *AT5G17920*, C_{389} of *AT1G71350*, C_{328} of *AT5G47480* and C_{3349} of *MAG5* were measured by RBS-amplicon-seq in WT, *atalkbh8-2* (*at1g31600-2*) and *attrm9* (*at1g36310)* after control conditions. The X-axis shows the assessed 5mC sites, and the Y-axis shows the 5mC level as a percentage. Wild type (WT) is shown in white; *atalkbh8-2* (*at1g31600-2*) is shown in black slash; *attrm9* is shown in golden slash. RBS-amplicon-seq was performed in triplicate, and the error bars shown are the standard deviation of the mean of three replicates.

3.5. Discussion

3.5.1. RNA 5mC may participate in heat stress responses in plants

Plants are sessile organisms that are often exposed to environmental stresses. They have evolved highly complex mechanisms at the transcriptional and post-transcriptional levels to respond to abiotic stresses (Atkinson & Urwin 2012). For example, in *A. thaliana* and rice, the YTH protein that 'reads' RNA *N6*-mA has

distinct expression patterns and increased transcription in response to abiotic and biotic stresses (Li, D et al. 2014). However, currently, there is limited evidence for the role of RNA 5mC in stress responses. Our RBS-amplicon-seq data showed that a 40°C heat shock for 30 min led to an increase of RNA 5mC at C_{3349} of *MAG5*. This could be a part of the early heat stress response in plants. Although the exact mechanism and purpose of up-regulated RNA 5mC are still unclear, this data provides an initial landscape of RNA 5mC. The increased methylation level may have roles in stabilizing RNA structures, promoting RNA-protein interactions, or regulating gene expression to meet the challenge of heat stress. Similar results also have been found in DNA that abiotic stresses can alter the gene expression via demethylation or increased methylation of DNA cytosine (Qiao & Fan 2011). Altered DNA methylation levels have been demonstrated to regulate transcription factor chromatin interactions (Lewis, J & Bird 1991; Tate & Bird 1993; Turek-Plewa & Jagodzinski 2005). Our research on increased RNA 5mC after heat shock treatment suggests another layer of regulation in plants that warrants further investigation.

3.5.2. Potential demethylation of *MAG5* C_{3349}

AlkB and AlkB-like proteins are responsible for most oxidative demethylation pathways. In mammals, AlkB human homologies, ALKBH1-8, FTO and TET have been found to regulate the methylation level of various substrates. However, the understanding of AlkB-like proteins and demethylation processes is still very limited in plants. To identify potential RNA 5mC demethylase in plants, we identified 14 candidates by aligning protein sequences of AlkB from *E. coli* and ALKBH1-8 from *H. sapiens* with *A. thaliana* protein database (TAIR 11) and identified knock-out T-DNA lines for 11 genes. Most of these candidates are in the 2-oxoglutarate-dependent dioxygenase family and contain the essential

HXD…H catalytic motif. The target substrates and functional roles of these proteins largely remain unknown. In my study, I explored the regulation activity of these 11 AlkB-like candidates toward the *MAG5* C_{3349}. Although no 5mC demethylase was identified, my result suggested some direct or indirect roles of AlkB-like candidates in regulating the 5mC level.

AT5G01780 is a candidate AtAlkB protein that contains a Fe (II)-2OG dioxygenase domain and an HXD…H motif. The increased 5mC level on *MAG5* C_{3349} detected in the *atalkb* mutant suggests that AtAlkB might target some specific RNA 5mC sites. Until now, the particular function of AT5G01780 is still unclear. Further study is recommended to reveal the mystery of AT5G01780 in regulating RNA 5mC.

Only recently, AT4G20350 was identified in *A. thaliana* as the AtALKBH6, an RNA demethylase that targets *N6*-mA and possibly 5mC (Huong, Ngoc & Kang 2020). The binding ability of ALKBH6 towards RNA 5mC was much lower than towards *N6*-mA. Similarly, our data showed that the methylation level of the *atalkbh6* mutant at the C_{3349} site of *MAG5* was 43%, slightly higher than that of WT at this site (29%). Huong, Ngoc and Kang (2020) also pointed out that ALKBH6 has significant roles in abiotic stress responses. The expression level of *AtALKBH6* was up-regulated after high salinity but down-regulated after cold or abscisic acid treatment. However, their data showed that ALKBH6 transcription is not sensitive to heat shock. Our data provided no evidence of *AtALKBH6* in regulating methylation at C_{3349} of *MAG5*. Whether *AtALKBH6* regulates the methylation level of other RNA 5mC sites requires further research.

ALKBH8 in mammals is a multifunctional enzyme that not only can methylate 5-carboxymethyluridine (c5mU) to 5-methoxycarbonylmethyluridine (mc5mU) but also hydroxylate mc5mU to (S)-5-methoxycarbonylhydroxymethyluridine (S)-

mchm^{5}U at the wobble position of tRNA$^{Gly}_{UCC}$ (Leihne et al. 2011; Songe-Moller et al. 2010; Zdzalik et al. 2014). In *A. thaliana*, the function of ALKBH8 is performed by two separate genes, *AtTRM9* (*AT1G36310*) as the methyltransferase and *AtALKBH8* (*AT3G31600*) as the hydroxylase. The only described substrate of *AtALKBH8* is mc5mU in a tRNA, and my data provides evidence of *AtALKBH8* in regulating mRNA 5mC levels in response to heat shock. I observed no effect of AtTRM9 on methylation levels. The higher 5mC level observed in *atalkbh8-1* and *atalkbh8-2* after heat shock treatment suggests the potential interaction between 5mC and mc5mU or (S)-mc5hmU, or some unknown activity of *AtALKBH8* in demethylating RNA 5mC. A more plausible explanation is an indirect effect of *AtALKBH8* on this 5mC site.

3.5.3. The potential role of *ALKBH8-like* genes on other RNA 5mC sites

To further investigate the role of ALKBH8 in regulating RNA 5mC level, we measured the methylation levels at C_{328} of *AT2G36120*, C_{3349} of *MAG5*, C_{1353} of *AT5G17920* and C_{389} of *AT1G71350* under control conditions. We did not observe any increase in methylation level at these four sites in the mutants, suggesting that AtALKBH8 might not be able to reduce RNA 5mC level after control conditions. Surprisingly, the loss of either *AtALKBH8* or *AtTRM9* leads to a consistently lower 5mC level than WT at C_{389} of *AT1G71350*. This could be caused by the direct role of AtALKBH8 and AtTRM9 in maintaining the 5mC level at C_{389} of *AT1G71350* or the indirect disruption of this 5mC site due to the alteration of c5mU and mc5mU level. Although the current understanding of AtALKBH8 and AtTRM9 is still limited, our data provide a new viewpoint on these two genes in regulating other RNA modifications.

3.5.4. Conclusion

In summary, this chapter provides an early study of putative RNA oxidation mutants in *Arabidopsis*. For the first time in plants, we investigated the role of *AlkB-like* genes in regulating RNA 5mC levels in response to heat shock at two 5mC sites, C_{3349} of *MAG5*/*AT5G47480* and C_{328} of *AT2G36120*. The *MAG5* C_{3349} site is a heat-response 5mC site, and the methylation level is significantly increased after thermal stress. In contrast, 5mC of *AT2G36120* C_{328} is unresponsive to thermal stress having the same m5C level under control conditions and after heat shock. In the tested *alkb-like* mutants, no significant effect of 5mC on the tested sites was observed under control conditions. However, we interestingly found increased 5mC levels at *MAG5* C_{3349} in mutants of *AT1G31600/AtALKBH8* and *AT5G01780/AtAlkB-C.*

Although it is currently challenging to conclude the exact roles of AT1G31600*/AtALKBH8* and AT5G01780*/AtAlkB-C* in regulating 5mC under thermal stress, my preliminary results provide interesting avenues to follow. As *AT1G31600/AtALKBH8* modifies tRNAs, the effect on mRNA 5mC after thermal stress may be indirect. One possible indirect explanation is that *AT1G31600/AtALKBH8* is required to efficiently translate a target protein regulating RNA 5mC during thermal stress. We also tested ALKBH8 homologs, AT1G31600 and AT1G36310, in regulating RNA 5mC levels but did not find any

remarkable effects in the mutants. *AT5G01780/AtAlkB-C* is an exciting gene to investigate further the protein's role in regulating 5mC during thermal stress. After validating my initial findings, a first experiment would be to determine the transcriptome-wide effects of *AT5G01780/AtAlkB-C* on 5mC after heat stress using RBS-seq. Assuming significant effects on m5C transcriptome-wide, it would be interesting to next test if *AT5G01780/AtAlkB-C* directly binds to 5mC containing RNA regions and then the *in vivo* biochemical activity and oxidative products. How RNA 5mC levels are regulated *in situ* remains largely unknown and requires further research.

4. General Discussion

This study aimed to fill the knowledge gaps of RNA 5mC in environmental-stress responses and the modification's regulatory pathways in plants. 5mC is a common modification identified in all types of cellular RNAs, especially on mRNAs. In mammals, the roles of RNA 5mC have been linked to cell differentiation and development, stress responses, various diseases, and cancers (Abbasi-Moheb et al. 2012; Blanco et al. 2016; Blanco & Frye 2014; Bohnsack, Höbartner & Bohnsack 2019; Flores et al. 2017; Trixl, L. et al. 2018). However, our current understanding of RNA 5mC in plants is far behind mammals. Our previous research found that the RNA 5mC writer (*TRM4b*) was linked to root stem cell proliferation and oxidative stress response in *A. thaliana* (David et al. 2017). Tang, Yongyan et al. (2020) also pointed out the role of RNA 5mC in responding to modest heat stress in rice. Although the exact role of 5mC in stress response is still elucidated, Tang, Yongyan et al. (2020) provided an initial framework linking the modification to photosynthesis. In Chapters 2 and 3, I explored the functional role and regulation pathways of RNA 5mC in response to heat shock treatment.

Plant heat stress response is a complex process involving transcriptional, post-transcriptional, translational, post-translational, and other cellular effects. The transcriptional and translational regulatory network involves the expression, activation and degradation of related proteins and enzymes (e.g., HSFs, HSPs, ROSs) to proceed with the acquisition of immediate heat response, thermotolerance and long-term heat-adaption (Echevarría-Zomeño et al. 2016; Ohama, N. et al. 2017; Wu, JR et al. 2017). While large is still unclear how related proteins are precisely generated and modulated to perform

corresponding tasks in the thermal response fleetly. One reasonable assumption is that gene expression and protein synthesis are after real-time monitoring and adjustment of RNA modifications, so-called post-transcription regulation. Different from direct DNA expression and transcription, the regulation of protein synthesis by post-transcriptional modifications is implicit and flexible. For example, the addition of *N6*-mA on mRNA could enhance mRNA structure stability, splicing of precursor mRNA into mature mRNA, and mRNA export from nuclear to the cytoplasm, so as to promote the transcription and translation of targeting mRNAs (Camper et al. 1984; Carroll, Narayan & Rottman 1990; Dominissini et al. 2012; Niu et al. 2013). While the removal of *N6*-mA might inhibit or even stop the transcription and translation of mRNA. To some extent, the dynamic change of *N6*-mA level can timely control the production of target mRNA and protein. This improvisational adjustment on mRNAs and protein by RNA modifications is also highly likely to be abundant in the heat stress response. My research focuses on how another widely presenting RNA modification, 5mC, responds to heat shock treatment.

4.1. Does RNA 5mC have a critical role in thermal tolerance?

Chapter 2 of my thesis is the first to describe the landscape of RNA 5mC after heat shock stress. Heat shock-induced a transcriptome-wide alteration of RNA 5mC abundance. Over 3,500 new 5mC sites emerged, and over 1,000 5mC sites were reduced after heat shock. Overall, the RNA 5mC level is also increased mainly after heat shock. Notably, in mRNAs, the 5mC coverage after heat shock treatment is about 3 times higher than that after control conditions. These data suggest that 5mC frequently participates in many co-transcriptional or post-transcriptional activities in response to heat stress. Before my study, little was

known about the dynamic nature and trajectory of RNA 5mC. My finding can be a crucial code to unlock the profile of RNA 5mC in the post-transcriptional regulatory network in heat stress response

In Chapter 2, I also investigate the effect of mRNA 5mCs on gene expression in heat shock response. According to our current knowledge, 5mC participates in RNA-protein binding to proceed with mRNA export, splicing, decay and other functions (Amort et al. 2017; Yang, X et al. 2017). There is not much information about the mRNA 5mC in regulating gene expression. In my study, I identified a small proportion of 5mC sites that can regulate the expression of the targeting gene. In particular, 7 highly methylated Cs on the 5'UTR of *At1g05340* can maintain the gene expression level to resist heat shock treatment. This finding suggests that RNA 5mC can act as an emergency regulator of gene expression for urgent heat shock response. For a very long time, DNA was thought to be the only initiator of gene coding. While in fact, the post-transcription network should be the first to sense and respond to any unexpected changes in protein requirements. DNA coding is often hysteretic and slow for urgent situations like heat shock. The transcriptome must have some self-autonomy to adjust protein production in response to emergencies urgently. RNA modifications like 5mC, so-called RNA coding, mainly perform this privilege. Instead of generating transcripts, RNA coding solves the demand in various ways, such as adjusting the transcription and translation efficiency. While at the current stage, this hypothesis remains theoretical. Further studies are encouraged to resolve the specific mechanism of RNA 5mC in managing gene expression.

Although most heat-response 5mC sites I detect have little or no effect on gene expression, their potential roles in post-transcriptional regulation cannot be ruled out. One possible explanation is that they are potential binding sites of RNA binding proteins (RBPs) that involve in heat shock response. Various post-

transcriptional controls, such as mRNA transport, mRNA localization, mRNA metabolism and RNA structure stabilization, require the participation of numerous RBPs. As many of these processes are transient or one-off, the binding of RBPs with targeting sites is usually dynamic and reversible. After the threat of heat shock, transcripts need to urgently recruit new RBPs or remove unneeded RBPs for stress response. Hence, the distribution and abundance of these RNA 5mC sites are temporarily altered to perform new post-transcriptional regulation on transcripts. If so, these 5mC sites can be seen as de facto manipulators of post-transcriptional control in response to heat shock.

4.2. RNA 5mC is after the regulation of multiple pathways

The dynamic alteration of RNA 5mC level after heat shock raises my interest in their regulatory pathways. The methylation pathway of RNA 5mCs is relatively well-studied, and many methyltransferases have been recognized in plants, such as TRM4B and TRDMT2. However, the demethylation pathway and demethylases have not yet been elucidated in plants. In chapter 3, I tentatively explored the potential demethylation activity of 11 AlkB-like candidates on RNA 5mC in *A. thaliana* after control and heat shock treatment. Unluckily, no specific evidence of 5mC demethylase is identified in my research other than clues of some candidates in modulating the 5mC level. It is highly possible that the RNA 5mC level can be regulated by multiple pathways, directly or indirectly.

AtALKBH6 (AT4G20305) can partially inhibit the methylation level at *MAG5* C_{3349} after both control and heat shock treatment. This result largely supports the idea of a recent study that AtALKBH6 had a low demethylation activity toward RNA 5mC (Huong, Ngoc & Kang 2020). Also, they found that AtALKBH6 can

control seed germination and seedling growth against abiotic stresses like cold, salt, or abscisic acid stress but not heat stress. This might explain why AtALKBH6 is not sensitive to heat stress in my result. Both Houng's and my studies suggest that AtALKBH6 is a minor or alternate demethylase of RNA 5mC, but its demethylation activity toward other substrates cannot be ruled out. To a large extent, AtALKBH6 defers seed germination and seedling growth through dynamic regulation of RNA methylation. This also suggests the vital status of RNA modifications or the post-transcription network in regulating the growth and development of plants and responding to environmental changes.

Another interesting point is AtALKBH8 (AT1G31600), which inhibits the methylation level of *MAG5* C_{3349} after heat shock treatment but enhances the methylation level of *AT1G71350* C_{389} after control conditions, suggesting that instead of demethylating 5mC, AtALKBH8 can adjust the 5mC level to some extent. This may be achieved by either RNA-RNA interaction or RNA-protein interaction. The current known function of ALKBH8 is the hydroxylation of mc5mU to (S)-mc5hmU at the wobble position of $tRNA^{Gly}_{UC}$ (Leihne et al. 2011; Songe-Moller et al. 2010; Zdzalik et al. 2014). The main function of tRNA is transcript decoding and protein synthesis. The transformation of mc5mU to (S)-mc5hmU may affect the paring of $tRNA^{Gly}_{UCC}$ with targeting transcript. Hence, some 5mC sites like *MAG5* C_{3349}, as a part of the post-transcriptional regulation, may also be regulated to respond to the translation change. Another explanation is that the mc5mU to (S)-mc5hmU transformation of $tRNA^{Gly}_{UCC}$ may regulate the synthesis of some RBPs that bind with 5mC sites, thereby indirectly adjusting the methylation level of binding 5mC sites to match RBPs. Either possibility illustrates the complexity and magnificence of the post-transcriptional regulation network. From RNA and protein to nucleotide bases and their chemical modifications, they are tightly linked like cogs, cooperating to

perform every complex regulation task that ensures the proper functioning of life.

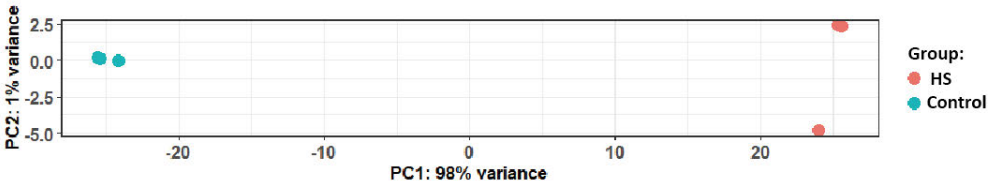
4.3. Conclusions and significance

In summary, this thesis highlights the potential role of RNA 5mC as a post-transcriptional regulator in response to heat shock in plants. For the first time, I describe in detail the landscape, function and potential demethylation pathway of RNA 5mC after heat shock treatment, and these findings lay a solid foundation for future research. Exploring RNA 5mC has far-reaching consequences for understanding the post-transcriptional regulatory network in managing gene expression, protein synthesis, and stress tolerance in crops. The gradual aggravation of global temperatures will significantly affect crops' growth and production. Improving crops' capacity to tolerate heat stress has become a hot topic in biology and agriculture, and this research provides a new avenue to address this immediate humanitarian challenge. In the future, manipulating specific 5mC sites may lead to heat tolerance and higher-yielding crops.

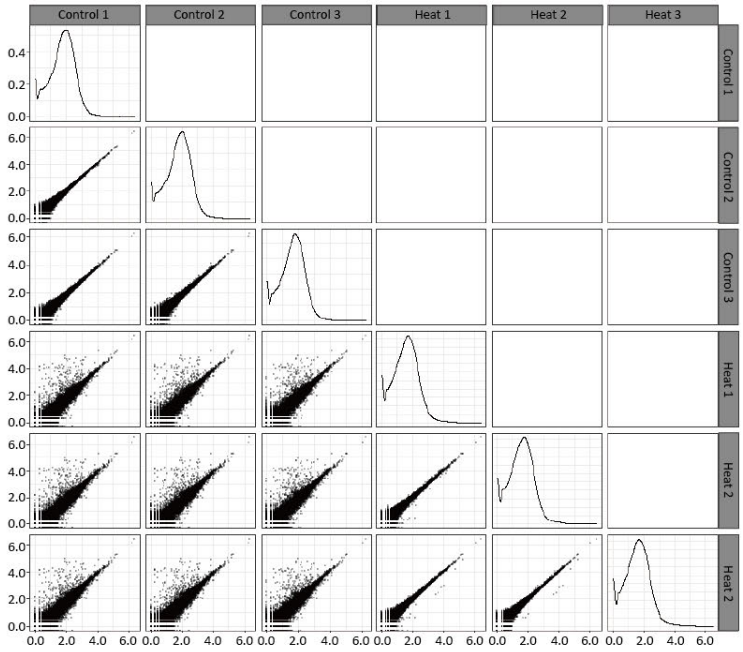
5. Supplementary Data

5.1. Supplementary Figures

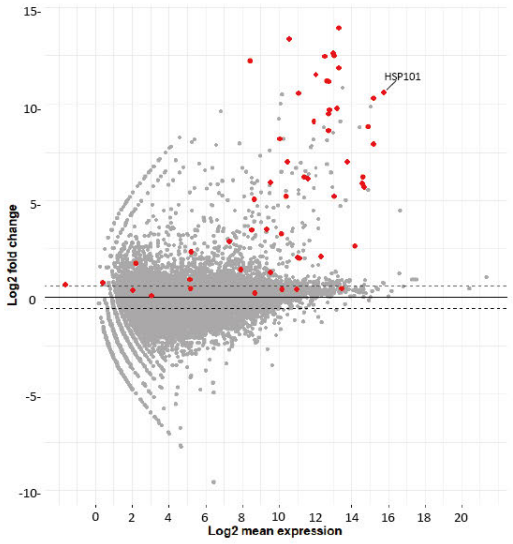
(A)



(B)



(C)



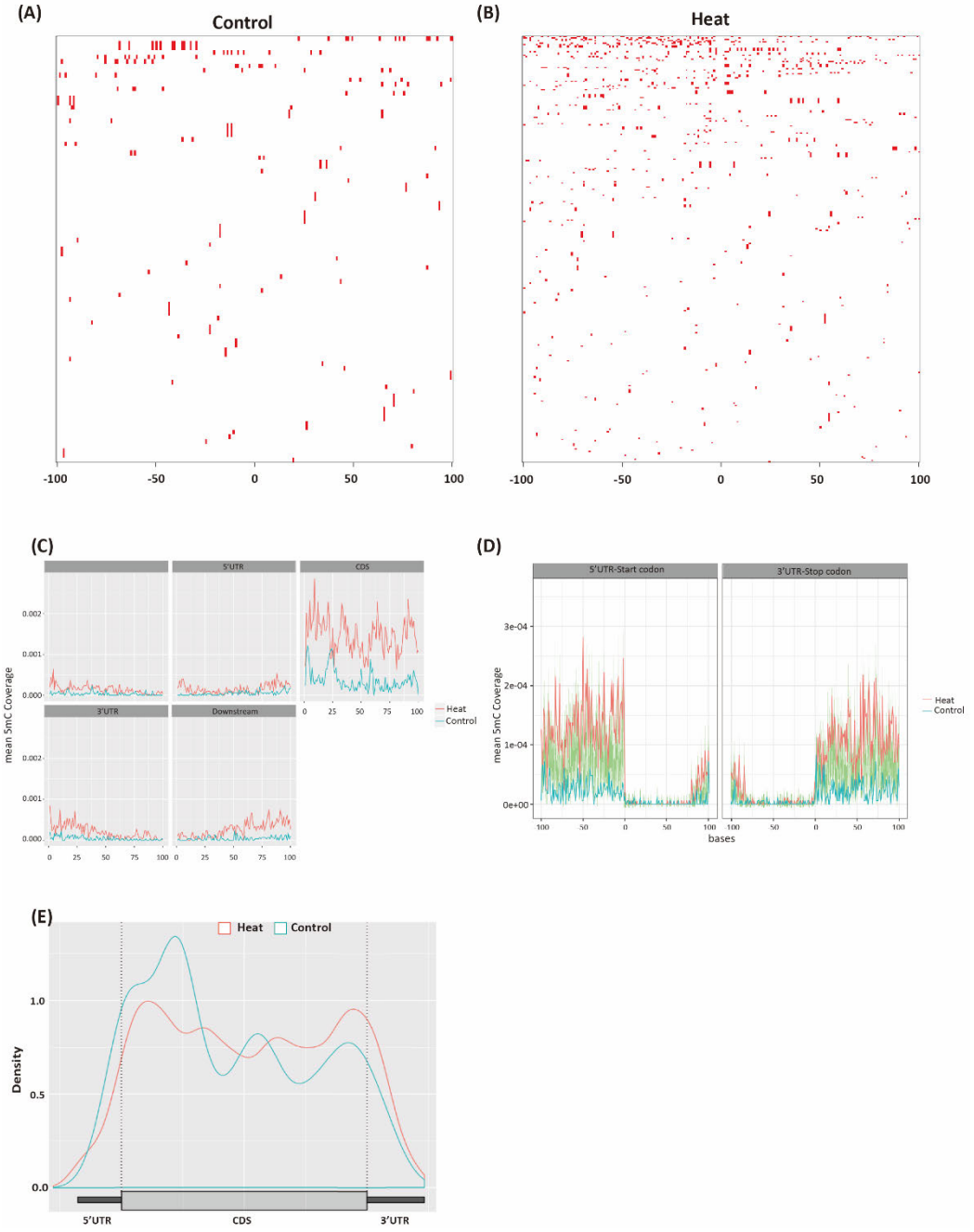
(D)

Gene name	TAIR ID	Log2 change
HsfA1d	AT1G32330	0.438
HsfA1e	AT3G02990	2.343
HsfA2	AT2G26150	12.532
HsfA3	AT5G03720	0.950
HsfA6b	AT3G22830	0.372
HsfA7a	AT3G51910	7.027
HsfA7b	AT3G63350	12.231
HsfA9	AT5G54070	0.655
HsfB1	AT4G36990	5.233
HsfB2a	AT5G62020	3.511
HsfB2b	AT4G11660	5.064
HsfB3	AT2G41690	0.760
HsfB4	AT1G46264	0.116
Hsp14.2-P(r)	AT5G47600	1.759
Hsp15.4-CI(r)	AT4G21870	0.909
Hsp15.7-CI(r)	AT5G37670	10.567
Hsp17.4-CI	AT3G46230	11.516
Hsp17.4-CIII	AT1G54050	9.525
Hsp17.6A-CI	AT1G59860	11.871
Hsp17.6B-CI	AT2G29500	9.771
Hsp17.6-CII	AT5G12020	12.473
Hsp17.6C-CI	AT1G53540	12.616
Hsp17.7-CII	AT5G12030	11.221
Hsp18.1-CI	AT5G59720	9.693
Hsp18.5-CI(r)	AT2G19310	6.133
Hsp22.0-ER	AT4G10250	11.187
Hsp23.5-M	AT5G51440	9.115
Hsp23.6-M	AT4G25200	13.920

Gene name	TAIR ID	Log2 change
Hsp25.4-P	AT4G27670	13.388
Hsp26.5-P(r)	AT1G52560	8.187
Hsp70-1	AT5G02500	2.662
Hsp70-2	AT5G02490	3.550
Hsp70-3	AT3G09440	6.246
Hsp70-4	AT3G12580	7.909
Hsp70-5	AT1G16030	10.289
Hsp70-6	AT4G24280	2.126
Hsp70-7	AT5G49910	2.019
Hsp70-8	AT2G32120	8.623
Hsp70-9	AT4G37910	0.233
Hsp70-10	AT5G09590	7.005
Hsp70-15	AT1G79920	5.255
Hsp70-17	AT4G16660	1.279
Hsp90-1	AT5G52640	8.820
Hsp90-2	AT5G56030	5.918
Hsp90-4	AT5G56000	1.452
Hsp90-5	AT2G04030	2.062
Hsp90-6	AT3G07770	2.394
Hsp90-7	AT4G24190	0.416
Hsp100-1	AT1G74310	10.619
Hsp100-3	AT5G15450	5.708
Hsp100-4	AT2G25140	6.235
Hsp100-5	AT5G50920	0.435
Hsp100-7	AT5G51070	0.431
HSP101	AT1G74310	10.619
HSA32	AT4G21320	5.943
PILS6	AT5G01990	2.925

Supplementary Figure 2.1: RNA transcript before and heat treatment.

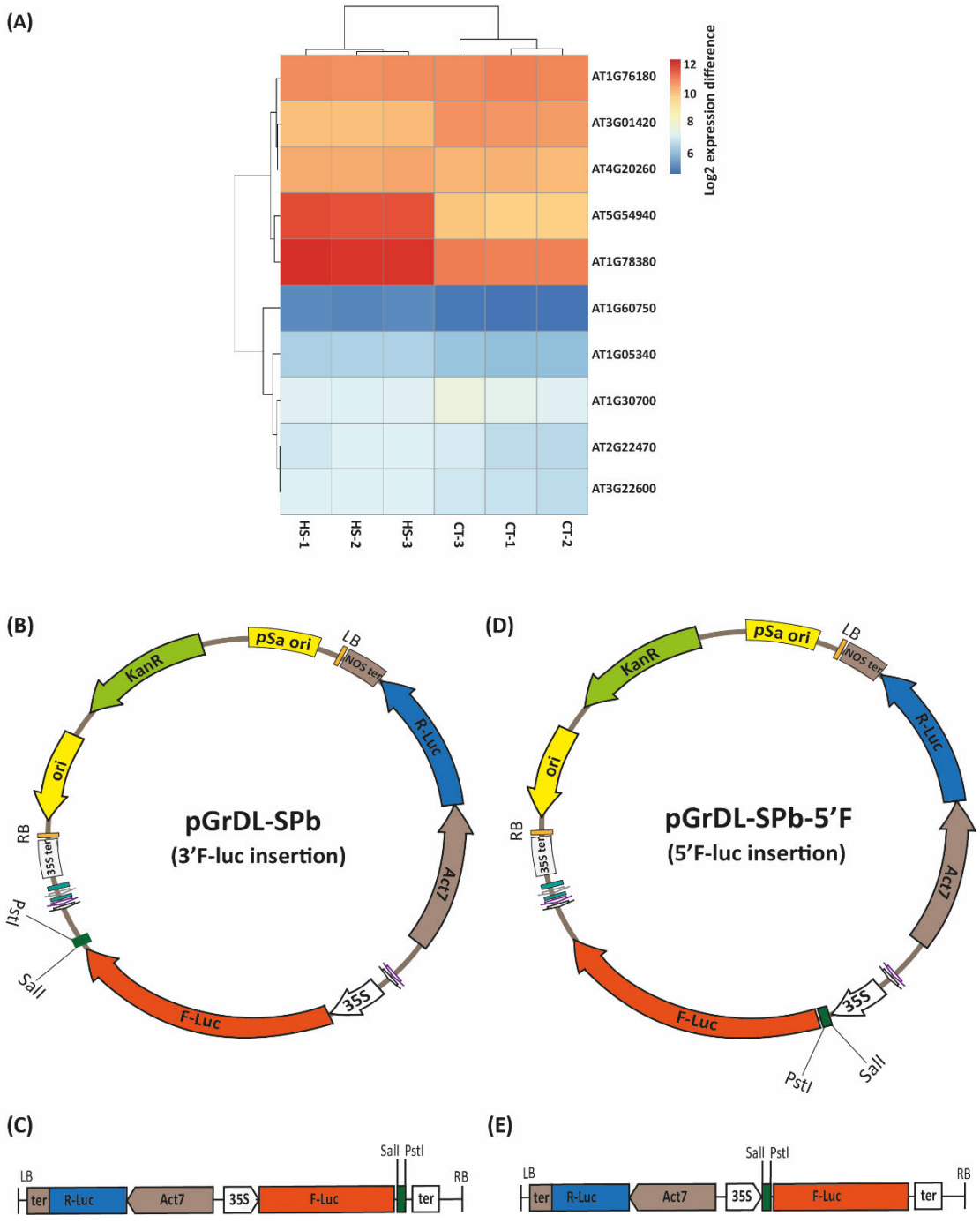
(A) Principal Component Analysis (PCA) of transcriptome-wide expression. Red dots: heat treatment samples; blue dots: control samples; X axis (PC1): variance distance between the heat treatment group and the control group; Y axis (PC2): variance distance among replicates within two groups. **(B)** Canonical Correlation Analysis (CCA) of transcriptome-wide expression among samples. X-axis & Y-axis: the score of each replicate among the control and heat treatments. **(C)** MA-plot analysis of 56 heat-related genes: each red dot represents one gene; X-axis (A-values): the mean expression (n=3) of the gene in log2 scale in heat treatments; Y-axis (M-values): the log2 fold change of gene expression before and after heat treatment; log2 fold change>0 means up-regulated expression after heat treatments, log2 fold change>0.5 means significantly up-regulated expression after heat treatment (P value <0.05). **(D)** 56 heat-response genes and log2 fold expression change after heat treatment.



Supplementary Figure 2.2: RNA 5mC distribution before and after heat treatment.

(A) Heat map of mRNA 5mC sites in control treatments. X-axis: the distance distribution on the transcript. **(B)** Heat map of mRNA 5mC sites after heat treatment. X-axis: the distance distribution on the transcript. **(C)** Spatial profile of 5mC coverage on different regions of mRNA after heat treatments and under control conditions. X-axis: the distance of the cytosine bases on the region, 0: start point, 100: endpoint; Y-axis: the mean

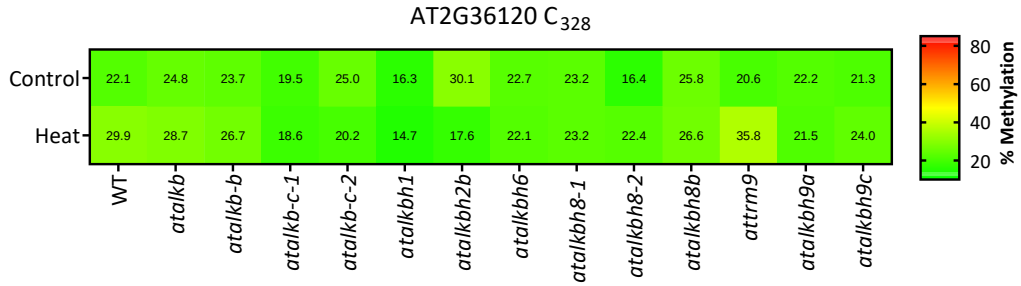
coverage of 5mC (5mC/(5mC+C)). **(D)** Spatial profile of 5mC coverage near start codon and stop codon of mRNA in heat treatments and control conditions. X-axis: the distance of C bases on the region, -100-0: 3'UTR/5'UTR, 0-100: start codon/stop codon; Y axis: the mean coverage of 5mC (5mC/(5mC+C)). **(E)** Spatial profile of 5mC density on mRNA in heat treatments and control conditions. X-axis: the mRNA region; Y axis: the density of 5mC (5mC base/5mC total *100)



Supplementary Figure 2.3: mRNA regions of interest and construction of F-Luc sensor vectors

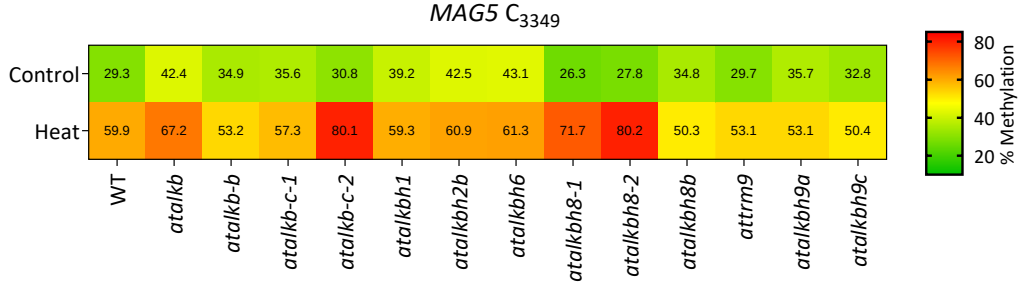
(A) The expression level of ten candidate genes containing RNA 5mC sites before and after heat treatment. The heat map shows the gene expression level of each sample in a log2. HS= heat treated replicates; CT= control replicates. **(B)** Schematic of pGrDL_SPb 3' vector, with *SalI* and *PstI* restriction sites at 3'UTR of *F-Luc* for target sequence insertion. **(C)** Schematic of the expression region of pGrDL_SPb 3' plasmid for target sequence insertion. **(D)** Schematics of pGrDL-SPb-5'F vector, with *SalI* and *PstI* restriction sites at 5'UTR of *F-*

Luc for target sequence insertion. **(E)** Schematic of the expression region of pGrDL_SPb 5' plasmid for target sequence insertion.



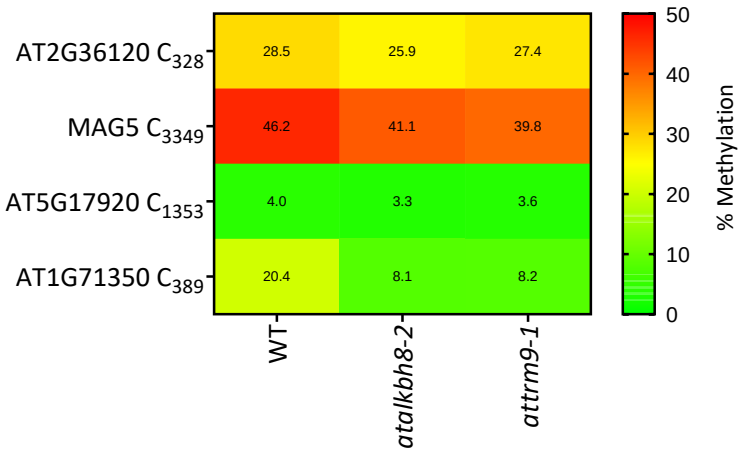
Supplementary Figure 3.1: The 5mC levels of *atalkb-like* mutants at *AT2G36120* C_{328}

The heat map shows the 5mC level of WT and *atalkb-like* mutants in percentage at *AT2G36120* C_{328} after control and heat shock treatment. Y-axis: the treatment conditions; X-axis: sample names.



Supplementary Figure 3.2: The 5mC levels in *atalkb-like* mutants at *MAG5* C_{3349}

The heat map shows the 5mC level of WT and *atalkb-like* mutants in percentage at *MAG5* C_{3349} after control and heat shock treatment. Y-axis: the treatment conditions; X-axis: sample names.



Supplementary Figure 3.3: The 5mC level of *atalkbh8* and *attrm9* at four 5mC sites

The heat map shows the 5mC level of WT and *atalkb-like* mutants in percentage at *AT2G36120* C_{328}, *MAG5* C_{3349}, *AT5G17920* C_{1353} and *AT1G71350* C_{389} after control treatment. Y-axis: the position of the 5mC site; X-axis: sample names.

5.2. Supplementary Tables

Supplementary Table 2.1: Oligonucleotide primers used for rt-PCR of HSPs and PDF2

Name	Primer Sequence	
HSP101	F	GTTCGACCCCCTTTCACATGAC
	R	CATCGATTTCCTCACGCACAAC
HSP20	F	GGC CCA CGT GTT TAA AGC TGA C
	R	ATCTTC ACATTCTCCGGCAACC
PDF2	F	GGGCAATGCAGCATATAGTTC
	R	TGGGTCTTCACTTAGCTCCAC

Supplementary Table 2.2: Candidate fragments with up-regulated 5mC sites after heat shock treatment

TAIR ID	refPos Chr (strand)	locus mRNA	5mC level (control)	5mC level (heat)	Fragment No.	Sequence
AT1G05340	1,559,674 (-)	5'UTR	0.00	0.98	1	acattcttgatagatacaaaaaaCatttttCagaCaCaaaatCataaaatCtttgccttttagtagatCaaagttcttta
	1,559,685 (-)		0.00	1.00		
	1,559,691 (-)		0.00	1.00		
	1,559,705 (-)		0.00	1.00		
	1,559,707 (-)		0.00	1.00		
	1,559,711 (-)		0.00	1.00		
	1,559,718 (-)		0.00	1.00		
	1,559,631 (-)	Exon 1	0.00	0.87	2	atgagCcagtacgatCacaaccagtctgcag
	1,559,641 (-)		0.00	0.93		
	1,559,427 (-)	Exon 2	0.00	0.38	3	gagCtaaCcCacCgCcacCgatgtCtacCtgtaCatCaccaccaCCgccgattggttaccCgaCtaaccaaCcgagtCatggttCggtagCtcagggtaaagtggaaaccaagtctaagggtgacggattcttcaaaggctg
	1,559,433 (-)		0.00	0.36		
	1,559,440 (-)		0.00	0.39		
	1,559,446 (-)		0.00	0.38		
	1,559,454 (-)		0.00	0.42		
	1,559,457 (-)		0.00	0.40		
	1,559,472 (-)		0.00	0.52		
	1,559,473 (-)		0.00	0.49		
	1,559,481 (-)		0.00	0.65		
	1,559,484 (-)		0.00	0.56		
	1,559,489 (-)		0.00	0.52		
	1,559,493 (-)		0.00	0.49		
	1,559,499 (-)		0.33	0.68		
	1,559,503 (-)		0.00	0.60		
	1,559,505 (-)		0.00	0.66		
	1,559,508 (-)		0.00	0.76		
	1,559,510 (-)		0.36	0.77		
	1,559,514 (-)		0.00	0.77		
	1,559,085 (-)	Exon 3	0.00	0.36	4	tcttgCggccatgtgttgctgttgtgcCCtggacatttgCttctaa
	1,559,096 (-)		0.00	0.36		
	1,559,097 (-)		0.00	0.34		
	1,559,119 (-)		0.00	0.28		
AT1G30700	10,894,302 (+)	Exon 2	0.00	0.84	5	gtttttcaattaccgtgatgttgatttgggtattaattctcataatggtaaaatCagtagttatgtggaaggtaaacgttactgcag
AT1G60750	22,363,347 (-)	Exon 2	0.00	0.91	6	aaggattgcttgtgaagcaagcttaaagcgtcttggtgtaaCctgcattgatctttattaccagcatcgaattgataccact
AT3G01420	160,023 (-)	Exon 9	0.00	0.23	7	gggacttatggcagagaagaaaatcaaaggttttgctatcagtgagactgctttttacattttcctcatcatggccaCaag
AT4G20260	10,940,805 (+)	5' UTR	0.00	0.09	8	gtttggagaagaagaagaacagatcaaatacgaggagagatctCtaaagagatttatcgtttCaagtaagtctctttatcaa
	10,940,824 (+)		0.00	0.08		
	10,941,628 (+)	Exon 2	0.00	0.09	9	aaagatgggttactggaattccaaggttgttccaaaattCaagaagttattcgagaaaaatagtgctaagaaggctgctgc
	10,943,279 (+)	Exon 5	0.00	0.11	10	agatggtatgattacttctcttgtatgaaaacatctttgtaCgtaacaaaaaatgaaaggaagaaacggaaaggaaca
AT1G76180	28,587,701 (-)	5' UTR	0.07	0.20	11	ccatcacctcttattttcatctgcttttgaatttCaaaCCttCaCataaaaaaaatactttgaatccgtgtttcattcat
	28,587,714 (-)		0.05	0.23		
	28,587,716 (-)		0.06	0.22		
	28,587,719 (-)		0.07	0.22		
	28,587,720 (-)		0.06	0.24		
	28,587,724 (-)		0.15	0.61		
AT2G22470	9,538,496 (-)	Exon 1	0.30	0.72	12	atgctcctgctcccggtccttccgctggattaactccaacttcgtCtccagctCcaggaccagaCggtgctgctgatgcacca
	9,538,507 (-)		0.35	0.80		
	9,538,515 (-)		0.35	0.77		
AT3G22600	8,007,207 (-)	Exon 1	0.35	0.73	13	ccgtgtctcaactacataactggaaactctacCtCtcctaatcagcaatgctgtaatcagttgagtagagtagtCcagtcttCtcctgactgtttgtgtcaagtccttaacggtgg
	8,007,215 (-)		0.40	0.86		
	8,007,255 (-)		0.34	0.70		
	8,007,257 (-)		0.32	0.70		

Supplementary Table 2.3: Candidate fragments with down-regulated 5mC sites after heat shock treatment

TAIR ID	refPos Chr (strand)	locus mRNA	5mC level (control)	5mC level (heat)	Fragment No.	Sequence
AT5G54940	22,308,682 (-)	Exon 3	0.05	0.00	14	atggttgagctagaCattCagattCcatCagCataCgatCCatttgCagaagCtaaagattCagatgcaCCaggagCtaaagagaacattcacattcgaatccagcagaggaatgggaaa
	22,308,688 (-)		0.06	0.00		
	22,308,689 (-)		0.07	0.00		
	22,308,697 (-)		0.06	0.00		
	22,308,706 (-)		0.06	0.00		
	22,308,712 (-)		0.07	0.00		
	22,308,718 (-)		0.08	0.00		
	22,308,719 (-)		0.08	0.00		
	22,308,723 (-)		0.09	0.00		
	22,308,727 (-)		0.08	0.00		
	22,308,730 (-)		0.08	0.00		
	22,308,734 (-)		0.08	0.00		
	22,308,740 (-)		0.10	0.00		
	22,308,744 (-)		0.11	0.00		
	22,308,759 (-)		0.11	0.00		
	22,308,767 (-)		0.12	0.00		
	22,308,768 (-)		0.12	0.00		
AT1G78380	29,487,568 (-)	Exon 1	0.07	0.00	15	ggttcagtacattgacgaggtctggtctcacaagaaccctatCcttccttCtgatccttacctgagagctcaagctag
	29,487,560 (-)		0.07	0.00		
	29,486,975 (-)	Exon 2	0.06	0.00	16	cgtggctaagtccttgcctgatcctgagaaggttactgaattcgtctctgagCtcaggaagaaatttgtacCtgagtaa
	29,486,685 (-)		0.06	0.00		
	29,486,666 (-)		0.06	0.00		
	29,486,634 (-)	3'UTR	0.05	0.00	17	tctcttgtcatatggttCggacttaagatgctctctgttttaCtttcgtgtgttgtCtttCttaaataatgcagagcatctgtgaggtggttgagtgtagagcatc
	29,486,609 (-)		0.05	0.00		
	29,486,595 (-)		0.07	0.00		
	29,486,591 (-)		0.06	0.00		

Supplementary Table 2.4: Control fragments

TAIR ID	locus mRNA	Sequence	Fragment No.
AT3G22260	Exon 1	aaatcaaacataacaaaaacCaacaaaacaaactcaaaaggattcttcttcaaacacaacatcttcaagaaaacat	18
AT1G13320	Exon 2	tgtctatggttgatgagcctttatacccgattgctgtgcttatcgacgagctaaaaaacgatgatattcagcgtagattgaac	19

Supplementary Table 2.5: Oligonucleotides primers used to anneal and produce candidate sensor fragments

Fragment Name	Gene Name & Position	Oligo Name	Oligo Sequence
1	AT1G05340-5'UTR	F	tcgacacattcttgatagatacaaaaaaCatttttCagaCaCaaaatCataaaatCtttgcctttagtagatCaaagttctttacctgca
		R	ggtaaagaactttGatctactaaaggcaaaGattttatGattttGtGtctGaaaaatGttttttgtatctatcaagaatgtg
2	AT1G05340-Exon 1	F	tcgacatgagCcagtacgatCacaaccagtctgcagctgca/78
		R	gctgcagactggttgtGatcgtactgGctcatg
4	AT1G05340-Exon 3	F	tcgactcttgCggccatgtgttgctgttgtgcCCtggacatttgCttctaactgca
		R	gttagaaGcaaatgtccaGGgcacaacagcaacacatggccGcaagag
5	AT1G30700-Exon 2	F	tcgacttttcaattaccgtgatgttgatttgggtattaattctcataatggtaaaatCagtagttatgtggaaggtaaacgttactgca
		R	gtaacgtttaccttccacataactactGattttaccattatgagaattaatacccaaatcaacatcacggtaattgaaaaag
6	AT1G60750-Exon 2	F	tcgacggattgcttgtgaagcaagcttaaagcgtcttggtgtaaCctgcattgatctttattaccagcatcgaattgataccactctgca
		R	gagtggtatcaattcgatgctggtaataaagatcaatgcagGttacaccaagacgctttaagcttgcttcacaagcaatccg
7	AT3G01420-Exon 9	F	tcgacggacttatggcagagaagaaaatcaaaggttttgctatcagtgagactgctttttacattttcctcatcatggccaCaagctgca
		R	gcttGtggccatgatgaggaaaatgtaaaaagcagtctcactgatagcaaaacctttgattttcttctctgccataagtccg
8	AT4G20260-5UTR	F	tcgacttggagaagaagaagaacagatcaaatacgaggagagatctCtaaagagatttatcgtttCaagtaagtctctttatcaactgca
		R	gttgataaagagacttacttGaaacgataaatctctttaGagatctctcctcgtatttgatctgttcttctcttctccaag
9	AT4G20260-Exon 2	F	tcgacaaagatgggttactggaattccaaggttgttccaaaattCaagaagttattcgagaaaaatagtgctaagaaggctgctgctgca
		R	gcagcagccttcttagcactatttttctcgaataacttcttGaattttggaacaaccttggaattccagtaacccatctttg
10	AT4G20260-3'UTR	F	tcgacagatggtatgattacttctcttgtatgaaaacatctttgtaCgtaacaaaaaaatgaaaggaagaaacggaaaggaacactgca
		R	gtgttcctttccgtttcttcctttcatttttttgttacGtacaaagatgttttcatacaagagaagtaatcataccatctg
11	AT1G76180-5'UTR	F	tcgaccatcacctcttattttcatctgcttttgaatttCaaaCCttCaCataaaaaaaaatacttttgaatccgtgtttcattcatctgca
		R	gatgaatgaaacacggattcaaaagtatttttttttatGtGaaGGtttGaaattcaaaagcagatgaaaataagaggtgatgg
12	AT2G22470-Exon 1	F	tcgacctcctgctcccggtccttccgctggattaactccaacttcgtCtccagctCcaggaccagaCggtgctgctgatgcaccactgca
		R	gtggtgcatcagcagcaccGtctggtcctgGagctggaGacgaagttggagttaatccagcggaaggaccgggagcaggagg
13	AT3G22600-3'UTR	F	tcgacaaatcaaacataacaaaaacCaacaaaacaaactcaaaaggattcttcttcaaacacaacatcttcaagaaaacatctgca
		R	gatgttttcttgaagatgttgtgtttgaagaagaatccttttgagtttgttttgttGgtttttgttatgtttgatttg
15	AT1G78380-Exon 1	F	tcgacggttcagtacattgacgaggtctggtctcacaagaaccctatCcttccttCtgatccttacctgagagctcaagctagctgca
		R	gctagcttgagctctcaggtaaggatcaGaaggaagGatagggttcttgtgagaccagacctcgtcaatgtactgaaccg
16	AT1G78380-Exon 2	F	tcgaccgtggctaagtccttgcctgatcctgagaaggttactgaattcgtctctgagCtcaggaagaaatttgtacCtgagtaactgca
		R	gttactcaGgtacaaatttcttcctgaGctcagagacgaattcagtaaccttctcaggatcaggcaaggacttagccacgg
17	AT1G78380-3'UTR	F	tcgactctcttgtcatatggttCggacttaagatgctctctgttttaCtttcgtgtgttgtCtttCttaaatatgcagagcatcctgca
		R	ggatgctctgcattatttaaGaaaGacaacacacgaaaGtaaaacagagagcatcttaagtccGaaccatatgacaagagag
19	AT1G13320-Exon 2	F	tcgactgtctatggttgatgagcctttatacccgattgctgtgcttatcgacgagctaaaaaacgatgatattcagcgtagattgctgca
		R	gcaatctacgctgaatatcatcgtttttagctcgtcgataagcacagcaatcgggtataaaggctcatcaaccatagacag

Supplementary Table 2.6: Oligonucleotide primers used to produce 5mC-mutated sensor fragments

Fragment Name	Gene Name & Position	Oligo Name	Oligo Sequence
mutated-1	AT1G05340-5'UTR	F	tcgacacattcttgatagatacaaaaaagatttttgagagagaaaatgataaaatgtttgccttttagtagatgaaagttctttacctgca
		R	ggtaaagaactttcatctactaaaggcaaacattttatcattttctctctcaaaaatctttttgtatctatcaagaatgtg
mutated-2	AT1G05340-Exon 1	F	tcgacatgaggcagtacgatgacaaccagtctgcagctgca
		R	gctgcagactggttgtcatcgtactgcctcatg
mutated-4	AT1G05340-Exon 3	F	tcgactcttggggccatgtgttgctgttgtgcggtggacatttggttctaactgca
		R	gttagaaccaaatgtccaccgcacaacagcaacacatggccccaagag
mutated-5	AT1G30700-Exon 2	F	tcgacttttcaattaccgtgatgttgatttgggtattaattctcataatggtaaaatgagtagttatgtggaaggtaaacgttactgca
		R	gtaacgtttaccttccacataactactcattttaccattatgagaattaatacccaaatcaacatcacggtaattgaaaaag
mutated-6	AT1G60750-Exon 2	F	tcgacggattgcttgtgaagcaagcttaaagcgtcttggtgtaagctgcattgatctttattaccagcatcgaattgataccactctgca
		R	gagtggtatcaattcgatgctggtaataaagatcaatgcagcttacaccaagacgctttaagcttgcttcacaagcaatccg
mutated-7	AT3G01420-Exon 9	F	tcgacggacttatggcagagaagaaaatcaaaggttttgctatcagtgagactgcttttacattttcctcatcatggccagaagctgca
		R	gcttctggccatgatgaggaaaatgtaaaaagcagtctcactgatagcaaaaccttgattttcttctctgccataagtccg
mutated-8	AT4G20260-5UTR	F	tcgacttggagaagaagaagaacagatcaaatacgaggagagatctgtaaagagatttatcgtttgaagtaagtctctttatcaactgca
		R	gttgataaagagacttacttcaaacgataaatctctttacagatctctcctcgtatttgatctgttcttcttcttctccaag
mutated-9	AT4G20260-Exon 2	F	tcgacaaagatgggttactggaattccaaggttgttccaaaattgaagaagttattcgagaaaaatagtgctaagaaggctgctgctgca
		R	gcagcagccttcttagcactatttttctcgaataacttcttcaattttggaacaaccttggaattccagtaaccccatctttg
mutated-10	AT4G20260-3'UTR	F	tcgacagatggtatgattacttctcttgtatgaaaacatctttgtaggtaacaaaaaaatgaaaggaagaaacggaaaggaacactgca
		R	gtgttcctttccgtttcttcctttcatttttttgttacctacaaagatgttttcatacaagagaagtaatcatacccatctg
mutated-11	AT1G76180-5'UTR	F	tcgaccatcacctcttattttcatctgcttttgaatttgaaaggttgagataaaaaaaatacttttgaatccgtgtttcattcatctgca
		R	gatgaatgaaacacggattcaaaagtatttttttttatctcaacctttcaaattcaaaagcagatgaaaataagaggtgatgg
mutated-12	AT2G22470-3'UTR	F	tcgacctcctgctcccggtccttccgctggattaactccaacttcgtgtccagctgcaggaccagagggtgctgctgatgcaccactgca
		R	gtggtgcatcagcagcaccctctggtcctgcagctggacacgaagttggagttaatccagcggaaggaccgggagcaggagg
mutated-15	AT1G78380-Exon 1	F	tcgacggttcagtacattgacgaggtctggtctcacaagaaccctatgcttccttgtgatccttacctgagagctcaagctagctgca
		R	gctagcttgagctctcaggtaaggatcacaaggaagcatagggttcttgtgagaccagacctcgtcaatgtactgaaccg
mutated-16	AT1G78380-Exon 2	F	tcgaccgtggctaagtccttgcctgatcctgagaaggttactgaattcgtctctgaggtcaggaagaaatttgtacgtgagtaactgca
		R	gttactcacgtacaaatttcttcctgacctcagagacgaattcagtaaccttctcaggatcaggcaaggacttagccacgg
mutated-17	AT1G78380-3'UTR	F	tcgactctcttgtcatatggttgggacttaagatgctctctgttttagtttcgtgtgttgtgtttgttaaataatgcagagcatcctgca
		R	ggatgctctgcattatttaacaaacacaacacacgaaactaaaacagagagcatcttaagtcccaaccatatgacaagagag
mutated-18	AT3G22600-3'UTR	F	tcgacaaatcaaacataacaaaaacgaacaaaacaaactcaaaaggattcttcttcaaacacaacatctcaagaaaacatctgca
		R	gatgttttcttgaagatgttgtgtttgaagaagaatccttttgagtttgttttgttcgtttttgttatgtttgatttg
mutated-19	AT1G13320-Exon 2	F	tcgactgtctatggttgatgagcctttatacccgattgctgtgcttatcgacgagctaaaaaacgatgatattcagcgtagattgctgca
		R	gcaatctacgctgaatatcatcgtttttttagctcgtcgataagcacagcaatcgggtataaaggctcatcaaccatagacag

Supplementary Table 2.7: dsDNA used to synthesise candidate fragments

Fragment Name	Gene Name & Position	Sequence
3	AT1G05340-Exon 2	gagCtaaCcCacCgCcacCgatgtCtacCtgtaCatCaccaccaCCgccgattggttaccCgaCtaaccaaCcga gtCatggttCggtagCtcagggtaaagtggaaaccaagtctaagggtgacggattcttcaaaggctg
13	AT3G22600-Exon 1	ccgtgtctcaactacataactggaaactctacCtCtcctaatcagcaatgctgtaatcagttgagtagagtagtCcagt cttCtcctgactgtttgtgtcaagtccttaacggtgg
14	AT5G54940-Exon 3	atggttgagctagaCattCagattCcatCagCataCgatCCatttgCagaagCtaaagattCagatgcaCCagga gCtaaagagaacattcacattcgaatccagcagaggaatgggaaa
mutated-3	AT1G05340-Exon 2	gaggtaagcgacgggcacggatgtgtacgtgtagatgaccaccagggccgattggttaccggagtaaccaagcgagt gatggttgggtaggtcagggtaaagtggaaaccaagtctaagggtgacggattcttcaaaggctg
mutated-13	AT3G22600-Exon 1	ccgtgtctcaactacataactggaaactctacgtgtcctaatcagcaatgctgtaatcagttgagtagagtagtgcagtc ttgtcctgactgtttgtgtcaagtccttaacggtgg
mutated-14	AT5G54940-Exon 3	atggttgagctagagattgagattgcatgaggataggatggatttggagaaggtaaagattgagatgcaggaggagg taaagagaacattcacattcgaatccagcagaggaatgggaaa

Supplementary Table 2.8: Oligonucleotide primers and dsDNA used to construct pGrDL-SPb-5' plasmid

Fragment	Sequence	
Mu-SalIPstI	Oligo -F	aattccttggcgacttccatgca
	Oligo -R	tggaagtcgccaag
In-SalIPstI	aaggggtaccgggcccaccccctactccaaaaatgtcaaagatacagtctcagaagaccaaagggctattgagacttttcaacaaagggtaatttcggg aaacctcctcggattccattgcccagctatctgtcacttcatcgaaaggacagtagaaaaggaaggtggctcctacaaatgccatcattgcgataaaggaa aggctatcattcaagatgcctctgccgacagtggtcccaaagatggacccccacccacgaggagcatcgtggaaaaagaagacgttccaaccacgtcttc aaagcaagtggattgatgtgacatctccactgacgtaagggatgacgcacaatcccactatccttcgcaagacccttcctctatataaggaagtcatttcat ttggagaggacctcgtcgactgatgccctgcaggagatggaagacgccaaaaacataaagaaaggcccggcgccaaga	

Supplementary Table 2.9: Oligonucleotide primers used to verify the recombinant plasmids

Primer name	Primer location	Sequence
3'_F-Luc_PCR_LP	F-Luc of SPb	GATCCTCATAAAGGCCAAGAAG
3'_F-Luc_PCR_RP	Down-stream F-Luc of SPb	CCCTTATCGGGAAACTACTCAC
3'_F-Luc_Seq_R	Down-stream of F-Luc	CAAGAGTAAAAGATAGTAAAACCGG
5'_F-Luc_Seq_R	F-Luc	GCCCATATCGTTTCATAGCTTC

Supplementary Table 3.1: Oligonucleotide primers used to screen T-DNA mutants

Name	Sequence (5'-3')	Name	Sequence (5'-3')
SAIL_223_C05_LP	TTGCCTTCGTCTTCTCAAATG	SAIL_223_C05_RP	GGGATGGCTAGAAAACTGACC
SAIL_68_G03_LP	CCCTAAATCAAGCAAATCGAG	SAIL_68_G03_RP	TCACTCACAAACGAAATTTTCAG
WiscDsLox469B10_LP	AAACCAGAGATTCTGGGTTCG	WiscDsLox469B10_RP	CTATTCCAACAGCCTGCAGG
SAIL_148_H12_LP	CGACGAGCTATGATGAGAGATG	SAIL_148_H12_RP	TTTTCAAAGGCGATGAATCAC
SAIL_1146_G04_LP	AAACCAGAGATTCTGGGTTCG	SAIL_1146_G04_RP	TTTCAGGAACTCACAGGCAAC
SAIL_32_F08_LP	AACCAAGTGGGAAGAAGAAGC	SAIL_32_F08_RP	TCAAAGAAATCGAATTGGAAATG
SAIL_60_C05_LP	AGTGGCAAATGCATCAAACTC	SAIL_60_C0_RP	ATTTAGGCGGTGTCCGTAAAC
SAIL_1239_H02_LP	TTCCGACAAAAGTGTGATTCC	SAIL_1239_H02_RP	CAAACTCATCCTCCGCTACAC
WiscDsLoxHs085_11E_LP	TGAACCGTACAAAACCTGTGC	WiscDsLoxHs085_11E_RP	CAGAAGACGGCTGAGAAAATG
WiscDsLoxHs038_10B_LP	TCAAATCACCAATGAGGAAGC	WiscDsLoxHs038_10B_RP	ATGGTTTTTCTAGAGGGAGCG
SALK_007964_LP	TCTTTTGAGGAAGCTTCGATG	SALK_007964_RP	TAGCAAAAACAGAAGGCAACG
SALK_102868_LP	TCCGCCTAAGAGGAAAATAGC	SALK_102868_RP	TACAGTTCGATTGGTCCAAGG
SK15155_LP	GCTCATCCTTCTCAAAAATGC	SK15155_RP	AATGCTCAGTCTCAGAGACCG
SALKseq_052340.3_LP	CAATGGTCATGCCTATGTTCC	SALKseq_052340.3_RP	GCAGAATGGAGGGGTTTAGTC
SALKseq_062071.2_LP	ACTCGCCTAGCCTGTCGTTAC	SALKseq_062071.2_RP	AATAGCGCGCAACAAGTAAAG
SAIL_125_B01_LP	GGTCGATGACAAAATACGACG	SAIL_125_B01_RP	CTCAGATGCTTGAGAACGTCC
FLAG_143G11_LP	TACAACACAGACCGTGTGTGG	FLAG_143G11_RP	TCCTTGATGTTTCAGGACAAAG
SALK_035671_LP	ATGGAGAAGGACACTATCGGC	SALK_035671_RP	TGCCGAGAACTTGGTGTTAAG
SALK_021477_LP	GAAGACGATGCGTCTCTTGTC	SALK_021477_RP	TCTTCGTCTTAGAGCCACCAG
SAIL_193_C03_LP	CAACGATTTCTTTTCATTTTATTTTG	SAIL_193_C03_RP	AGATGCGGAGAGGATACTTCC
SALK_021775_LP	GCTTCTGGTTTTCTTTCATTGC	SALK_021775_RP	TAAGAACCGGAGAATTCACCAG
WiscDsLoxHs188_10A_LP	TCATCGAACAAGTTTTACACGG	WiscDsLoxHs188_10A_RP	GATTTGGAGGAGATTAAGCCG
SALK_111811_LP	AAAGCGAAACAAACCACAGAAC	SALK_111811_RP	ACCGGTCTGGTTCTACTTCAAG
WiscDsLox501D02_LP	ACTGGTGTTTTTAGCGCAGTG	WiscDsLox501D02_RP	AGAAAGTCGTGGTTGTCGATG
SALK_204823_LP	AGGAAGCAAAGAACGCATATTC	SALK_204823_RP	ACATGCCATTTAACCAACCTTC
SAIL_402_E12_LP	GCACAAGAACTTACCCAACAGG	SAIL_402_E12_RP	GGCAAAGGACGTTCAACTATTC
GABI_423_C08_LP	GCCAAGGAACATGTTAGAGGAC	GABI_423_C08_RP	GTTGAACAGAACGAAGGTCTCC
FLAG_156D06_LP	CGAGTCGAGTTGAGTTTGTCC	FLAG_156D06_RP	GTCGAGATGATGATGAAAGCC
SALK_004215_LP	GAAAGGCTGTGAGTATTCACCC	SALK_004215_RP	TTTGTGGCAGAACTACGAGAAG
SALK_107289_LP	CATGACCATTGGAGGACGTAG	SALK_107289_RP	TCAGGGTGAATACTCACAGCC
SALK_135308_LP	AATACTTCCCGTTACCGCAAC	SALK_135308_RP	CTTGCTCTGTGCTCATCGTTC
SALK_003181_LP	AAAACCGAAAGAACCGAAATC	SALK_003181_RP	TTAACAGGTTGGCATCTGGAG
SALK_094502_LP	CAAGGATGATTTCTCACCTGC	SALK_094502_RP	TGGGATTATCGCCTCATATTG
SALK_083838_LP	CAAGGATGATTTCTCACCTGC	SALK_083838_RP	TGGGATTATCGCCTCATATTG
SALK_024872_LP	CTGGCTATGACTGGTTATCTGC	SALK_024872_RP	TGAGATTTGGTGATCTCCCTTC
SAIL_756_A08_LP	TCACAATTTATCAAACGATCAACC	SAIL_756_A08_RP	ATGCCATGGATCTTTCAACAG
SALK_138864_LP	GTCTTGTTGATGCCTCAAAGC	SALK_138864_RP	GTCTTGTTGATGCCTCAAAGC
SALK_023985_LP	TTGTTATGGACTTCACTCCGC	SALK_023985_RP	ACGGCAAGTAAGTGAAACTCG
FLAG_564H09_LP	CCAATTGTTTAAGCATTTGACATC	FLAG_564H09_RP	AGCTGCGTTTACAGCAAAAAC
SALK _LB3	ATTTTGCCGATTTCGGAAC	WiscDs-LB	TCCTCGAGTTTCTCCATAATAATGT
SAIL _LB3	TAGCATCTGAATTTCATAACCAATCTC GATACAC	WiscDsLoxHs-LB	TGATCCATGTAGATTTCCCGGACATG C
FLAG _LB1	CTACAAATTGCCTTTTCTTATCGAC	GABI_LB1	ATATTGACCATCATACTCATTGC
pSKTAIL(SK)-LB1	TTCTCATCTAAGCCCCCATTTGG		
RT-AT2G22270-F	GTGCAACACAATCAGAAGTT	RT-AT2G22270-R	GAGCTTTGCATCTTCTCTGA
RT-AT4G39860-F	GCAAGAAGCTAGAGGTAACA	RT-AT4G39860-R	CATTTCACGGAGTTTAGCTG
RT-AT2G22260-F	ATACAAAGGTGCCAGTGATT	RT-AT2G22260-R	GCTGCTTCTCTTTAACCTCT
RT-AT1G11780-F	TCTGATTCAAGTGCACCAAT	RT-AT1G11780-R	TATGGAAGCTGACTGGAGTA
RT-AT3G14140-F	CTTGGTCTCCATCAGGTTAG	RT-AT3G14140-R	AAGATTAACGTGTCTGCCTT
RT-AT3G14160-F	AGGGAGGAGATGAAATACCA	RT-AT3G14160-R	ATCCCTCTGATCACCATACA
RT-AT5G01780-F	CAGAAACTGGAAGACTTGGT	RT-AT5G01780-R	ATCACTCCTTGAGCTTCTTC
RT-AT4G36090-F	CTGTCGGATCAGTTCTAGTG	RT-AT4G36090-R	CCTACGATACTTGCTTCCTC
RT-AT2G17970-F	CTGATGTAGCTAAGCACTGT	RT-AT2G17970-R	TGAGTCATAACCATTGCCTC
RT-AT1G48980-F	GTTGAGGAAACCGGTGAATA	RT-AT1G48980-R	GGCCATTTTGACTCATCCA
RT-AT2G48080-F	AACCCTTGTCCAAAGTGTTA	RT-AT2G48080-R	CTTAAGTGGGAGAGTGAGTG
RT-AT4G02940-F	GAACAACCAATCTCCACTCT	RT-AT4G02940-R	GTCAGGTCGAATACGAAAGA
RT-AT1G36310-F	GAGTACGAGAGAGCAGAAAG	RT-AT1G36310-R	CCTTCGCTAAAGACATGGTA
RT-AT1G31600-F	ACATGGAAAGCCTCAACTAC	RT-AT1G31600-R	ACATTTGTTGCTGAGAGTCA
RT-AT4G02485-F	AATCTGTTGATCTTCCGGTC	RT-AT4G02485-R	TCCACAGCTTCATATTCGTT
RT-AT4G20350-F	GCCTCAAAGCTTACTCATCT	RT-AT4G20350-R	TACGGACTTTGGGAACTAGA

Supplementary Table 3.2: Oligonucleotide primers used to amplify BS-amplicon regions

Name	C-T converted RT primers		C-T converted PCR primers	
MAG5	F	ACACTGACGACATGGTTCTACA	F	TAAAGAGATGGGTGGAAAGAGG
C3349	R	TACGGTAGCAGAGACTTGGTCT	R	CAACTCCCACTAAATAAAAACATC
AT2G36120	F	ACACTGACGACATGGTTCTACA	F	GGGTATGGTGAGGGATATATTGG
C328	R	TACGGTAGCAGAGACTTGGTCT	R	CTCCACCACCACCTACTC
AT5G17920	F	ACACTGACGACATGGTTCTACA	F	TAATGTTGTGGTTTTGGTTTTAAG
C1353	R	TACGGTAGCAGAGACTTGGTCT	R	CTCACAACAAACTCTCCTAAACTC
AT1G71350	F	ACACTGACGACATGGTTCTACA	F	TATGTTGAAAGGAGGTGAG
C389	R	TACGGTAGCAGAGACTTGGTCT	R	CTACTTTACAAACTTCTTCACTAC

Supplementary Table 3.3: C-T converted sequences of the RBS-amplicon-seq

>gi|ref|AT2G36120|C328|
TGGGGGTGGAGGTGGTGGATATGGTGGTGGTGTTGGTGGAGGAGGAGGTGGTGGCTTTGGAG GAGGCTATGGAGGTGGAAGTGGTGAAGGTGGTGGAGTTGGATATGGAGGTGGAGAAGTTGGT GGGTATGGTGGAGGTGGAGGAGGTG
>gi|ref|MAG5|C3349|
TGTGGAATTT**C**TAGTTGAGGAAGTTGTATTATTTTTTTTTTTAATGTTAGTTTTATTTTGAAGTAATT TATTGGGTTATGAAAATAAGTTTGAGATAAAGAATGATA
>gi|ref|AT5G17920|C1353|
GAGATTTTTTTCAAGAGTTATTAATGAGGGTGTTTAGAAGGTTGTTGTTGTTTTGAAGGGATTTGA TTATTGTTGTGTAATTAATGTTAGTGTTAGGTTAGATGTTTAGTAGAAGAAGTTTAATTTTTTAATT TTATTAATTATAATTATTGGATTTTTTTTATGGATTGTA
>gi|ref|AT1G71350|C389|
GTTTTATGTTATATTATAGGTGGTGTAGATTTGATGTTTTTAGGTATTTTGATTTTTTTTCAAGGTTA TTTTTTTTTTTTTGTTGGAGAAATATGGGTTGTTAAAGTTTTTGGAAATTTAGTATTTATTGTGGTT GGGTGTATTATAATGA

Supplementary Table 3.4: RBS-amplicon-seq result of *atalkb-like* mutants after control and heat shock treatments

Name	T-DNA line	5mC site	Treatment	Replicates	Total reads	C account	T account	methyl %
WT		AT2G36120 C328	Control	1	47533	8315	38773	17.66
				2	39854	10467	28933	26.57
			Heat	1	37076	9236	27454	25.17
				2	39861	13662	25885	34.55
		MAG5 C3349	Control	1	34573	7805	26747	22.59
				2	34526	13078	21404	37.93
			Heat	1	33246	19220	13989	57.88
				2	25933	16047	9879	61.90
atalkb	Flag143G11	AT2G36120 C328	Control	1	26329	6510	19560	24.97
				2	54510	13343	40689	24.69
			Heat	1	38299	12869	25001	33.98
				2	39677	9202	30109	23.41
		MAG5 C3349	Control	1	46423	21611	24678	46.69
				2	75171	28906	46224	38.47
			Heat	1	34746	21593	13110	62.22
				2	78504	56934	21465	72.62
atalkb-b	SK15155	AT2G36120 C328	Control	1	47514	11759	35755	24.70
				2	80919	18369	62550	22.70
			Heat	1	41939	8861	33078	21.10
				2	23370	7530	15840	32.20
		MAG5 C3349	Control	1	147019	40858	106161	27.80
				2	196054	85983	110071	43.90
			Heat	1	260278	135754	124524	52.20
				2	215892	117093	98799	54.20
atalkb-c-1	SAIL_193_C03	AT2G36120 C328	Control	1	31691	5143	26260	16.38
				2	45512	10189	34869	22.61
			Heat	1	32475	5149	27004	16.01
				2	38210	8003	29819	21.16
		MAG5 C3349	Control	1	59734	18882	40802	31.64
				2	47168	14081	33018	29.90
			Heat	1	39351	34574	4700	88.03
				2	77504	56439	20990	72.89
atalkb-c-2	SALK_021477C	AT2G36120 C328	Control	1	32472	11577	20544	36.04
				2	21333	2966	18215	14.00
			Heat	1	68632	13758	54100	20.27
				2	10356	2069	8180	20.19
		MAG5 C3349	Control	1	96616	32387	64140	33.55
				2	48576	18341	30144	37.83
			Heat	1	56642	30904	25675	54.62
				2	69010	41458	27471	60.15
atalkbh1	SALK_007964	AT2G36120 C328	Control	1	7624	327	7231	4.33
				2	17774	4953	12611	28.20
			Heat	1	24837	4025	20594	16.35
				2	55053	7146	47387	13.10
		MAG5 C3349	Control	1	38265	15102	23111	39.52
				2	53184	20689	32436	38.94
			Heat	1	31423	21791	9601	69.42
				2	100210	50755	49366	50.69
atalkbh2b	SAIL_68_G03	AT2G36120 C328	Control	1	6442	2479	3905	38.83
				2	33244	7052	25848	21.43
			Heat	1	7034	829	6119	11.93
				2	25134	5814	19096	23.34
		MAG5 C3349	Control	1	28546	13278	15215	46.60
				2	177121	68624	108370	38.77
			Heat	1	96917	72206	24589	74.60
				2	87218	43343	43803	49.74

Continued Supplementary Table 3.4

Name	T-DNA line	5mC site	Treatment	Replicates	Total reads	C account	T account	methyl %
atalkbh6	SALK_138864	AT2G36120 C328	Control	1	89485	20545	68060	23.19
				2	80907	17791	62291	22.22
			Heat	1	50384	11180	38683	22.42
				2	64825	13968	50187	21.77
		MAG5 C3349	Control	1	37594	16381	21180	43.61
				2	47436	20160	27205	42.56
			Heat	1	50515	30638	19822	60.72
				2	66246	41002	25178	61.96
atalkbh8-1	SALK_094502C	AT2G36120 C328	Control	1	49844	7634	41705	15.47
				2	64527	19714	44202	30.84
			Heat	1	94574	21750	71971	23.21
				2	58495	13411	44482	23.17
		MAG5 C3349	Control	1	66215	18519	47632	28.00
				2	50032	12361	37642	24.72
			Heat	1	50649	40371	10229	79.78
				2	83296	53546	29638	64.37
atalkbh8-2	SALK_083838C	AT2G36120 C328	Control	1	30451	5579	24596	18.49
				2	68917	9823	58590	14.36
			Heat	1	21257	3394	17704	16.09
				2	4775	1350	3367	28.62
		MAG5 C3349	Control	1	35784	13719	22028	38.38
				2	58123	11687	46401	20.12
			Heat	1	57863	46592	11163	80.67
				2	18013	14352	3640	79.77
atalkbh8b	SALK_024872	AT2G36120 C328	Control	1	14606	11759	35755	21.30
				2	23833	7125	16409	30.28
			Heat	1	60380	15874	43861	26.57
				2	60831	15991	44240	26.55
		MAG5 C3349	Control	1	54769	22666	32055	41.62
				2	65595	19168	46333	29.26
			Heat	1	27828	16498	11302	59.35
				2	84457	35938	48450	42.59
attrm9	SALK_135308C	AT2G36120 C328	Control	1	11503	1612	9817	14.10
				2	3924	1053	2839	27.06
			Heat	1	2719	1033	1673	38.17
				2	11189	3687	7362	33.37
		MAG5 C3349	Control	1	106138	24077	81972	22.70
				2	49953	19333	30585	38.73
			Heat	1	58885	34072	24757	57.92
				2	83116	40469	42574	48.73
atalkbh9a	SALK_204823	AT2G36120 C328	Control	1	10758	8260	2498	23.22
				2	20561	16223	4338	21.10
			Heat	1	17187	13298	3889	22.63
				2	16587	13203	3384	20.40
		MAG5 C3349	Control	1	26951	17408	9543	35.41
				2	17635	11272	6363	36.08
			Heat	1	34786	18197	16589	47.69
				2	20563	8410	12153	59.10
atalkbh9C	SALK_021775	AT2G36120 C328	Control	1	6230	4730	1500	24.07
				2	10771	8779	1992	18.49
			Heat	1	26989	20242	6747	25.00
				2	16977	13072	3905	23.00
		MAG5 C3349	Control	1	66597	43475	23122	34.72
				2	55471	38236	17235	31.07
			Heat	1	88754	39851	48903	55.10
				2	47532	25605	21927	46.13

Supplementary Table 3.5: Repeat RBS-amplicon-seq result of *atalkb8* and *attrm9* after control treatments

Name	T-DNA line	5mC site	Replicates	Total reads	C account	T account	methyl %
WT	-	AT5G17920 C1353	1	2769	99	2668	3.58
			2	2759	125	2631	3.99
			3	2410	96	2309	3.99
			4	2235	89	2142	3.99
			5	2301	93	2207	4.04
		MAG5 C3349	1	4389	2031	2358	46.27
			2	6874	3159	3714	31.25
			3	4256	1330	2926	49.57
			4	4882	2420	2462	47.91
			5	4310	2065	2245	45.96
		AT1G71350 C389	1	862	6	856	0.70
			2	1934	520	1414	13.47
			3	1151	155	996	22.95
			4	1222	280	940	22.95
			5	492	87	405	17.68
		AT2G36120 C328	1	29841	8484	21280	28.50
			2	5931	2407	3508	40.69
			3	22315	4928	17327	22.14
			4	16399	4664	11697	28.51
			5	34200	7919	26192	23.22
atalkbh8-2	SALK_083838C	AT5G17920 C1353	1	1924	50	1873	2.60
			2	3003	84	2917	2.80
			3	2582	107	2472	4.15
			4	2264	74	2188	3.27
			5	2033	91	1940	4.48
		MAG5 C3349	1	2945	1172	1773	39.80
			2	2972	1221	1751	41.08
			3	3930	1763	2167	44.86
			4	205	131	74	63.90
			5	3543	1148	2895	28.39
		AT1G71350 C389	1	371	29	342	7.82
			2	300	25	275	8.33
			3	664	57	607	8.58
			4	245	16	229	6.53
			5	79	26	53	32.91
		AT2G36120 C328	1	32842	6463	26321	19.71
			2	14367	2563	11777	17.87
			3	40561	16492	23980	40.75
			4	26905	6952	19889	25.90
			5	36336	10521	26742	28.23
attrm9	SALK_135308C	AT5G17920 C1353	1	1924	50	1873	2.60
			2	2424	108	2313	4.46
			3	2582	107	2472	4.15
			4	3696	134	3561	3.63
			5	8700	272	8423	3.13
		MAG5 C3349	1	2945	1172	1773	39.80
			2	4641	1717	2923	37.00
			3	3930	1763	2167	44.86
			4	4507	1995	2512	44.26
			5	17720	6695	11024	37.78
		AT1G71350 C389	1	371	29	342	7.82
			2	543	106	437	19.52
			3	664	57	607	8.58
			4	175	6	169	3.43
			5	1297	3	1294	0.23
		AT2G36120 C328	1	32842	6463	26321	19.71
			2	32385	11246	21049	34.82
			3	40561	16492	23980	40.75
			4	19336	5279	14020	27.35
			5	69560	14945	54463	21.53

6. Reference

Abbasi-Moheb, L, Mertel, S, Gonsior, M, Nouri-Vahid, L, Kahrizi, K, Cirak, S, Wieczorek, D, Motazacker, MM, Esmaeeli-Nieh, S, Cremer, K, Weißmann, R, Tzschach, A, Garshasbi, M, Abedini, SS, Najmabadi, H, Ropers, HH, Sigrist, SJ & Kuss, AW 2012, 'Mutations in NSUN2 cause autosomal-recessive intellectual disability', *Am J Hum Genet*, vol. 90, no. 5, May 4, pp. 847-855.

Adhikary, H, Bakos, O & Biggar, KK 2019, 'The Role of Protein Lysine Methylation in the Regulation of Protein Function: Looking Beyond the Histone Code', in *The DNA, RNA, and Histone Methylomes*, pp. 453-477.

Agarwal, M, Katiyar-Agarwal, S, Sahi, C, Gallie, DR & Grover, A 2001, 'Arabidopsis thaliana Hsp100 proteins: kith and kin', *Cell Stress & Chaperones*, vol. 6, no. 3, p. 219.

Agris, PF, Sierzputowska-Gracz, H & Smith, C 1986, 'Transfer RNA contains sites of localized positive charge: carbon NMR studies of 13C methyl-enriched Escherichia coli and yeast tRNAPhe', *Biochemistry*, vol. 25, no. 18, pp. 5126-5131.

Agris, PF, Vendeix, FA & Graham, WD 2007, 'tRNA's wobble decoding of the genome: 40 years of modification', *J Mol Biol*, vol. 366, no. 1, Feb 9, pp. 1-13.

Aik, W, Scotti, JS, Choi, H, Gong, L, Demetriades, M, Schofield, CJ & McDonough, MA 2014, 'Structure of human RNA N(6)-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation', *Nucleic Acids Res*, vol. 42, no. 7, Apr, pp. 4741-4754.

Alberts, B, Johnson, A, Lewis, J, Raff, M, Roberts, K & Walter, P 2003, 'Molecular biology of the cell', *SCANDINAVIAN JOURNAL OF RHEUMATOLOGY*, vol. 32, no. 2, pp. 125-125.

Alexandrov, A, Chernyakov, I, Gu, W, Hiley, SL, Hughes, TR, Grayhack, EJ & Phizicky, EM 2006, 'Rapid tRNA decay can result from lack of nonessential modifications', *Molecular cell*, vol. 21, no. 1, pp. 87-96.

Amort, T, Rieder, D, Wille, A, Khokhlova-Cubberley, D, Riml, C, Trixl, L, Jia, X-Y, Micura, R & Lusser, A 2017, 'Distinct 5-methylcytosine profiles in poly (A) RNA from mouse embryonic stem cells and brain', *Genome Biology*, vol. 18, no. 1, p. 1.

Amort, T, Soulière, MF, Wille, A, Jia, X-Y, Fiegl, H, Wörle, H, Micura, R & Lusser, A 2013, 'Long non-coding RNAs as targets for cytosine methylation', *RNA biology*, vol. 10, no. 6, pp. 1002-1008.

An, J, Rao, A & Ko, M 2017, 'TET family dioxygenases and DNA demethylation in stem cells and cancers', *Exp Mol Med*, vol. 49, no. 4, Apr 28, p. e323.

Atkinson, NJ & Urwin, PE 2012, 'The interaction of plant biotic and abiotic stresses: from genes to the field', *Journal of Experimental Botany*, vol. 63, no. 10, pp. 3523-3543.

Barik, S 1993, 'The structure of the 5′ terminal cap of the respiratory syncytial virus mRNA', *Journal of general virology*, vol. 74, no. 3, pp. 485-490.

Baxter-Roshek, JL, Petrov, AN & Dinman, JD 2007, 'Optimization of ribosome structure and function by rRNA base modification', *PloS one*, vol. 2, no. 1, p. e174.

Begum, T, Reuter, R & Schöffl, F 2013, 'Overexpression of AtHsfB4 induces specific effects on root development of Arabidopsis', *Mechanisms of development*, vol. 130, no. 1, pp. 54-60.

Berka, M, Kopecká, R, Berková, V, Brzobohatý, B & Černý, M 2022, 'Regulation of Heat Shock Proteins 70 and Their Role in Plant Immunity', *Journal of Experimental Botany*.

Birkedal, U, Christensen-Dalsgaard, M, Krogh, N, Sabarinathan, R, Gorodkin, J & Nielsen, H 2015, 'Profiling of Ribose Methylations in RNA by High-Throughput Sequencing', *Angewandte Chemie International Edition*, vol. 54, no. 2, pp. 451-455.

Björk, GR & Hagervall, TG 2014, 'Transfer RNA Modification: Presence, Synthesis, and Function', *EcoSal Plus*, vol. 6, no. 1, 2014/05//.

Björk, GR, Jacobsson, K, Nilsson, K, Johansson, MJO, Byström, AS & Persson, OP 2001, 'A primordial tRNA modification required for the evolution of life?', *The EMBO Journal*, vol. 20, no. 1-2, pp. 231-239.

Blanco, S, Bandiera, R, Popis, M, Hussain, S, Lombard, P, Aleksic, J, Sajini, A, Tanna, H, Cortés-Garrido, R, Gkatza, N, Dietmann, S & Frye, M 2016, 'Stem cell function and stress response are controlled by protein synthesis', *Nature*, vol. 534, no. 7607, 2016/06/01, pp. 335-340.

Blanco, S, Dietmann, S, Flores, JV, Hussain, S, Kutter, C, Humphreys, P, Lukk, M, Lombard, P, Treps, L & Popis, M 2014, 'Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders', *The EMBO Journal*, vol. 33, no. 18, pp. 2020-2039.

Blanco, S & Frye, M 2014, 'Role of RNA methyltransferases in tissue renewal and pathology', *Current opinion in cell biology*, vol. 31, pp. 1-7.

Boccaletto, P, Stefaniak, F, Ray, A, Cappannini, A, Mukherjee, S, Purta, E, Kurkowska, M, Shirvanizadeh, N, Destefanis, E, Groza, P, Avşar, G, Romitelli, A, Pir, P, Dassi, E, Conticello, SG, Aguilo, F & Bujnicki, JM 2021, 'MODOMICS: a database of RNA modification pathways. 2021 update', *Nucleic Acids Research*, vol. 50, no. D1, pp. D231-D235.

Bohnsack, KE, Höbartner, C & Bohnsack, MT 2019, 'Eukaryotic 5-methylcytosine (m5C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease', *Genes*, vol. 10, no. 2, p. 102.

Bujnicki, JM, Feder, M, Ayres, CL & Redman, KL 2004, 'Sequence–structure–function studies of tRNA: m5C methyltransferase Trm4p and its relationship to DNA: m5C and RNA: m5U methyltransferases', *Nucleic Acids Research*, vol. 32, no. 8, pp. 2453-2463.

Burgess, AL, David, R & Searle, IR 2015, 'Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae', *BMC Plant Biol*, vol. 15, p. 199.

Camper, S, Albers, R, Coward, J & Rottman, F 1984, 'Effect of undermethylation on mRNA cytoplasmic appearance and half-life', *Molecular and cellular biology*, vol. 4, no. 3, pp. 538-543.

Cao, G, Li, H-B, Yin, Z & Flavell, RA 2016, 'Recent advances in dynamic m6A RNA modification', *Open biology*, vol. 6, no. 4, p. 160003.

Carroll, S, Narayan, P & Rottman, F 1990, 'N6-methyladenosine residues in an intron-specific region of prolactin pre-mRNA', *Molecular and cellular biology*, vol. 10, no. 9, pp. 4456-4465.

Chan, CT, Pang, YLJ, Deng, W, Babu, IR, Dyavaiah, M, Begley, TJ & Dedon, PC 2012, 'Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins', *Nature communications*, vol. 3, no. 1, pp. 1-9.

Chen, Zhang, L, Zheng, G, Fu, Y, Ji, Q, Liu, F, Chen, H & He, C 2014, 'Crystal structure of the RNA demethylase ALKBH5 from zebrafish', *FEBS letters*, vol. 588, no. 6, pp. 892-898.

Chen, X, Li, A, Sun, B-F, Yang, Y, Han, Y-N, Yuan, X, Chen, R-X, Wei, W-S, Liu, Y & Gao, C-C 2019, '5-methylcytosine promotes pathogenesis of bladder cancer through stabilizing mRNAs', *Nature cell biology*, vol. 21, no. 8, pp. 978-990.

Chen, Y, Sierzputowska-Gracz, H, Guenther, R, Everett, K & Agris, PF 1993, '5-Methylcytidine is required for cooperative binding of magnesium (2+) and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA', *Biochemistry*, vol. 32, no. 38, pp. 10249-10253.

Chmielowska-Bąk, J, Arasimowicz-Jelonek, M & Deckert, J 2019, 'In search of the mRNA modification landscape in plants', *BMC Plant Biology*, vol. 19, no. 1, pp. 1-8.

Chong, LP, Wang, Y, Gad, N, Anderson, N, Shah, B & Zhao, R 2015, 'A highly charged region in the middle domain of plant endoplasmic reticulum (ER)-localized heat-shock protein 90 is required for resistance to tunicamycin or high calcium-induced ER stresses', *Journal of Experimental Botany*, vol. 66, no. 1, pp. 113-124.

Choudhury, FK, Rivero, RM, Blumwald, E & Mittler, R 2017, 'Reactive oxygen species, abiotic stress and stress combination', *The Plant Journal*, vol. 90, no. 5, pp. 856-867.

Cowling, VH 2010, 'Regulation of mRNA cap methylation', *Biochemical Journal*, vol. 425, no. 2, pp. 295-302.

Cui, X, Liang, Z, Shen, L, Zhang, Q, Bao, S, Geng, Y, Zhang, B, Leo, V, Vardy, LA, Lu, T, Gu, X & Yu, H 2017, '5-Methylcytosine RNA Methylation in Arabidopsis Thaliana', *Mol Plant*, vol. 10, no. 11, Nov 6, pp. 1387-1399.

Dalluge, JJ, Hashizume, T & McCloskey, JA 1996, 'Quantitative Measurement of Dihydrouridine in RNA Using Psotope Dilution Liquid Chromatography-Mass Spectrometry (LC/MS)', *Nucleic Acids Research*, vol. 24, no. 16, pp. 3242-3245.

Dalluge, JJ, Hashizume, T, Sopchik, AE, McCloskey, JA & Davis, DR 1996, 'Conformational Flexibility in RNA: The Role of Dihydrouridine', *Nucleic Acids Research*, vol. 24, no. 6, pp. 1073-1079.

David, R, Burgess, A, Parker, B, Li, J, Pulsford, K, Sibbritt, T, Preiss, T & Searle, IR 2017, 'Transcriptome-wide mapping of RNA 5-methylcytosine in Arabidopsis mRNAs and noncoding RNAs', *The Plant Cell*, vol. 29, no. 3, pp. 445-460.

Decatur, WA & Fournier, MJ 2002, 'rRNA modifications and ribosome function', *Trends in biochemical sciences*, vol. 27, no. 7, pp. 344-351.

Delatte, B, Wang, F, Ngoc, LV, Collignon, E, Bonvin, E, Deplus, R, Calonne, E, Hassabi, B, Putmans, P & Awe, S 2016, 'Transcriptome-wide distribution and function of RNA hydroxymethylcytosine', *Science*, vol. 351, no. 6270, pp. 282-285.

Dominissini, D, Moshitch-Moshkovitz, S, Schwartz, S, Salmon-Divon, M, Ungar, L, Osenberg, S, Cesarkas, K, Jacob-Hirsch, J, Amariglio, N & Kupiec, M 2012, 'Topology of the human and mouse m 6 A RNA methylomes revealed by m 6 A-seq', *Nature*, vol. 485, no. 7397, pp. 201-206.

Dominissini, D, Nachtergaele, S, Moshitch-Moshkovitz, S, Peer, E, Kol, N, Ben-Haim, MS, Dai, Q, Di Segni, A, Salmon-Divon, M & Clark, WC 2016, 'The dynamic N 1-methyladenosine methylome in eukaryotic messenger RNA', *Nature*, vol. 530, no. 7591, pp. 441-446.

Driedonks, N, Xu, J, Peters, JL, Park, S & Rieu, I 2015, 'Multi-level interactions between heat shock factors, heat shock proteins, and the redox system regulate acclimation to heat', *Frontiers in plant science*, vol. 6, p. 999.

Duan, H-C, Wei, L-H, Zhang, C, Wang, Y, Chen, L, Lu, Z, Chen, PR, He, C & Jia, G 2017, 'ALKBH10B is an RNA N 6-methyladenosine demethylase affecting Arabidopsis floral transition', *The Plant Cell*, vol. 29, no. 12, pp. 2995-3011.

Dubin, DT & Taylor, RH 1975, 'The methylation state of poly A-containing-messenger RNA from cultured hamster cells', *Nucleic Acids Research*, vol. 2, no. 10, pp. 1653-1668.

Durant, PC, Bajji, AC, Sundaram, M, Kumar, RK & Davis, DR 2005, 'Structural effects of hypermodified nucleosides in the Escherichia coli and human tRNALys anticodon loop: the effect of nucleosides s2U, mcm5U, mcm5s2U, mnm5s2U, t6A, and ms2t6A', *Biochemistry*, vol. 44, no. 22, pp. 8078-8089.

Ebrahimi, M, Hosseinkhani, S, Heydari, A, Khavari-Nejad, RA & Akbari, J 2012, 'Improvement of Thermostability and Activity of Firefly Luciferase Through [TMG][Ac] Ionic Liquid Mediator', *Applied Biochemistry and Biotechnology*, vol. 168, no. 3, 2012/10/01, pp. 604-615.

Echevarría-Zomeño, S, Fernández-Calvino, L, Castro-Sanz, AB, López, JA, Vázquez, J & Castellano, MM 2016, 'Dissecting the proteome dynamics of the early heat stress response leading to plant survival or death in Arabidopsis', *Plant, cell & environment*, vol. 39, no. 6, pp. 1264-1278.

Edelheit, S, Schwartz, S, Mumbach, MR, Wurtzel, O & Sorek, R 2013, 'Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs', *PLoS genetics*, vol. 9, no. 6.

Elkon, R, Ugalde, AP & Agami, R 2013, 'Alternative cleavage and polyadenylation: extent, regulation and function', *Nature Reviews Genetics*, vol. 14, no. 7, pp. 496-506.

Fahlberg, P, Buhot, N, Johansson, ON & Andersson, MX 2019, 'Involvement of lipid transfer proteins in resistance against a non-host powdery mildew in Arabidopsis thaliana', *Molecular plant pathology*, vol. 20, no. 1, pp. 69-77.

Fang, L, Wang, W, Li, G, Zhang, L, Li, J, Gan, D, Yang, J, Tang, Y, Ding, Z, Zhang, M, Zhang, W, Deng, D, Song, Z, Zhu, Q, Cui, H, Hu, Y & Chen, W 2020, 'CIGAR-seq, a CRISPR/Cas-based method for unbiased screening of novel mRNA modification regulators', *Mol Syst Biol*, vol. 16, no. 11, Nov, p. e10025.

Fleming, AM, Alenko, A, Kitt, JP, Orendt, AM, Flynn, PF, Harris, JM & Burrows, CJ 2019, 'Structural elucidation of bisulfite adducts to pseudouridine that result in deletion signatures during reverse transcription of RNA', *Journal of the American Chemical Society*, vol. 141, no. 41, pp. 16450-16460.

Flores, JV, Cordero-Espinoza, L, Oeztuerk-Winder, F, Andersson-Rolf, A, Selmi, T, Blanco, S, Tailor, J, Dietmann, S & Frye, M 2017, 'Cytosine-5 RNA Methylation Regulates Neural Stem Cell Differentiation and Motility', *Stem Cell Reports*, vol. 8, no. 1, Jan 10, pp. 112-124.

Foster, PG, Nunes, CR, Greene, P, Moustakas, D & Stroud, RM 2003, 'The first structure of an RNA m 5 C methyltransferase, Fmu, provides insight into catalytic mechanism and specific binding of RNA substrate', *Structure*, vol. 11, no. 12, pp. 1609-1620.

Franco-Zorrilla, JM, Valli, A, Todesco, M, Mateos, I, Puga, MI, Rubio-Somoza, I, Leyva, A, Weigel, D, García, JA & Paz-Ares, J 2007, 'Target mimicry provides a new mechanism for regulation of microRNA activity', *Nature genetics*, vol. 39, no. 8, pp. 1033-1037.

Frye, M & Watt, FM 2006, 'The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors', *Current biology*, vol. 16, no. 10, pp. 971-981.

Fu, L, Guerrero, CR, Zhong, N, Amato, NJ, Liu, Y, Liu, S, Cai, Q, Ji, D, Jin, SG, Niedernhofer, LJ, Pfeifer, GP, Xu, GL & Wang, Y 2014, 'Tet-mediated formation of 5-hydroxymethylcytosine in RNA', *J Am Chem Soc*, vol. 136, no. 33, Aug 20, pp. 11582-11585.

Fu, Y, Jia, G, Pang, X, Wang, RN, Wang, X, Li, CJ, Smemo, S, Dai, Q, Bailey, KA, Nobrega, MA, Han, KL, Cui, Q & He, C 2013, 'FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA', *Nat Commun*, vol. 4, p. 1798.

Gao, H-N, Jiang, H, Lian, X-Y, Cui, J-Y, You, C-X, Hao, Y-J & Li, Y-Y 2021, 'Identification and Functional Analysis of the MdLTPG Gene Family in Apple', *Plant Physiology and Biochemistry*.

Gigova, A, Duggimpudi, S, Pollex, T, Schaefer, M & Koš, M 2014, 'A cluster of methylations in the domain IV of 25S rRNA is required for ribosome stability', *RNA*, vol. 20, no. 10, pp. 1632-1644.

Gilbert, WV, Bell, TA & Schaening, C 2016, 'Messenger RNA modifications: Form, distribution, and function', *Science*, vol. 352, no. 6292, pp. 1408-1412.

Goll, MG, Kirpekar, F, Maggert, KA, Yoder, JA, Hsieh, C-L, Zhang, X, Golic, KG, Jacobsen, SE & Bestor, TH 2006, 'Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2', *Science*, vol. 311, no. 5759, pp. 395-398.

Goujon, M, McWilliam, H, Li, W, Valentin, F, Squizzato, S, Paern, J & Lopez, R 2010, 'A new bioinformatics analysis tools framework at EMBL-EBI', *Nucleic Acids Research*, vol. 38, no. 2, pp. W695-W699.

Guan, Q, Lu, X, Zeng, H, Zhang, Y & Zhu, J 2013, 'Heat stress induction of miR398 triggers a regulatory loop that is critical for thermotolerance in Arabidopsis', *The Plant Journal*, vol. 74, no. 5, pp. 840-851.

Haag, S, Sloan, KE, Ranjan, N, Warda, AS, Kretschmer, J, Blessing, C, Hübner, B, Seikowski, J, Dennerlein, S & Rehling, P 2016, 'NSUN3 and ABH1 modify the wobble position of mt-tRNAMet to expand codon recognition in mitochondrial translation', *The EMBO Journal*, vol. 35, no. 19, pp. 2104-2119.

Haag, S, Warda, AS, Kretschmer, J, Günnigmann, MA, Höbartner, C & Bohnsack, MT 2015, 'NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs', *RNA*, vol. 21, no. 9, Sep, pp. 1532-1543.

Hasanuzzaman, M, Nahar, K, Alam, MM, Roychowdhury, R & Fujita, M 2013, 'Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants', *Int J Mol Sci*, vol. 14, no. 5, May 3, pp. 9643-9684.

He, S, Zhang, Y, Wang, J, Wang, Y, Ji, F, Sun, L, Zhang, G & Hao, F 2022, 'H3K4me2, H4K5ac and DNA methylation function in short- and long-term heat stress responses through affecting the expression of the stress-related genes in G. hirsutum', *Environmental and Experimental Botany*, vol. 194.

He, XJ, Chen, T & Zhu, JK 2011, 'Regulation and function of DNA methylation in plants and animals', *Cell Res*, vol. 21, no. 3, Mar, pp. 442-465.

He, Y-F, Li, B-Z, Li, Z, Liu, P, Wang, Y, Tang, Q, Ding, J, Jia, Y, Chen, Z & Li, L 2011, 'Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA', *Science*, vol. 333, no. 6047, pp. 1303-1307.

Helm, M 2006, 'Post-transcriptional nucleotide modification and alternative folding of RNA', *Nucleic Acids Research*, vol. 34, no. 2, pp. 721-733.

Hon, GC, Song, CX, Du, T, Jin, F, Selvaraj, S, Lee, AY, Yen, CA, Ye, Z, Mao, SQ, Wang, BA, Kuan, S, Edsall, LE, Zhao, BS, Xu, GL, He, C & Ren, B 2014, '5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation', *Mol Cell*, vol. 56, no. 2, Oct 23, pp. 286-297.

Hong, B, Brockenbrough, JS, Wu, P & Aris, JP 1997, 'Nop2p is required for pre-rRNA processing and 60S ribosome subunit synthesis in yeast', *Molecular and cellular biology*, vol. 17, no. 1, pp. 378-388.

Hou, Y-M, Gamper, H & Yang, W 2015, 'Post-transcriptional modifications to tRNA—a response to the genetic code degeneracy', *RNA*, vol. 21, no. 4, pp. 642-644.

Hu, L, Lu, J, Cheng, J, Rao, Q, Li, Z, Hou, H, Lou, Z, Zhang, L, Li, W, Gong, W, Liu, M, Sun, C, Yin, X, Li, J, Tan, X, Wang, P, Wang, Y, Fang, D, Cui, Q, Yang, P, He, C, Jiang, H, Luo, C & Xu, Y 2015, 'Structural insight into substrate preference for TET-mediated oxidation', *Nature*, vol. 527, no. 7576, Nov 5, pp. 118-122.

Hu, S, Ding, Y & Zhu, C 2020, 'Sensitivity and responses of chloroplasts to heat stress in plants', *Frontiers in plant science*, vol. 11, p. 375.

Hu, W, Hu, G & Han, B 2009, 'Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice', *Plant Science*, vol. 176, no. 4, pp. 583-590.

Huang, T, Chen, W, Liu, J, Gu, N & Zhang, R 2019, 'Genome-wide identification of mRNA 5-methylcytosine in mammals', *Nat Struct Mol Biol*, vol. 26, no. 5, May, pp. 380-388.

Huang, Y-C, Niu, C-Y, Yang, C-R & Jinn, T-L 2016, 'The heat stress factor HSFA6b connects ABA signaling and ABA-mediated heat responses', *Plant physiology*, vol. 172, no. 2, pp. 1182-1199.

Huang, Y, Chavez, L, Chang, X, Wang, X, Pastor, W, Kang, J, Zepeda-Martínez, J, Pape, U, Jacobsen, S, Peters, B & Rao, A 2014, 'Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, 01/28, pp. 1361-1366.

Huber, SM, van Delft, P, Mendil, L, Bachman, M, Smollett, K, Werner, F, Miska, EA & Balasubramanian, S 2015, 'Formation and abundance of 5-hydroxymethylcytosine in RNA', *Chembiochem*, vol. 16, no. 5, Mar 23, pp. 752-755.

Huong, TT, Ngoc, LNT & Kang, H 2020, 'Functional Characterization of a Putative RNA Demethylase ALKBH6 in Arabidopsis Growth and Abiotic Stress Responses', *Int J Mol Sci*, vol. 21, no. 18, Sep 13.

Hussain, S, Sajini, AA, Blanco, S, Dietmann, S, Lombard, P, Sugimoto, Y, Paramor, M, Gleeson, JG, Odom, DT & Ule, J 2013, 'NSun2-mediated cytosine-5

methylation of vault noncoding RNA determines its processing into regulatory small RNAs', *Cell reports*, vol. 4, no. 2, pp. 255-261.

Ito, S, Shen, L, Dai, Q, Wu, SC, Collins, LB, Swenberg, JA, He, C & Zhang, Y 2011, 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science*, vol. 333, no. 6047, pp. 1300-1303.

Iyer, LM, Tahiliani, M, Rao, A & Aravind, L 2009, 'Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids', *Cell cycle*, vol. 8, no. 11, pp. 1698-1710.

Jackman, JE & Alfonzo, JD 2013, 'Transfer RNA modifications: nature's combinatorial chemistry playground', *WIREs RNA*, vol. 4, no. 1, pp. 35-48.

Jaenisch, R & Bird, A 2003, 'Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals', *Nat Genet*, vol. 33 Suppl, Mar, pp. 245-254.

Jia, G, Fu, Y, Zhao, X, Dai, Q, Zheng, G, Yang, Y, Yi, C, Lindahl, T, Pan, T & Yang, Y-G 2011, 'N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO', *Nature chemical biology*, vol. 7, no. 12, pp. 885-887.

Joshi, JR, Singh, V & Friedman, H 2020, 'Arabidopsis cysteine-rich trans-membrane module (CYSTM) small proteins play a protective role mainly against heat and UV stresses', *Funct Plant Biol*, vol. 47, no. 3, Feb, pp. 195-202.

Jühling, T, Duchardt-Ferner, E, Bonin, S, Wöhnert, J, Pütz, J, Florentz, C, Betat, H, Sauter, C & Mörl, M 2018, 'Small but large enough: structural properties of

armless mitochondrial tRNAs from the nematode Romanomermis culicivorax', *Nucleic Acids Research*, vol. 46, no. 17, pp. 9170-9180.

Kaplan, F, Kopka, J, Haskell, DW, Zhao, W, Schiller, KC, Gatzke, N, Sung, DY & Guy, CL 2004, 'Exploring the temperature-stress metabolome of Arabidopsis', *Plant Physiol*, vol. 136, no. 4, Dec, pp. 4159-4168.

Katahira, J 2012, 'mRNA export and the TREX complex', *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1819, no. 6, pp. 507-513.

Khoddami, V & Cairns, BR 2013, 'Identification of direct targets and modified bases of RNA cytosine methyltransferases', *Nature biotechnology*, vol. 31, no. 5, pp. 458-464.

Khoddami, V, Yerra, A, Mosbruger, TL, Fleming, AM, Burrows, CJ & Cairns, BR 2019, 'Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution', *Proceedings of the National Academy of Sciences*, vol. 116, no. 14, pp. 6784-6789.

King, MY & Redman, KL 2002, 'RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine', *Biochemistry*, vol. 41, no. 37, pp. 11218-11225.

Ko, M, An, J, Bandukwala, HS, Chavez, L, Äijö, T, Pastor, WA, Segal, MF, Li, H, Koh, KP, Lähdesmäki, H, Hogan, PG, Aravind, L & Rao, A 2013, 'Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX', *Nature*, vol. 497, no. 7447, pp. 122-126.

Koffa, MD, Clements, JB, Izaurralde, E, Wadd, S, Wilson, SA, Mattaj, IW & Kuersten, S 2001, 'Herpes simplex virus ICP27 protein provides viral mRNAs with access to the cellular mRNA export pathway', *The EMBO Journal*, vol. 20, no. 20, pp. 5769-5778.

Koh, CS & Sarin, LP 2018, 'Transfer RNA modification and infection - Implications for pathogenicity and host responses', *Biochim Biophys Acta Gene Regul Mech*, vol. 1861, no. 4, Apr, pp. 419-432.

Kohler, A & Hurt, E 2007, 'Exporting RNA from the nucleus to the cytoplasm', *Nat Rev Mol Cell Biol*, vol. 8, no. 10, Oct, pp. 761-773.

Kohli, RM & Zhang, Y 2013, 'TET enzymes, TDG and the dynamics of DNA demethylation', *Nature*, vol. 502, no. 7472, pp. 472-479.

Kolde, R 2012, 'Pheatmap: pretty heatmaps', *R package version*, vol. 1, no. 2, p. 726.

Kowalak, JA, Pomerantz, SC, Crain, PF & McCloskey, JA 1993, 'A novel method for the determination of posttranscriptional modification in RNA by mass spectrometry', *Nucleic Acids Research*, vol. 21, no. 19, pp. 4577-4585.

Kriaucionis, S & Heintz, N 2009, 'The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain', *Science*, vol. 324, no. 5929, pp. 929-930.

Krishna, P & Gloor, G 2001, 'The Hsp90 family of proteins in Arabidopsis thaliana', *Cell Stress & Chaperones*, vol. 6, no. 3, p. 238.

Krueger, F 2012, 'Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries', *URL http://www. bioinformatics. babraham. ac. uk/projects/trim_galore/.(Date of access: 28/04/2016)*.

Kurowski, MA, Bhagwat, AS, Papaj, G & Bujnicki, JM 2003, 'Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB', *BMC Genomics*, vol. 4, no. 1, pp. 1-6.

Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H, Valentin, F, Wallace, IM, Wilm, A, Lopez, R, Thompson, JD, Gibson, TJ & Higgins, DG 2007, 'Clustal W and Clustal X version 2.0', *Bioinformatics*, vol. 23, no. 21, pp. 2947-2948.

Leihne, V, Kirpekar, F, Vågbø, CB, Van den Born, E, Krokan, HE, Grini, PE, Meza, TJ & Falnes, PØ 2011, 'Roles of Trm9-and ALKBH8-like proteins in the formation of modified wobble uridines in Arabidopsis tRNA', *Nucleic Acids Research*, vol. 39, no. 17, pp. 7688-7701.

Lesk, C, Rowhani, P & Ramankutty, N 2016, 'Influence of extreme weather disasters on global crop production', *Nature*, vol. 529, no. 7584, 2016/01/01, pp. 84-87.

Lewis, CJT, Pan, T & Kalsotra, A 2017, 'RNA modifications and structures cooperate to guide RNA–protein interactions', *Nature Reviews Molecular Cell Biology*, vol. 18, no. 3, 2017/03/01, pp. 202-210.

Lewis, J & Bird, A 1991, 'DNA methylation and chromatin structure', *FEBS letters*, vol. 285, no. 2, pp. 155-159.

Li, B, Gao, K, Ren, H & Tang, W 2018, 'Molecular mechanisms governing plant responses to high temperatures', *J Integr Plant Biol*, vol. 60, no. 9, Sep, pp. 757-779.

Li, D, Zhang, H, Hong, Y, Huang, L, Li, X, Zhang, Y, Ouyang, Z & Song, F 2014, 'Genome-wide identification, biochemical characterization, and expression analyses of the YTH domain-containing RNA-binding protein family in Arabidopsis and Rice', *Plant Molecular Biology Reporter*, vol. 32, no. 6, pp. 1169-1186.

Li, J, Wu, X, Do, T, Nguyen, V, Zhao, J, Ng, PQ, Burgess, A, David, R & Searle, I 2021, 'Quantitative and Single-Nucleotide Resolution Profiling of RNA 5-Methylcytosine', in *RNA Modifications*, Springer, pp. 135-151.

Li, S & Mason, CE 2014, 'The pivotal regulatory landscape of RNA modifications', *Annu Rev Genomics Hum Genet*, vol. 15, pp. 127-150.

Li, X, Xiong, X, Wang, K, Wang, L, Shu, X, Ma, S & Yi, C 2016, 'Transcriptome-wide mapping reveals reversible and dynamic N 1-methyladenosine methylome', *Nature chemical biology*, vol. 12, no. 5, pp. 311-316.

Li, X, Xiong, X & Yi, C 2016, 'Epitranscriptome sequencing technologies: decoding RNA modifications', *Nat Methods*, vol. 14, no. 1, Dec 29, pp. 23-31.

Liang, X-h, Liu, Q & Fournier, MJ 2009, 'Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing', *RNA*, vol. 15, no. 9, pp. 1716-1728.

Liao, Y, Smyth, GK & Shi, W 2013, 'The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote', *Nucleic Acids Research*, vol. 41, no. 10, pp. e108-e108.

Lin, B-L, Wang, J-S, Liu, H-C, Chen, R-W, Meyer, Y, Barakat, A & Delseny, M 2001, 'Genomic analysis of the Hsp70 superfamily in Arabidopsis thaliana', *Cell Stress & Chaperones*, vol. 6, no. 3, p. 201.

Liu, J, Feng, L, Li, J & He, Z 2015, 'Genetic and epigenetic control of plant heat responses', *Frontiers in plant science*, vol. 6, 2015-April-24.

Liu, Q & Gregory, RI 2019, 'RNAmod: an integrated system for the annotation of mRNA modifications', *Nucleic Acids Research*, vol. 47, no. W1, pp. W548-W555.

Liu, R & Lang, Z 2020, 'The mechanism and function of active DNA demethylation in plants', *Journal of Integrative Plant Biology*, vol. 62, no. 1, pp. 148-159.

Long, HK, Blackledge, NP & Klose, RJ 2013, 'ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection', *Biochem Soc Trans*, vol. 41, no. 3, Jun, pp. 727-740.

Lorenz, R, Bernhart, SH, Zu Siederdissen, CH, Tafer, H, Flamm, C, Stadler, PF & Hofacker, IL 2011, 'ViennaRNA Package 2.0', *Algorithms for Molecular Biology*, vol. 6, no. 1, p. 26.

Love, MI, Huber, W & Anders, S 2014, 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, vol. 15, no. 12, p. 550.

Lv, H, Dao, FY, Zhang, D, Yang, H & Lin, H 2021, 'Advances in mapping the epigenetic modifications of 5‐methylcytosine (5mC), N6‐methyladenine (6mA), and N4‐methylcytosine (4mC)', *Biotechnology and Bioengineering*, vol. 118, no. 11, pp. 4204-4216.

Ma, Y, Min, L, Wang, M, Wang, C, Zhao, Y, Li, Y, Fang, Q, Wu, Y, Xie, S & Ding, Y 2018, 'Disrupted genome methylation in response to high temperature has distinct affects on microspore abortion and anther indehiscence', *The Plant Cell*, vol. 30, no. 7, pp. 1387-1403.

Madore, E, Florentz, C, Giegé, R, Sekine, Si, Yokoyama, S & Lapointe, J 1999, 'Effect of modified nucleotides on Escherichia coli tRNAGlu structure and on its aminoacylation by glutamyl‐tRNA synthetase: Predominant and distinct roles of the mnm5 and s2 modifications of U34', *European Journal of Biochemistry*, vol. 266, no. 3, pp. 1128-1135.

Maekawa, S & Yanagisawa, S 2021, 'Ribosome biogenesis factor OLI2 and its interactor BRX1-2 are associated with morphogenesis and lifespan extension in Arabidopsis thaliana', *Plant Biotechnology*, p. 20.1224 a.

Marcinkowski, M, Pilžys, T, Garbicz, D, Steciuk, J, Zugaj, D, Mielecki, D, Sarnowski, TJ & Grzesiuk, E 2020, 'Human and Arabidopsis alpha‐ketoglutarate

- dependent dioxygenase homolog proteins — New players in important regulatory processes', *IUBMB life*, vol. 72, no. 6, pp. 1126-1144.

Martin, C & Zhang, Y 2005, 'The diverse functions of histone lysine methylation', *Nat Rev Mol Cell Biol*, vol. 6, no. 11, Nov, pp. 838-849.

Martínez-Pérez, M, Aparicio, F, López-Gresa, MP, Bellés, JM, Sánchez-Navarro, JA & Pallás, V 2017, 'Arabidopsis m6A demethylase activity modulates viral infection of a plant virus and the m6A abundance in its genomic RNAs', *Proceedings of the National Academy of Sciences*, vol. 114, no. 40, pp. 10755-10760.

Masiello, I & Biggiogera, M 2017, 'Ultrastructural localization of 5-methylcytosine on DNA and RNA', *Cell Mol Life Sci*, vol. 74, no. 16, Aug, pp. 3057-3064.

McDonough, MA, Loenarz, C, Chowdhury, R, Clifton, IJ & Schofield, CJ 2010, 'Structural studies on human 2-oxoglutarate dependent oxygenases', *Curr Opin Struct Biol*, vol. 20, no. 6, Dec, pp. 659-672.

Meng, S, Zhou, H, Feng, Z, Xu, Z, Tang, Y & Wu, M 2019, 'Epigenetics in Neurodevelopment: Emerging Role of Circular RNA', *Frontiers in Cellular Neuroscience*, vol. 13, no. 327, 2019-July-19.

Metodiev, MD, Spåhr, H, Polosa, PL, Meharg, C, Becker, C, Altmueller, J, Habermann, B, Larsson, N-G & Ruzzenente, B 2014, 'NSUN4 is a dual function mitochondrial protein required for both methylation of 12S rRNA and coordination of mitoribosomal assembly', *PLoS genetics*, vol. 10, no. 2.

Meza, TJ, Moen, MN, Vågbø, CB, Krokan, HE, Klungland, A, Grini, PE & Falnes, PØ 2012, 'The DNA dioxygenase ALKBH2 protects Arabidopsis thaliana against methylation damage', *Nucleic Acids Research*, vol. 40, no. 14, pp. 6620-6631.

Min, L, Li, Y, Hu, Q, Zhu, L, Gao, W, Wu, Y, Ding, Y, Liu, S, Yang, X & Zhang, X 2014, 'Sugar and auxin signaling pathways respond to high-temperature stress during anther development as revealed by transcript profiling analysis in cotton', *Plant physiology*, vol. 164, no. 3, pp. 1293-1308.

Mlynárová, L, Nap, J-P & Bisseling, T 2007, 'The SWI/SNF chromatin-remodeling gene AtCHR12 mediates temporary growth arrest in Arabidopsis thaliana upon perceiving environmental stress', *The Plant Journal*, vol. 51, no. 5, pp. 874-885.

Molinier, J, Ries, G, Zipfel, C & Hohn, B 2006, 'Transgeneration memory of stress in plants', *Nature*, vol. 442, no. 7106, pp. 1046-1049.

Motorin, Y & Grosjean, H 1999, 'Multisite-specific tRNA: m5C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: identification of the gene and substrate specificity of the enzyme', *RNA*, vol. 5, no. 8, pp. 1105-1118.

Motorin, Y & Helm, M 2010, 'tRNA stabilization by modified nucleotides', *Biochemistry*, vol. 49, no. 24, Jun 22, pp. 4934-4944.

Motorin, Y, Lyko, F & Helm, M 2009, '5-methylcytosine in RNA: detection, enzymatic formation and biological functions', *Nucleic Acids Research*, vol. 38, no. 5, pp. 1415-1430.

Moyle, RL, Carvalhais, LC, Pretorius, LS, Nowak, E, Subramaniam, G, Dalton-Morgan, J & Schenk, PM 2017, 'An Optimized Transient Dual Luciferase Assay for Quantifying MicroRNA Directed Repression of Targeted Sequences', *Front Plant Sci*, vol. 8, p. 1631.

Nachtergaele, S & He, C 2018, 'Chemical Modifications in the Life of an mRNA Transcript', *Annu Rev Genet*, vol. 52, Nov 23, pp. 349-372.

Nakano, S, Suzuki, T, Kawarada, L, Iwata, H, Asano, K & Suzuki, T 2016, 'NSUN3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNA Met', *Nature chemical biology*, vol. 12, no. 7, p. 546.

Nishad, A & Nandi, AK 2021, 'Recent advances in plant thermomemory', *Plant Cell Rep*, vol. 40, no. 1, Jan, pp. 19-27.

Nishimura, K, Apitz, J, Friso, G, Kim, J, Ponnala, L, Grimm, B & van Wijk, KJ 2015, 'Discovery of a unique Clp component, ClpF, in chloroplasts: a proposed binary ClpF-ClpS1 adaptor complex functions in substrate recognition and delivery', *The Plant Cell*, vol. 27, no. 10, pp. 2677-2691.

Niu, Y, Zhao, X, Wu, Y-S, Li, M-M, Wang, X-J & Yang, Y-G 2013, 'N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function', *Genomics, Proteomics & Bioinformatics*, vol. 11, no. 1, pp. 8-17.

Nobles, KN, Yarian, CS, Liu, G, Guenther, RH & Agris, PF 2002, 'Highly conserved modified nucleosides influence Mg2+-dependent tRNA folding', *Nucleic Acids Research*, vol. 30, no. 21, pp. 4751-4760.

Nover, N, Bharti, K, Döring, P, Mishra, SK, Ganguli, A & Scharf, K-D 2001, 'Arabidopsis and the Heat Stress Transcription Factor World: How Many Heat Stress Transcription Factors Do We Need?', *Cell Stress & Chaperones*, vol. 6, no. 3, pp. 177-189.

O'Malley, RC, Barragan, CC & Ecker, JR 2015, 'A user's guide to the Arabidopsis T-DNA insertion mutant collections', *Methods in molecular biology (Clifton, N.J.)*, vol. 1284, pp. 323-342.

Ohama, N, Kusakabe, K, Mizoi, J, Zhao, H, Kidokoro, S, Koizumi, S, Takahashi, F, Ishida, T, Yanagisawa, S & Shinozaki, K 2016, 'The transcriptional cascade in the heat stress response of Arabidopsis is strictly regulated at the level of transcription factor expression', *The Plant Cell*, vol. 28, no. 1, pp. 181-201.

Ohama, N, Sato, H, Shinozaki, K & Yamaguchi-Shinozaki, K 2017, 'Transcriptional Regulatory Network of Plant Heat Stress Response', *Trends Plant Sci*, vol. 22, no. 1, Jan, pp. 53-65.

Oliva, R, Cavallo, L & Tramontano, A 2006, 'Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions', *Nucleic Acids Research*, vol. 34, no. 3, pp. 865-879.

Oliva, R, Tramontano, A & Cavallo, L 2007, 'Mg2+ binding and archaeosine modification stabilize the G15–C48 Levitt base pair in tRNAs', *RNA*, vol. 13, no. 9, pp. 1427-1436.

Pagamas, P & Nawata, E 2008, 'Sensitive stages of fruit and seed development of chili pepper (Capsicum annuum L. var. Shishito) exposed to high-temperature stress', *Scientia Horticulturae*, vol. 117, no. 1, pp. 21-25.

Pavlopoulou, A & Kossida, S 2009, 'Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases', *Genomics*, vol. 93, no. 4, pp. 350-357.

Pavlopoulou, A & Kossida, S 2009, 'Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases', *Genomics*, vol. 93, no. 4, Apr, pp. 350-357.

Pfaff, C, Ehrnsberger, HF, Flores-Tornero, M, Sorensen, BB, Schubert, T, Langst, G, Griesenbeck, J, Sprunck, S, Grasser, M & Grasser, KD 2018, 'ALY RNA-Binding Proteins Are Required for Nucleocytosolic mRNA Transport and Modulate Plant Growth and Development', *Plant Physiol*, vol. 177, no. 1, May, pp. 226-240.

Qiao, W & Fan, L 2011, 'Epigenetics, a mode for plants to respond to abiotic stresses', *Frontiers in Biology*, vol. 6, no. 6, pp. 477-481.

Quinlan, AR & Hall, IM 2010, 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, vol. 26, no. 6, pp. 841-842.

Rai, K, Chidester, S, Zavala, CV, Manos, EJ, James, SR, Karpf, AR, Jones, DA & Cairns, BR 2007, 'Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish', *Genes & development*, vol. 21, no. 3, pp. 261-266.

Reid, R, Greene, PJ & Santi, DV 1999, 'Exposition of a family of RNA m5C methyltransferases from searching genomic and proteomic sequences', *Nucleic Acids Research*, vol. 27, no. 15, pp. 3138-3145.

Rieder, D, Amort, T, Kugler, E, Lusser, A & Trajanoski, Z 2016, 'meRanTK: methylated RNA analysis ToolKit', *Bioinformatics*, vol. 32, no. 5, pp. 782-785.

Rivero, L, Scholl, R, Holomuzki, N, Crist, D, Grotewold, E & Brkljacic, J 2014, 'Handling Arabidopsis plants: growth, preservation of seeds, transformation, and genetic crosses', in *Arabidopsis protocols*, Springer, pp. 3-25.

Rodriguez-Hernandez, A, Spears, JL, Gaston, KW, Limbach, PA, Gamper, H, Hou, Y-M, Kaiser, R, Agris, PF & Perona, JJ 2013, 'Structural and mechanistic basis for enhanced translational efficiency by 2-thiouridine at the tRNA anticodon wobble position', *Journal of molecular biology*, vol. 425, no. 20, pp. 3888-3906.

Rościsław, K, Karol, Z & Sebastian, G 2019, 'Charging the code — tRNA modification complexes', *Current Opinion in Structural Biology*, vol. 55, pp. 138-146.

Ruelland, E & Zachowski, A 2010, 'How plants sense temperature', *Environmental and Experimental Botany*, vol. 69, no. 3, pp. 225-232.

Sakita-Suto, S, Kanda, A, Suzuki, F, Sato, S, Takata, T & Tatsuka, M 2007, 'Aurora-b regulates rna methyltransferase nsun2', *Molecular biology of the cell*, vol. 18, no. 3, pp. 1107-1117.

Salditt-Georgieff, M, Jelinek, W, Darnell, JE, Furuichi, Y, Morgan, M & Shatkin, A 1976, 'Methyl labeling of HeLa cell hnRNA: a comparison with mRNA', *Cell*, vol. 7, no. 2, pp. 227-237.

Saletore, Y, Meyer, K, Korlach, J, Vilfan, ID, Jaffrey, S & Mason, CE 2012, 'The birth of the Epitranscriptome: deciphering the function of RNA modifications', *Genome Biology*, vol. 13, no. 10, 2012/10/31, p. 175.

Savchenko, GE, Klyuchareva, EA, Abramchik, LM & Serdyuchenko, EV 2002, 'Effect of Periodic Heat Shock on the Inner Membrane System of Etioplasts', *Russian Journal of Plant Physiology*, vol. 49, no. 3, 2002/05/01, pp. 349-359.

Schaefer, M 2015, 'Chapter Fourteen - RNA 5-Methylcytosine Analysis by Bisulfite Sequencing', in C He (ed.), *Methods in Enzymology*, vol. 560, Academic Press, pp. 297-329.

Schaefer, M, Pollex, T, Hanna, K & Lyko, F 2008, 'RNA cytosine methylation analysis by bisulfite sequencing', *Nucleic Acids Research*, vol. 37, no. 2, pp. e12-e12.

Schaefer, M, Pollex, T, Hanna, K, Tuorto, F, Meusburger, M, Helm, M & Lyko, F 2010, 'RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage', *Genes & development*, vol. 24, no. 15, pp. 1590-1595.

Scharf, K-D, Siddique, M & Vierling, E 2001, 'The Expanding Family of Arabidopsis thaliana Small Heat Stress Proteins and a New Family of Proteins Containing α-Crystallin Domains (Acd Proteins)', *Cell Stress & Chaperones*, vol. 6, no. 3, pp. 225-237.

Schosserer, M, Minois, N, Angerer, TB, Amring, M, Dellago, H, Harreither, E, Calle-Perez, A, Pircher, A, Gerstl, MP, Pfeifenberger, S, Brandl, C, Sonntagbauer, M, Kriegner, A, Linder, A, Weinhausel, A, Mohr, T, Steiger, M, Mattanovich, D, Rinnerthaler, M, Karl, T, Sharma, S, Entian, KD, Kos, M, Breitenbach, M, Wilson, IB, Polacek, N, Grillari-Voglauer, R, Breitenbach-Koller, L & Grillari, J 2015,

'Methylation of ribosomal RNA by NSUN5 is a conserved mechanism modulating organismal lifespan', *Nat Commun*, vol. 6, Jan 30, p. 6158.

Segref, A, Sharma, K, Doye, V, Hellwig, A, Huber, J, Lührmann, R & Hurt, E 1997, 'Mex67p, a novel factor for nuclear mRNA export, binds to both poly (A)+ RNA and nuclear pores', *The EMBO Journal*, vol. 16, no. 11, pp. 3256-3271.

Shapiro, R, Braverman, B, Louis, JB & Servis, RE 1973, 'Nucleic Acid Reactivity and Conformation: II. REACTION OF CYTOSINE AND URACIL WITH SODIUM BISULFITE', *Journal of Biological Chemistry*, vol. 248, no. 11, 1973/06/10/, pp. 4060-4064.

Sharma, S, Yang, J, Watzinger, P, Kötter, P & Entian, K-D 2013, 'Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively', *Nucleic Acids Research*, p. 679.

Shen, L, Song, C-X, He, C & Zhang, Y 2014, 'Mechanism and function of oxidative reversal of DNA and RNA methylation', *Annual review of biochemistry*, vol. 83, pp. 585-614.

Sibbritt, T, Patel, HR & Preiss, T 2013, 'Mapping and significance of the mRNA methylome', *Wiley Interdisciplinary Reviews: RNA*, vol. 4, no. 4, pp. 397-422.

Sloan, KE, Warda, AS, Sharma, S, Entian, KD, Lafontaine, DLJ & Bohnsack, MT 2017, 'Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function', *RNA Biol*, vol. 14, no. 9, Sep 2, pp. 1138-1152.

Smith, JE, Cooperman, BS & Mitchell, P 1992, 'Methylation sites in Escherichia coli ribosomal RNA: Localization and identification of four new sites of methylation in 23 S rRNA', *Biochemistry*, vol. 31, no. 44, pp. 10825-10834.

Song, J, Zhai, J, Bian, E, Song, Y, Yu, J & Ma, C 2018, 'Transcriptome-Wide Annotation of m(5)C RNA Modifications Using Machine Learning', *Front Plant Sci*, vol. 9, p. 519.

Songe-Moller, L, van den Born, E, Leihne, V, Vagbo, CB, Kristoffersen, T, Krokan, HE, Kirpekar, F, Falnes, PO & Klungland, A 2010, 'Mammalian ALKBH8 possesses tRNA methyltransferase activity required for the biogenesis of multiple wobble uridine modifications implicated in translational decoding', *Mol Cell Biol*, vol. 30, no. 7, Apr, pp. 1814-1827.

Sowers, LC, Shaw, BR & Sedwick, WD 1987, 'Base stacking and molecular polarizability: Effect of a methyl group in the 5-position of pyrimidines', *Biochemical and Biophysical Research Communications*, vol. 148, no. 2, 1987/10/29/, pp. 790-794.

Squires, JE, Patel, HR, Nousch, M, Sibbritt, T, Humphreys, DT, Parker, BJ, Suter, CM & Preiss, T 2012, 'Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA', *Nucleic Acids Research*, vol. 40, no. 11, pp. 5023-5033.

Squires, JE, Patel, HR, Nousch, M, Sibbritt, T, Humphreys, DT, Parker, BJ, Suter, CM & Preiss, T 2012, 'Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA', *Nucleic Acids Research*, p. 144.

Stephanie, LN & Timothy, ROC 2013, 'Direct Repair in Mammalian Cells', in C Clark (ed.), *New Research Directions in DNA Repair*, IntechOpen, Rijeka.

Stief, A, Altmann, S, Hoffmann, K, Pant, BD, Scheible, W-R & Bäurle, I 2014, 'Arabidopsis miR156 Regulates Tolerance to Recurring Environmental Stress through SPL Transcription Factors ', *The Plant Cell*, vol. 26, no. 4, pp. 1792-1807.

Strenkert, D, Schmollinger, S, Sommer, F, Schulz-Raffelt, M & Schroda, M 2011, 'Transcription Factor–Dependent Chromatin Remodeling at Heat Shock and Copper-Responsive Promoters in Chlamydomonas reinhardtii ', *The Plant Cell*, vol. 23, no. 6, pp. 2285-2301.

Sugio, A, Dreos, R, Aparicio, F & Maule, AJ 2009, 'The cytosolic protein response as a subcomponent of the wider heat shock response in Arabidopsis', *The Plant Cell*, vol. 21, no. 2, pp. 642-654.

Sun, K, Zheng, Y & Zhu, Z 2017, 'Luciferase Complementation Imaging Assay in Nicotiana benthamiana Leaves for Transiently Determining Protein-protein Interaction Dynamics', *J Vis Exp*, no. 129, Nov 20.

Swindell, WR, Huebner, M & Weber, AP 2007, 'Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways', *BMC Genomics*, vol. 8, May 22, p. 125.

Tahiliani, M, Koh, KP, Shen, Y, Pastor, WA, Bandukwala, H, Brudno, Y, Agarwal, S, Iyer, LM, Liu, DR, Aravind, L & Rao, A 2009, 'Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1', *Science*, vol. 324, no. 5929, pp. 930-935.

Tan, L & Shi, YG 2012, 'Tet family proteins and 5-hydroxymethylcytosine in development and disease', *Development*, vol. 139, no. 11, pp. 1895-1902.

Tang, Y, Gao, C-C, Gao, Y, Yang, Y, Shi, B, Yu, J-L, Lyu, C, Sun, B-F, Wang, H-L & Xu, Y 2020, 'OsNSUN2-mediated 5-methylcytosine mRNA modification enhances rice adaptation to high temperature', *Developmental cell*, vol. 53, no. 3, pp. 272-286. e277.

Tang, Y, Wen, X, Lu, Q, Yang, Z, Cheng, Z & Lu, C 2007, 'Heat stress induces an aggregation of the light-harvesting complex of photosystem II in spinach plants', *Plant physiology*, vol. 143, no. 2, pp. 629-638.

Taniguchi, I & Ohno, M 2008, 'ATP-dependent recruitment of export factor Aly/REF onto intronless mRNAs by RNA helicase UAP56', *Mol Cell Biol*, vol. 28, no. 2, Jan, pp. 601-608.

Tate, PH & Bird, AP 1993, 'Effects of DNA methylation on DNA-binding proteins and gene expression', *Current opinion in genetics & development*, vol. 3, no. 2, pp. 226-231.

Thieme, CJ, Rojas-Triana, M, Stecyk, E, Schudoma, C, Zhang, W, Yang, L, Miñambres, M, Walther, D, Schulze, WX & Paz-Ares, J 2015, 'Endogenous Arabidopsis messenger RNAs transported to distant tissues', *Nature Plants*, vol. 1, no. 4, pp. 1-9.

Tian, F, Hu, X-L, Yao, T, Yang, X, Chen, J-G, Lu, M-Z & Zhang, J 2021, 'Recent Advances in the Roles of HSFs and HSPs in Heat Stress Response in Woody Plants', *Frontiers in plant science*, vol. 12, 2021-July-09.

TISNÉ, C, RIGOURD, M, MARQUET, R, EHRESMANN, C & DARDEL, F 2000, 'NMR and biochemical characterization of recombinant human tRNA3Lys expressed in Escherichia coli: identification of posttranscriptional nucleotide modifications required for efficient initiation of HIV-1 reverse transcription', *RNA*, vol. 6, no. 10, pp. 1403-1412.

Toh, JD, Crossley, SW, Bruemmer, KJ, Eva, JG, He, D, Iovan, DA & Chang, CJ 2020, 'Distinct RNA N-demethylation pathways catalyzed by nonheme iron ALKBH5 and FTO enzymes enable regulation of formaldehyde release rates', *Proceedings of the National Academy of Sciences*, vol. 117, no. 41, pp. 25284-25292.

Tripp, J, Mishra, SK & SCHARF, KD 2009, 'Functional dissection of the cytosolic chaperone network in tomato mesophyll protoplasts', *Plant, cell & environment*, vol. 32, no. 2, pp. 123-133.

Trixl, L, Amort, T, Wille, A, Zinni, M, Ebner, S, Hechenberger, C, Eichin, F, Gabriel, H, Schoberleitner, I, Huang, A, Piatti, P, Nat, R, Troppmair, J & Lusser, A 2018, 'RNA cytosine methyltransferase Nsun3 regulates embryonic stem cell differentiation by promoting mitochondrial activity', *Cell Mol Life Sci*, vol. 75, no. 8, Apr, pp. 1483-1497.

Trixl, L & Lusser, A 2018, 'The dynamic RNA modification 5‐methylcytosine and its emerging role as an epitranscriptomic mark', *Wiley Interdisciplinary Reviews: RNA*, vol. 10, no. 1, p. e1510.

Tschudi, C & Ullu, E 2002, 'Unconventional rules of small nuclear RNA transcription and cap modification in trypanosomatids', *Gene Expression, The Journal of Liver Research*, vol. 10, no. 1-2, pp. 3-16.

Tuck, MT, Wiehl, PE & Pan, T 1999, 'Inhibition of 6-methyladenine formation decreases the translation efficiency of dihydrofolate reductase transcripts', *The international journal of biochemistry & cell biology*, vol. 31, no. 8, pp. 837-851.

Tuorto, F, Herbst, F, Alerasool, N, Bender, S, Popp, O, Federico, G, Reitter, S, Liebers, R, Stoecklin, G, Grone, HJ, Dittmar, G, Glimm, H & Lyko, F 2015, 'The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis', *EMBO J*, vol. 34, no. 18, Sep 14, pp. 2350-2362.

Tuorto, F, Liebers, R, Musch, T, Schaefer, M, Hofmann, S, Kellner, S, Frye, M, Helm, M, Stoecklin, G & Lyko, F 2012, 'RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis', *Nature structural & molecular biology*, vol. 19, no. 9, p. 900.

Turek-Plewa, J & Jagodzinski, P 2005, 'The role of mammalian DNA methyltransferases in the regulation of gene expression', *Cellular and Molecular Biology Letters*, vol. 10, no. 4, p. 631.

Ueda, M & Seki, M 2019, 'Histone Modifications Form Epigenetic Regulatory Networks to Regulate Abiotic Stress Response1 [OPEN]', *Plant physiology*, vol. 182, no. 1, pp. 15-26.

Van Haute, L, Dietmann, S, Kremer, L, Hussain, S, Pearce, SF, Powell, CA, Rorbach, J, Lantaff, R, Blanco, S & Sauer, S 2016, 'Deficient methylation and formylation of mt-tRNA Met wobble cytosine in a patient carrying mutations in NSUN3', *Nature communications*, vol. 7, no. 1, pp. 1-10.

Vandenabeele, S, Vanderauwera, S, Vuylsteke, M, Rombauts, S, Langebartels, C, Seidlitz, HK, Zabeau, M, Van Montagu, M, Inzé, D & Van Breusegem, F 2004,

'Catalase deficiency drastically affects gene expression induced by high light in Arabidopsis thaliana', *The Plant Journal*, vol. 39, no. 1, pp. 45-58.

Vare, VY, Eruysal, ER, Narendran, A, Sarachan, KL & Agris, PF 2017, 'Chemical and Conformational Diversity of Modified Nucleosides Affects tRNA Structure and Function', *Biomolecules*, vol. 7, no. 1, Mar 16.

Voigts-Hoffmann, F, Hengesbach, M, Kobitski, AY, Van Aerschot, A, Herdewijn, P, Nienhaus, GU & Helm, M 2007, 'A methyl group controls conformational equilibrium in human mitochondrial tRNALys', *Journal of the American Chemical Society*, vol. 129, no. 44, pp. 13382-13383.

Walbott, H, Auxilien, S, Grosjean, H & Golinelli-Pimpaneau, B 2007, 'The carboxyl-terminal extension of yeast tRNA m5C methyltransferase enhances the catalytic efficiency of the amino-terminal domain', *Journal of Biological Chemistry*, vol. 282, no. 32, pp. 23663-23671.

Wang, L, Ma, K-B, Lu, Z-G, Ren, S-X, Jiang, H-R, Cui, J-W, Chen, G, Teng, N-J, Lam, H-M & Jin, B 2020, 'Differential physiological, transcriptomic and metabolomic responses of Arabidopsis leaves under prolonged warming and heat shock', *BMC Plant Biology*, vol. 20, no. 1, pp. 1-15.

Wang, X, Wang, M, Dai, X, Han, X, Zhou, Y, Lai, W, Zhang, L, Yang, Y, Chen, Y & Wang, H 2021, 'RNA 5-methylcytosine regulates YBX2-dependent liquid-liquid phase separation', *Fundamental Research*.

Warren, L, Manos, PD, Ahfeldt, T, Loh, Y-H, Li, H, Lau, F, Ebina, W, Mandal, PK, Smith, ZD & Meissner, A 2010, 'Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA', *Cell stem cell*, vol. 7, no. 5, pp. 618-630.

Weigel, D & Glazebrook, J 2009, 'Quick miniprep for plant DNA isolation', *Cold Spring Harbor Protocols*, vol. 2009, no. 3, p. pdb. prot5179.

Weston, DJ, Karve, AA, Gunter, LE, Jawdy, SS, Yang, X, Allen, SM & Wullschleger, SD 2011, 'Comparative physiology and transcriptional networks underlying the heat shock response in Populus trichocarpa, Arabidopsis thaliana and Glycine max', *Plant, cell & environment*, vol. 34, no. 9, pp. 1488-1506.

Winter, D, Vinegar, B, Nahal, H, Ammar, R, Wilson, GV & Provart, NJ 2007, 'An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets', *PloS one*, vol. 2, no. 8, p. e718.

Wu, H, D'Alessio, AC, Ito, S, Xia, K, Wang, Z, Cui, K, Zhao, K, Eve Sun, Y & Zhang, Y 2011, 'Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells', *Nature*, vol. 473, no. 7347, 2011/05/01, pp. 389-393.

Wu, H, Qu, X, Dong, Z, Luo, L, Shao, C, Forner, J, Lohmann, JU, Su, M, Xu, M & Liu, X 2020, 'WUSCHEL triggers innate antiviral immunity in plant stem cells', *Science*, vol. 370, no. 6513, pp. 227-231.

Wu, JR, Wang, LC, Lin, YR, Weng, CP, Yeh, CH & Wu, SJ 2017, 'The Arabidopsis heat-intolerant 5 (hit5)/enhanced response to aba 1 (era1) mutant reveals the crucial role of protein farnesylation in plant responses to heat stress', *New Phytol*, vol. 213, no. 3, Feb, pp. 1181-1193.

Xiong, X, Yi, C & Peng, J 2017, 'Epitranscriptomics: Toward A Better Understanding of RNA Modifications', *Genomics, Proteomics & Bioinformatics*.

Xu, H, Iwashiro, R, Li, T & Harada, T 2013, 'Long-distance transport of Gibberellic Acid Insensitive mRNA in Nicotiana benthamiana', *BMC Plant Biology*, vol. 13, no. 1, pp. 1-9.

Xu, Y, Xu, C, Kato, A, Tempel, W, Abreu, JG, Bian, C, Hu, Y, Hu, D, Zhao, B, Cerovina, T, Diao, J, Wu, F, He, HH, Cui, Q, Clark, E, Ma, C, Barbara, A, Veenstra, GJ, Xu, G, Kaiser, UB, Liu, XS, Sugrue, SP, He, X, Min, J, Kato, Y & Shi, YG 2012, 'Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development', *Cell*, vol. 151, no. 6, Dec 7, pp. 1200-1213.

Xu, Y, Yu, Z, Zhang, D, Huang, J, Wu, C, Yang, G, Yan, K, Zhang, S & Zheng, C 2018, 'CYSTM, a Novel Non-Secreted Cysteine-Rich Peptide Family, Involved in Environmental Stresses in Arabidopsis thaliana', *Plant Cell Physiol*, vol. 59, no. 2, Feb 1, pp. 423-438.

Xue, G-P, Sadat, S, Drenth, J & McIntyre, CL 2013, 'The heat shock factor family from Triticum aestivum in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes', *Journal of Experimental Botany*, vol. 65, no. 2, pp. 539-557.

Yamagata, K & Kobayashi, A 2017, 'The cysteine-rich domain of TET2 binds preferentially to mono- and dimethylated histone H3K36', *J Biochem*, vol. 161, no. 4, Apr 1, pp. 327-330.

Yang, L, Perrera, V, Saplaoura, E, Apelt, F, Bahin, M, Kramdi, A, Olas, J, Mueller-Roeber, B, Sokolowska, E, Zhang, W, Li, R, Pitzalis, N, Heinlein, M, Zhang, S, Genovesio, A, Colot, V & Kragler, F 2019, 'm(5)C Methylation Guides Systemic Transport of Messenger RNA over Graft Junctions in Plants', *Curr Biol*, vol. 29, no. 15, Aug 5, pp. 2465-2476 e2465.

Yang, X, Yang, Y, Sun, BF, Chen, YS, Xu, JW, Lai, WY, Li, A, Wang, X, Bhattarai, DP, Xiao, W, Sun, HY, Zhu, Q, Ma, HL, Adhikari, S, Sun, M, Hao, YJ, Zhang, B, Huang, CM, Huang, N, Jiang, GB, Zhao, YL, Wang, HL, Sun, YP & Yang, YG 2017, '5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader', *Cell Res*, vol. 27, no. 5, May, pp. 606-625.

Ye, F, Zhang, L, Jin, L, Zheng, M, Jiang, H & Luo, C 2014, 'Repair of methyl lesions in RNA by oxidative demethylation', *MedChemComm*, vol. 5, no. 12, pp. 1797-1803.

Yi, C, Yang, C-G & He, C 2009, 'A non-heme iron-mediated chemical demethylation in DNA and RNA', *Accounts of chemical research*, vol. 42, no. 4, pp. 519-529.

Yu, B, Edstrom, WC, Benach, J, Hamuro, Y, Weber, PC, Gibney, BR & Hunt, JF 2006, 'Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB', *Nature*, vol. 439, no. 7078, pp. 879-884.

Zdzalik, D, Vagbo, CB, Kirpekar, F, Davydova, E, Puscian, A, Maciejewska, AM, Krokan, HE, Klungland, A, Tudek, B, van den Born, E & Falnes, PO 2014, 'Protozoan ALKBH8 oxygenases display both DNA repair and tRNA modification activities', *PloS one*, vol. 9, no. 6, p. e98729.

Zhang, H, Lang, Z & Zhu, J-K 2018, 'Dynamics and function of DNA methylation in plants', *Nature Reviews Molecular Cell Biology*, vol. 19, no. 8, pp. 489-506.

Zhang, H, Zhang, X, Clark, E, Mulcahey, M, Huang, S & Shi, YG 2010, 'TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription

via hydroxylation of 5-methylcytosine', *Cell Res*, vol. 20, no. 12, Dec, pp. 1390-1393.

Zhang, Q, Zheng, Q, Yu, X, He, Y & Guo, W 2020, 'Overview of distinct 5-methylcytosine profiles of messenger RNA in human hepatocellular carcinoma and paired adjacent non-tumor tissues', *J Transl Med*, vol. 18, no. 1, Jun 22, p. 245.

Zhang, S, Zhang, H, Xia, Y & Xiong, L 2018, 'The caseinolytic protease complex component CLPC1 in Arabidopsis maintains proteome and RNA homeostasis in chloroplasts', *BMC Plant Biology*, vol. 18, no. 1, pp. 1-16.

Zhao, L-Y, Song, J, Liu, Y, Song, C-X & Yi, C 2020, 'Mapping the epigenetic modifications of DNA and RNA', *Protein & Cell*, vol. 11, no. 11, pp. 792-808.

Zhao, Q, Chen, W, Bian, J, Xie, H, Li, Y, Xu, C, Ma, J, Guo, S, Chen, J & Cai, X 2018, 'Proteomics and phosphoproteomics of heat stress-responsive mechanisms in spinach', *Frontiers in plant science*, vol. 9, p. 800.

Zhou, J, Wan, J, Gao, X, Zhang, X, Jaffrey, SR & Qian, SB 2015, 'Dynamic m(6)A mRNA methylation directs translational control of heat shock response', *Nature*, vol. 526, no. 7574, Oct 22, pp. 591-594.