Conservation and Function of RNA 5-methylcytosine in Plants

Alice Louise Burgess

A thesis submitted for the degree of Doctor of Philosophy

The Department of Genetics & Evolution
School of Biological Sciences
March 2016

Table of Contents

Abstract

Post-transcriptional methylation of RNA cytosine residues to 5-methylcytosine (m^5C) is an important modification that regulates RNA metabolism, translation and stress responses and occurs in both eukaryotes and prokaryotes. Yet, to date, no transcriptome-wide identification of m^5C sites has been undertaken in plants. Here, we present over a thousand m^5C sites transcriptome-wide in mRNAs and ncRNAs (non-coding RNAs) in three tissue types; siliques, shoots and roots of *Arabidopsis thaliana* at single nucleotide resolution using high-throughput Illumina sequencing of bisulfite treated RNA (RBS-seq). We show that m^5C methylation sites can be tissue-specific, or shared among the tissue types investigated. Among the shared m^5C sites, some are differentially regulated between tissue types, while others are constitutively methylated at the same level across all three tissue types. Within mRNAs, the majority of m^5C sites are located within coding sequences. A small, significant enrichment of m^5C sites in 3'UTRs of mRNAs was observed when normalizing for length and sequence coverage. We also investigated ncRNAs and demonstrate conservation of rRNA and tRNA m^5C sites across six species in the kingdom *Plantae*, suggesting important and highly conserved roles of this post-transcriptional modification.

We identified over 100 m^5C sites in diverse RNA classes such as mRNAs, lncRNAs (long non-coding RNAs), snoRNAs (small nucleolar RNAs) and tRNAs mediated by *Arabidopsis* tRNA methyltransferase 4B (TRM4B) in siliques, shoots and roots. TRM4 plays broad roles in many organisms for mediating oxidative stress tolerance and balancing stem cell self-renewal and differentiation. We discovered that these roles are also conserved in plants, as *Arabidopsis trm4b* mutants have shorter primary roots, which is linked to a reduced capacity for cells to divide in the root meristem. Furthermore, *trm4b* mutants are also more sensitive to oxidative stress and have reduced stability of non-methylated tRNAs. Here, we extend the known m^5C sites in tRNAs mediated by Transfer RNA aspartic acid methyltransferase 1 (TRDMT1) and find no evidence of m^5C sites mediated by TRDMT1 in other RNA classes. Additionally we demonstrate that rRNA methylation requires the conserved RNA methyltransferase (RMTase) NSUN5. Our results also suggest functional

redundancy of the three predicted RMTase NOP2 paralogs in *Arabidopsis.* This thesis provides the first maps of the *Arabidopsis* m^5C epitranscriptome and characterization of *Arabidopsis* genetic mutants needed to further probe functions of this new layer of gene regulation in plants.

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 give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also 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.

Signed…… Date……..

List of Publications

David R., Burgess A., Parker B., Pulsford K., Sibbritt T., Preiss T., and Searle, I. (2016) Transcriptome-wide mapping of RNA 5-methylcytosine in Arabidopsis mRNAs and ncRNAs. *In preparation*

Burgess A., David R. and Searle, I. (2016) Deciphering the epitranscriptome: a green perspective. *Journal of Integrative Plant Biology* DOI: 10.1111/jipb.12483

Burgess A., David R. and Searle, I. (2015) Conservation of tRNA and rRNA methylation in the kingdom Plantae. *BMC Plant Biology* 15(1), 199

Acknowledgements

Special thanks to my PhD supervisor Iain Searle. Thank you for your guidance and encouragement over the years, it really means a lot to me. I've learnt a lot from you, and like you joked about the other day, as soon as you've finished teaching me and I'm useful, that's when it's time for me to leave!

To Rakesh David, you are the best co-supervisor ever and a great friend. I've really enjoyed working with you, and you've taught me a lot. Thank you.

I would also like to thank the other members of the Searle laboratory, past and present, for making it such a supportive and enjoyable place to work. I would also like to thank our collaborators; Brian Parker (Bioinformatics analysis), Tennille Sibbritt (for teaching me how to perform RNA bisulfite treatment and answering lots of questions) and Thomas Preiss (valuable suggestions and intellectual input). Thank you to Joy Raison, for generating beautiful circle plots and to Simon Baxter for critical reading for the tRNA and rRNA methylation paper.

I would like to thank my best friends Sara Salehi, Namita Mehta, Nida Khan, Reza Shebassi, Saeideh Kazerooni, Rachel Geard and Christian Maskey for encouraging and supporting me throughout my PhD and being my 'Adelaide' family away from home.

Thank you to my loving family for always being there, for encouraging and supporting me throughout my PhD. Thank you for listening to my whining and cheering me up when experiments were failing epically.

Many thanks to the Grains Research & Development Corporation (GRDC) and the Australian Government, who provided me with postgraduate scholarships to support me during my PhD.

Abbreviations

Ψ: Pseudouridine
ABA: Abscisic Acid
ACRF: Australian Centre for Plant Functional Genomics
AdoMet: S-adenosyl-L-methionine
ALKBH: ALKB Dioxygenase Homologue
ALKBH5: Alkylation Repair Homologue Protein 5
ANOVA: Analysis of Variance
AtCPSF30: Arabidopsis thaliana Cleavage and Polyadenylation Specificity Factor 30
AtELP1: Arabidopsis thaliana Elongator Protein 1
AtFIB1: Arabidopsis thaliana Fibrillarin 1
AtFIB2: Arabidopsis thaliana Fibrillarin 2
AtFIB3: Arabidopsis thaliana Fibrillarin 3
AtFIP37: Arabidopsis thaliana FKBP12 Interacting Protein 37
ATMS1: Arabidopsis thaliana Methionine Synthase 1
AtSRp30: Arabidopsis thaliana SR Protein 30
AtTAD1: Arabidopsis thaliana tRNA-specific Adenosine Deaminase 1
AtTRM7: Arabidopsis thaliana Transfer RNA Methyltransferase 7
AtTRM10: Arabidopsis thaliana Transfer RNA Methyltransferase 10
AtTRM11: Arabidopsis thaliana Transfer RNA Methyltransferase 11
AtTRM82: Arabidopsis thaliana Transfer RNA Methyltransferase 82
Aza-IP: 5-Azacytidine Immunoprecipitation
bp: base pair
BS: Bisulfite
bsRNA-seq: RNA bisulfite sequencing
Ca^5C: 5-Carboxylcytosine
Can-0: Canary Isles-0
CDS: Coding Sequence
cDNA: Complementary DNA
Cm: 2'-O-Cytosine Methylation
CMS: Cytosine-5-methylenesulfonate
Col-0: Columbia-0

CPSF30: Cleavage and Polyadenylation Specificity Factor 30

CYCB1: Cyclin B1

CYCB1;1:GUS: Cyclin B1 Promoter Driving the GUS Reporter Construct

D: Dihydrouridine

DAG: Days After Germination

DAP: Days After Pollination

dCAPS: Derived Cleaved Amplified Polymorphic Sequences

DIM1: Adenosine Dimethyl Transferase 1

DIM1A: Adenosine Dimethyl Transferase 1A

DIM1B: Adenosine Dimethyl Transferase 1B

DIM1C: Adenosine Dimethyl Transferase 1C

DM: Differential Methylation

DNA: Deoxyribonucleic Acid

DNMT2: DNA Methyltransferase 2

F_1: First Generation Hybrid

f^5C: 5-Formylcytosine

FDR: False Discovery Rate

FTO: Fat Mass and Obesity Associated Protein

G2-M: Cell Growth Stage 2 - Mitosis

GFP: Green Fluorescent Protein

GO: Gene Ontology

GUS: β–Glucuronidase

H_2O_2: Hydrogen Peroxide

HAMR: High-throughput Annotation of Modified Ribonucleotides

HeLa: Henrietta Lacks Human Cervical Cancer Cells

Hen-16: Henriksjfall-16

hm^5C: 5-Hydroxymethylcytosine

HNRNP: Heterogeneous Nuclear Ribonucleoprotein

HNRNPA2B1: Heterogeneous Nuclear Ribonucleoprotein A2B1

HNRNPC: Heterogeneous Nuclear Ribonucleoprotein C

HPLC: High-Pressure Liquid Chromatography

HuR: Human Antigen R

I: Inosine

i^6A: N^6-Isopentenyladenosylation

IGV: Integrative Genomics viewer

lncRNA: Long Non-coding RNA

m^1G: 1-Methylguanosine

m^2G: 2-Methylguanosine

m_2^6A: N-6 Dimethylation

m^3C: 3-Methylcytosine

m^3T: 3-Methylthymidine

m^3U: 3-Methyluridine

m^4C: N4-Methylcytosine

m^4Cm: N4, 2'-O-Dimethylcytosine

m^5C: 5-Methylcytosine

m^6A: N^6-Methyladenosine

m^7G: 7-Methylguanosine

MAG5: MAIGO5

MALAT1: Metastasis Associated Lung Adenocarcinoma Transcript 1

METTL3: Methyltransferase Like 3

METTL14: Methyltransferase Like 14

miCLIP: Methylation Individual-Nucleotide-Resolution Crosslinking and Immunoprecipitation

miRNA: Micro RNA

mRNA: Messenger RNA

MS: Mass Spectrometry

MS Media: Murashige and Skoog Media

MTA: Adenosine Methyltransferase A

MTB: Adenosine Methyltransferase B

NAT: Natural Antisense Transcript

NaCl: Sodium Chloride

ncRNA: Non-coding RNA

NOP2: Nucleolar Protein 2

NOP2A: Nucleolar Protein 2A

NOP2B: Nucleolar Protein 2B

NOP2C: Nucleolar Protein 2C

NSUN2: NOP2/Sun Domain Protein 2

NSUN5: NOP2/Sun Domain Protein 5

nt: Nucleotide
OLI2: Oligocellula 2
PAR: Photosynthetic Active Radiation
PAMPs: Pathogen Associated Molecular Patterns
PCR: Polymerase Chain Reaction
PRC2: Polycomb Repressive Complex 2
PTC: Peptidyl Transferase Center
PUS: Pseudouridine Synthase
PUS1: Pseudouridine Synthase 1
PUS4: Pseudouridine Synthase 4
PUS7: Pseudouridine Synthase 7
QC: Quiescent Center
q-RT-PCR: Quantitative - Reverse Transcription - Polymerase Chain Reaction
R: Resistance
RAM: Root Apical Meristem
RBS-seq: Illumina RNA BiSulfite Sequencing
RBS-amp-seq: RNA Bisulfite Amplicon Sequencing
RBP: RNA Binding Protein
RCM1: rRNA Cytosine Methyltransferase 1
RCMT9: RNA Cytosine Methyltransferase 9
RIP: RNA Immunoprecipitation
R-Luc: Renilla Luciferase
RMTase: RNA Methyltransferase
RNA: Ribonucleic Acid
RNA-seq: Illumina RNA Sequencing
RNMT: RNA Methyltransferase
RPKM: Reads per Kilobase of Transcript per Million Mapped Reads
rRNA: Ribosomal RNA
SCS9: Suppressor of CSB3 9
s.d: Standard Deviation
s.e: Standard Error
S1: Sensor 1
S2: Sensor 2
S3: Sensor 3

siRNA: Small Interfering RNA

snoRNA: Small Nucleolar RNA

snoRNP: Small Nucleolar Ribonucleoprotein Complex

SNP: Single Nucleotide Polymorphism

snRNA: Small Non-coding RNA

SR: Serine/Arginine Rich

SRSF2: SR Splicing Factor 2

SRSF3: SR Splicing Factor 3

SRSF10: SR Splicing Factor 10

SVR1: Suppressor of Variegation 1

t^6A: Threonylcarbamoyladenosylation

TAIR: The Arabidopsis Information Resource

T-DNA: Transfer-DNA

TERC: Telomerase RNA Component

TET: Ten-Eleven Translocation

TLC: Thin Layer Chromatography

TRDMT1: Transfer RNA Asp Methyltransferase 1

TRM: Transfer RNA Methyltransferase

TRM4: Transfer RNA Methyltransferase 4

TRM4A: Transfer RNA Methyltransferase 4A

TRM4B: Transfer RNA Methyltransferase 4B

TRM4B-OX: TRM4B – Over Expressor

tRNA: Transfer RNA

3'UTR: 3' Untranslated Region

5'UTR: 5' Untranslated Region

WT: Wild Type

WTAP: Wilm's Tumor 1 Associating Protein

X-gal: 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactopyranoside

XIST: X-inactive Specific Transcript

YTHDC1: YTH Domain Containing 1

YTHDF1: YTH Domain Family 1

YTHDF2: YTH Domain Family 2

Chapter 1 Introduction

Deciphering the epitranscriptome: a green perspective

Alice L. Burgess[1,2], Rakesh David[1,2] and Iain R. Searle[1,2,3]

[1]School of Biological Sciences, The University of Adelaide, South Australia, 5005 Australia

[2]School of Agriculture, Food and Wine, The Waite Research Institute, The University of Adelaide, South Australia, 5005 Australia

[3]The University of Adelaide and Shanghai Jiao Tong University Joint International Centre for Agriculture and Health

Journal of Integrative Plant Biology, May 2016

Statement of Authorship

Title of paper	Deciphering the epitranscriptome: a green perspective
Publication Status	Under review manuscript style
Publication details	submitted to the *Journal of Integrative Plant Biology*

Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed below);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Candidate	Alice L. Burgess		
Contribution to paper	Wrote the manuscript and produced the figures.		
Overall percentage (%)	88%		
Signature		Date	15/3/16

Name of Co-Author	Rakesh David		
Contribution to paper	Supervised development of work and edited the manuscript. Overall contribution 2 %.		
Signature		Date	15/03/16

Name of Co-Author	Iain R. Searle		
Contribution to paper	Supervised development of work and edited the manuscript. Overall contribution 10 %.		
Signature		Date	15/3/16

JIPB Journal of Integrative Plant Biology

Deciphering the epitranscriptome: A green perspective

Alice Burgess[1,2], Rakesh David[1,2] and Iain Robert Searle[1,2,3]*

[1]School of Biological Sciences, The University of Adelaide, South Australia, 5005 Australia, [2]School of Agriculture, Food and Wine, The Waite Research Institute, The University of Adelaide, South Australia, 5005 Australia, [3]The University of Adelaide and Shanghai Jiao Tong University Joint International Centre for Agriculture and Health, Joint International Research Laboratory of Metabolic & Developmental Sciences, Adelaide, Australia.

Iain Robert Searle
*Correspondence:
iain.searle@adelaide.edu.au

Abstract The advent of high-throughput sequencing technologies coupled with new detection methods of RNA modifications has enabled investigation of a new layer of gene regulation – the epitranscriptome. With over 100 known RNA modifications, understanding the repertoire of RNA modifications is a huge undertaking. This review summarizes what is known about RNA modifications with an emphasis on discoveries in plants. RNA ribose modifications, base methylations and pseudouridylation are required for normal development in *Arabidopsis*, as mutations in the enzymes modifying them have diverse effects on plant development and stress responses. These modifications can regulate RNA structure, turnover and translation. Transfer RNA and ribosomal RNA modifications have been mapped extensively and their functions investigated in many organisms, including plants. Recent work exploring the locations, functions and targeting of N^6-methyladenosine (m^6A), 5-methylcytosine (m^5C), pseudouridine (Ψ), and additional modifications in mRNAs and ncRNAs are highlighted, as well as those previously known on tRNAs and rRNAs. Many questions remain as to the exact mechanisms of targeting and functions of specific modified sites and whether these modifications have distinct functions in the different classes of RNAs.

Keywords: RNA modifications; epitranscriptome; RNA 5-methylcytosine (m^5C); N^6-methyladenosine (m^6A); Pseudouridine (Ψ); *Arabidopsis*
Citation: Burgess A, David R, Searle IR (2016) Deciphering the epitranscriptome: A green perspective. **J Integr Plant Biol** XX: XX–XX doi: 10.1111/jipb.12483
Edited by: Zhizhong Gong, China Agricultural University, China
Received Mar. 2, 2016; **Accepted** May 10, 2016
Available online on May 12, 2016 at www.wileyonlinelibrary.com/journal/jipb

Invited Expert Review

OnlineOpen

INTRODUCTION

Chemical modifications of DNA and proteins such as histones have been established as important regulators of gene expression, eukaryotic development and stress responses (Suzuki and Bird 2008; Lawrence et al. 2016). More recently, a new level of gene regulation, the epitranscriptome, or RNA modifications has gained interest and momentum. There are over 100 different RNA modifications found in different RNA species, the most abundant and most intensively studied are transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). Loss of modifications on tRNAs and rRNAs is linked to multiple human diseases (Blanco and Frye 2014; Torres et al. 2014) and detrimental effects on development and stress responses in other organisms, including plants, underscoring their vital roles (Motorin and Helm 2011; El Yacoubi et al. 2012). However, new functions and interactions are also being discovered for RNA modifications in mRNAs and other non-coding RNAs (ncRNAs) such as long non-coding RNAs (lncRNAs), micro RNAs (miRNAs) and other small RNAs.

Although the presence of RNA modifications such as the 5' cap structure and internal N^6-methyladenosine (m^6A) in mRNAs has been known for decades (Desrosiers et al. 1974; Perry and Kelley 1974; Dubin and Taylor 1975; Shatkin 1976), the flood gates have only just been opened for a new wave of research describing other modifications and their impact on gene regulation. Several recently developed high-throughput sequencing methods for detecting RNA modifications have allowed investigation of low abundance mRNA and ncRNAs on an unprecedented scale thereby enabling deciphering of their functions in RNA metabolism, gene regulation, translation, development and stress responses (methods are reviewed in Shafik et al. 2016). Interestingly, many of the RNA modifications and the enzymes responsible for 'writing' and 'reading' the modifications are conserved across the three domains of life (Jackman and Alfonzo 2013), suggesting

important, conserved biological functions of this added layer of complexity and flexibility for RNA regulation.

In this article, we provide an update on the research on tRNAs and rRNAs, highlighting the discoveries in plants before discussing the recent studies investigating the dynamic role of RNA modifications in regulating the function of tRNAs and rRNAs, and the epitranscriptomic landscape of other classes of RNA, including mRNAs and lncRNAs. Continuing on from Fray and Simpson (2015), this review extends and discusses recent developments utilizing transcriptome-wide sequencing to explore the RNA modification landscapes of N^6-methyladenosine (m^6A), 5-methylcytosine (m^5C), pseudouridine (Ψ) and other modifications which perturb Watson-Crick base pairing. Finally, we will discuss conclusions and future perspectives to shed light on the *Arabidopsis* epitranscriptome.

PLANT TRANSFER RNA MODIFICATIONS AND THEIR FUNCTIONS

Transfer RNAs are considered the most heavily modified types of RNA, and these modifications are highly conserved in bacteria, yeast, mammals and plants, consistent with their central role in translation. At least 92 unique chemical modifications have been identified in tRNAs with varied chemical properties and effects on the stability and function of tRNAs (Machnicka et al. 2013). Transfer RNA modifications include RNA editing of adenosine to inosine (A-I), methylation or acetylation of RNA bases, isomerization or reduction of uridine to pseudouridine (Ψ) or dihydrouridine (D) to name a few. These post-transcriptional modifications can occur on the base, or on the ribose sugar backbone of the RNA molecule.

Transfer RNAs can vary in length from 70–90 nucleotides long with RNA modifications occurring at different positions on the iconic clover-leaf secondary structure. The functional roles of tRNA modifications are determined by their position on the clover leaf structure and by the chemical properties of the RNA modification. The functions of these modifications affecting tRNA biogenesis can be divided into three major groups, (i) modifications that affect amino-acylation on the acceptor stem; (ii) modifications on or near the anticodon loop can affect anti-codon binding, wobble base pairing and frame shifting; and (iii) other positions on D-stem, TΨC stem and variable loop/junction affecting stability, structure, translation and tRNA cleavage/degradation. The positions and types of modifications and the enzymes responsible for mediating them throughout different domains of life have been reviewed extensively in (Phizicky and Hopper 2010; El Yacoubi et al. 2012; Towns and Begley 2012). Here we focus on tRNA modifications and tRNA modifying enzymes investigated in plants.

The identities of RNA modifications present in tRNAs of several plant species (*Arabidopsis*, tobacco, maize, hybrid aspen and wheat) have been investigated using a combination of chromatography and mass spectroscopy techniques on purified tRNAs (Shugart 1972; Jones and Scott 1981; Chen et al. 2010; Hienzsch et al. 2013). In two independent studies on *Arabidopsis* tRNAs, a total of 26 known tRNA RNA modifications were identified and four novel, potentially plant specific RNA modifications (Chen et al. 2010; Hienzsch et al. 2013). In addition, bioinformatics approaches have been used to predict hundreds of RNA base modification sites in *Arabidopsis* miRNAs and tRNAs, based on the ability of certain RNA modifications to introduce mismatches in sequences after reverse transcription (Iida et al. 2009). Evidence has also shown dynamic regulation of tRNA modifications. When comparing different plant tissues, differences in the abundance and types of tRNA modifications were found when comparing different plant tissues and cell cultures (Jones and Scott 1981; Hienzsch et al. 2013) and new and old leaves (Shugart 1972). Recently, the tRNA modification 2'-o-cytosine methylation (Cm) was shown to be increased in response to pathogen infection in *Arabidopsis* (Ramirez et al. 2015). In other organisms, several tRNA modifications were shown to be induced under stress conditions such as oxidative stress (Chan et al. 2010; Chan et al. 2012), nutrient starvation (Preston et al. 2013) and toxins (Hertz et al. 2014).

Identification of RNA modifications present in tRNAs is only half the story – the tRNA modifying enzymes, or 'writers' are just beginning to be characterized in plants. A combination of bioinformatics and reverse genetics approaches have been used to predict and identify tRNA modifying enzymes in *Arabidopsis* (Golovko et al. 2002; Chen et al. 2006; Miyawaki et al. 2006; Pavlopoulou and Kossida 2009; Zhou et al. 2009; Chen et al. 2010; Hu et al. 2010; Mehlgarten et al. 2010; Leihne et al. 2011; Zhou et al. 2013; Burgess et al. 2015; Ramirez et al. 2015). Transfer RNA modifying enzymes have been characterized for mediating base methylations of guanine and cytosine residues, and modifying 2′-O-ribose methylations.

RNA methylation of guanine and cytosine residues commonly occurs in *Arabidopsis* tRNAs and have roles in mediating RNA structure and stabilization through for example Mg^{2+} binding (Chen et al. 1993; David et al. 2016). Based on homologous genes in yeast, three guanosine transfer RNA methyltransferase (TRM) enzymes have been identified in *Arabidopsis*, namely AtTRM10 (At5g47680), AtTRM11 (At3g26410) and AtTRM82 (At1g03110), which mediate m^1G, m^2G and m^7G in tRNAs, respectively (Chen et al. 2010). Of these, a biological role in plant development was only identified for AtTRM11, as the mutant showed an early-flowering phenotype. 5-methylcytosine in tRNAs is mediated by *Arabidopsis* transfer RNA aspartic acid methyltransferase 1 (TRDMT1, At5g25480) at position 38 and by tRNA specific methyltransferase 4B (TRM4B, At2g22400) at the variable loop/TΨC stem junction (Goll et al. 2006; Burgess et al. 2015). Loss of both TRDMT1 and TRM4B results in increased sensitivity to the antibiotic Hygromycin B, suggesting roles for these modifications in translation (Burgess et al. 2015). Similar functions were found for TRM4 in yeast (Wu et al. 1998) and translation efficiency was reduced in mammals (Tuorto et al. 2012). Another tRNA methylation modification is 2′-O-ribose methylation. Recently, an *Arabidopsis* homolog of yeast TRM7 (At5g01230), a 2′-O-ribose methyltransferase, was identified to be required for efficient immune response to *Pseudomonas syringae* (Ramirez et al. 2015).

Several modifications in the anticodon loop fine-tune translation by reducing frame shift mutations and mediating codon binding stringency at the third 'wobble' base pair position. RNA editing of adenosine to inosine (A-I) at the first position of the anticodon allows a single tRNA to decode multiple codons for the same amino acid, because I can base pair with A, C or U. RNA editing (A-I) by AtTAD1 (homologous to yeast Tad1p tRNA-specific adenosine deaminase) (At1g01760)

at the position 3'-adjacent to the anticodon in nuclear $tRNA^{Ala(AGC)}$ has been shown to be required for efficient translation under stress conditions, as *Arabidopsis attad1* mutants have reduced biomass when exposed to heat and cold stress treatments (Zhou et al. 2013). The molecular function of this specific RNA editing event is unclear. A conserved multi-protein Elongator complex mediates acetylation of histones and tRNA wobble uridine modifications (Mehlgarten et al. 2010). Four components of the Elongator complex have been characterized in plants, demonstrating roles for the Elongator complex in ABA and oxidative stress response in *Arabidopsis* (Chen et al. 2006; Zhou et al. 2009). In addition, Elongator mutants such as *atelp1* (At5g13680) display pleiotropic growth defects (Chen et al. 2010).

RIBOSOMAL RNA MODIFICATIONS AND FUNCTIONS IN *ARABIDOPSIS*

Ribosomes are multi-subunit complexes of non-coding ribosomal RNAs and proteins. In eukaryotes, three of the four rRNAs present in the small and large rRNA subunits are encoded in a single, polycistronic, pre-rRNA transcript. Multiple processing steps involving cleavage and RNA modifications are required for maturation and assembly of the rRNAs with ribosomal proteins (Henras et al. 2015). Ribosomal RNA modifications tend to be clustered around conserved structural and functional regions of the ribosomes such as the peptidyl transferase center (PTC) and are required for efficient translation (Decatur and Fournier 2002). Ribosomal RNAs contain three broad types of RNA modifications, ribose methylation, pseudouridylation and several types of base methylations (e.g. m^5C, m^3U, m^6A) reviewed in (Decatur and Fournier 2002; Baxter-Roshek et al. 2007).

The most abundant rRNA modifications are Ψ and 2′-O-ribose methylations. The majority of these rRNA modifications are mediated by small nucleolar ribonucleoprotein complexes (snoRNPs) composed of multiple conserved proteins and a small nucleolar RNA (snoRNA), which directs sequence-specific targeting. These two modifications are guided by two different classes of snoRNAs, (i) box-C/D snoRNAs which guide 2′-O-ribose methylations mediated by the methyltransferase NOP1 (yeast)/Fibrillarin (human) and (ii) box-H/ACA snoRNAs which direct conversion of uridine to pseudouridine by Cbf5/NAP57/Dyskerin (human) (Kiss 2001; Brown et al. 2003). Three genes encoding homologues of the essential yeast and human Fibrillarin 2′-O-ribose methyltransferase were identified in *Arabidopsis*, AtFIB1 (At5g52470), AtFIB2 (At4g25630) and AtFIB3 (At5g52490) (Barneche et al. 2000; Pih et al. 2000). Of these three genes, transcripts were only detected from AtFIB1 and AtFIB2, and both proteins are able to partially complement a conditional yeast NOP1/Fibrillarin mutant. This suggests that the *Arabidopsis* 2′-O-ribose methyltransferase snoRNPs might be heterogeneous, and contain either AtFIB1 or AtFIB2, and these different snoRNPs may have specialized functions in plants. Similarly, removal of rRNA pseudouridylation in yeast and *Arabidopsis* by deletion of CBF5, is also lethal (Lermontova et al. 2007) while defects in the human homolog Dyskerin result in dyskeratosis congenita, a disease characterized by abnormal skin pigmentation and bone marrow failure (Heiss et al. 1998). Moreover, patients with this condition were recently found to have reduced Ψ in rRNA and the ncRNA telomerase component TERC (Telomerase RNA component) (Schwartz et al. 2014a). Another chloroplast specific rRNA Ψ synthase was identified in *Arabidopsis* in a suppressor screen for mutants complementing a chloroplast variegation mutation, SUPPRESSOR OF VARIAGATION1 (SVR1, At2g39140) (Yu et al. 2008). *Arabidopsis svr1* mutants are small and pale green, with defects in chloroplast rRNA processing and translation. SVR1 is predicted to target chloroplast rRNA Ψ independently of a snoRNA guide in a similar manner to other tRNA and mitochondrial rRNA Ψ synthases from yeast and bacteria (Ansmant et al. 2000).

Unlike most of the rRNA Ψ and 2′-O-ribose methylations, which are catalyzed by snoRNPs, the rRNA base methylations are all performed by site-specific base methyltransferases. The nuclear large subunit 25S rRNA in yeast and *Arabidopsis* contains two m^5C sites, which are methylated by RNA methyltransferases RCM1 (rRNA cytosine methyltransferase 1) and NOP2 (nucleolar protein 2) (Sharma et al. 2013; Gigova et al. 2014). The two methylation sites have roles in antibiotic sensitivity and rRNA biogenesis and processing in yeast, respectively (Hong et al. 1997; Sharma et al. 2013). In *Arabidopsis* the RCM1 homolog, NOP2/Sun domain protein 5 (NSUN5), was found to methylate the orthologous position in 25S rRNA (Burgess et al. 2015). The second m^5C site unexpectedly remained unchanged in single mutants for all three *Arabidopsis* NOP2 homologs, NOP2A (At5g55920), NOP2B (At4g26600) and NOP2C (At1g06560), as *nop2a* mutants have a leaf phenotype (Fujikura et al. 2009; Burgess et al. 2015). The unchanged methylation level in the single mutants may suggest functional redundancy (Burgess et al. 2015). Two adjacent adenosines are N-6 dimethylated (m_2^6A) in small subunit rRNAs of eukaryotes and prokaryotes by adenosine dimethyl transferase 1 (DIM1) homologs. Similar to the case of NOP2, the *Arabidopsis* genome encodes three rRNA dimethyl transferase enzymes: DIM1A (At2g47420), the nuclear 18S rRNA dimethyl transferase required for organized root growth and epidermal patterning (Wieckowski and Schiefelbein 2012), DIM1B (At5g66360), the mitochondrial rRNA dimethyl transferase (Richter et al. 2010) and DIM1C/PALEFACE1 (At1g01860), which is located in the chloroplast and is required for chloroplast development in the cold (Tokuhisa et al. 1998).

In organisms such as yeast, with only one copy of DIM1 and NOP2, loss of either of these enzymes results in lethality (Lafontaine et al. 1994; Hong et al. 1997). Surprisingly, the presence of catalytically inactivated modifying enzymes rescues the phenotype in several organisms, suggesting other important roles for DIM1 and NOP2 in ribosome biogenesis (Lafontaine et al. 1995; King and Redman 2002; Zorbas et al. 2015). *Arabidopsis dim1a* and *nop2a* mutants both display small, malformed leaves, slow growth and other phenotypes (Fujikura et al. 2009; Wieckowski and Schiefelbein 2012), reminiscent of many other *Arabidopsis* mutants with roles in rRNA biogenesis and of ribosomal protein mutants (Nishimura et al. 2005; Byrne 2009; Abbasi et al. 2010), pointing to additional functions besides RNA methylation for these proteins in plants. Another predicted rRNA m^5C methyltransferase is *Arabidopsis* RNA methyltransferase (RNMT, At3g13180), which is related to the bacterial Fmu 16S rRNA methyltransferase (Pavlopoulou and Kossida 2009).

Arabidopsis rnmt mutants have reduced global cytosine methylation, however, the specific nucleotide position is yet to be identified (Hebrard et al. 2013). In addition to m^5C and $m_2{}^6A$, *Arabidopsis* rRNA also contains several m^6A base methylations (Wan et al. 2015).

MESSENGER RNA AND OTHER NON-CODING RNA MODIFICATIONS

In the following sections we review and discuss the RNA modifications discovered in *Arabidopsis*, animal, yeast and bacterial epitranscriptomes to date and their diverse functions. Recently, two modifications have been identified transcriptome-wide using direct detection methods, m^6A (Immunoprecipitation and next generation sequencing) and m^5C (RNA Bisulfite sequencing) in plants (Luo et al. 2014; Li et al. 2014b; Wan et al. 2015; David et al. 2016). The first discovered and globally most abundant RNA modification, pseudouridine (Ψ) has been mapped transcriptome-wide by several recent studies in mammals and yeast (Carlile et al. 2014; Lovejoy et al. 2014; Schwartz et al. 2014a; Li et al. 2015). Although Ψ sites have not been mapped in plants transcriptome-wide to date, the enzymatic functions required for transcriptome-wide Ψ have been investigated in *Arabidopsis* (Lermontova et al. 2007; Yu et al. 2008; Chen et al. 2010). Interestingly, these high throughput studies have revealed m^6A, m^5C and Ψ to show distinct distribution patterns along mRNA transcripts and are associated with specific functions as discussed in the following sections (Figure 1). Additional modifications in the *Arabidopsis* epitranscriptome such as 3-methyl cytosine (m^3C) and 1-methyl guanosine (m^1G) have been computationally predicted transcriptome-wide based on common nucleotide substitution and reverse transcription errors caused by these modifications during RNA-seq library preparation (Ryvkin et al. 2013; Vandivier et al. 2015). The modifications m^6A, m^5C and Ψ are unable to be detected using this method, as they do not alter Watson-Crick base pairing.

N^6-METHYLADENOSINE (M^6A)

m^6A 'writers'

Although the presence of m^6A in mRNAs was first discovered in the 1970s (Desrosiers et al. 1974; Perry and Kelley 1974), many questions still remain unanswered about the roles of m^6A in protein coding transcripts. While Ψ is the globally most abundant RNA modification, m^6A is the most highly abundant RNA modification in mRNAs and is enriched in poly-adenylated RNA fractions in plants and animals and has recently also been identified in bacterial mRNAs (Zhong et al. 2008; Meyer et al. 2012; Deng et al. 2015). A multi-protein complex mediates these m^6A sites. The catalytic core is composed of a heterodimer of methyltransferase like 3 (METTL3) and methyltransferase like 14 (METTL14) in mammals (Bokar et al. 1994; Liu et al. 2014). Recently, the mammalian splicing factor Wilm's tumor 1 associating protein (WTAP) and KIAA1429 were identified as additional components of the m^6A 'writer' complex (Ping et al. 2014; Schwartz et al. 2014b). WTAP may have roles in targeting the m^6A activity of METTL3 and METTL14, in a site-specific manner, as m^6A sites were mediated by WTAP dependent or independent mechanisms (Schwartz et al. 2014b).

Likewise, the *Arabidopsis* m^6A 'writer' complex contains the adenosine methyltransferase MTA (At4g10760), which is predicted to form a heterodimer with MTB (At4g09980) (Bujnicki et al. 2002; Zhong et al. 2008). The *Arabidopsis* homolog of mammalian WTAP is known as *Arabidopsis thaliana* FKBP12 interacting protein 37 (AtFIP37, At3g54170), and was identified as a binding partner of MTA several years prior to similar studies in mammals (Faure et al. 1998; Zhong et al. 2008). This is shown in Figure 2A. Further studies are required to identify additional components and interacting proteins. All three known components of the *Arabidopsis* m^6A writer complex are essential, as loss results in embryo lethality (Bujnicki et al. 2002; Vespa et al. 2004; Zhong et al. 2008). The lethality of *mta* mutants in *Arabidopsis* can be rescued by expressing MTA during embryo development using the *ABI3* promoter (Bodi et al. 2012). Use of this system allowed investigation of the requirement for m^6A in vegetative development, floral architecture and cell specification. The importance of m^6A methylation for gene regulation is underscored by disorders caused by loss of m^6A 'writer' complex components in human, yeast, mouse and fly (Clancy et al. 2002; Hongay and Orr-Weaver 2011; Bodi et al. 2012; Wang et al. 2014b; Chen et al. 2015b).

In order to elucidate why m^6A is essential to plant development, three independent studies have mapped m^6A epitranscriptomes in *Arabidopsis* and rice (Luo et al. 2014; Li et al. 2014b; Wan et al. 2015). As reported in the earlier studies in mammalian mRNAs, m^6A sites were found to occur all along transcripts, with low signals observed across coding sequences and high enrichment in 3′UTRs and around stop codons (Dominissini et al. 2012; Meyer et al. 2012). Specifically in plants, there was a slight enrichment for m^6A peaks at the start codon (Luo et al. 2014; Li et al. 2014b). However, a more recent study with greater sequencing depth and stringency conditions for m^6A antibody binding, did not detect enrichment at start codons in *Arabidopsis* (Wan et al. 2015). Thousands of methylated transcripts were detected in different tissue types in *Arabidopsis* and rice and even in different *Arabidopsis* ecotypes. While many sites were specific to a particular tissue or ecotype, a large number of these sites were also conserved, even between animals and plants (Luo et al. 2014; Wan et al. 2015). As the deposition patterns and even specific m^6A sites are conserved, the targeting mechanisms of the m^6A 'writer' complex and functions of m^6A are also likely conserved (Figure 2A).

In support of this, transcriptome-wide mapping studies of m^6A have confirmed earlier reports that the m^6A 'writer' complex methylates sites within a highly conserved consensus sequence 'RR**A**CH', (R = A/G and H = A/C/U) and this mostly occurs in G**A**C or less commonly in the A**A**C context (Wei and Moss 1977; Csepany et al. 1990). This consensus sequence is conserved in yeast, mammals and plants (Dominissini et al. 2012; Meyer et al. 2012; Schwartz et al. 2013; Luo et al. 2014; Li et al. 2014b; Wan et al. 2015). Interestingly, this is not the case for prokaryotes as unique distribution patterns and potential targeting were discovered in bacteria, as unlike animals and plants, m^6A is enriched in coding sequences and at a

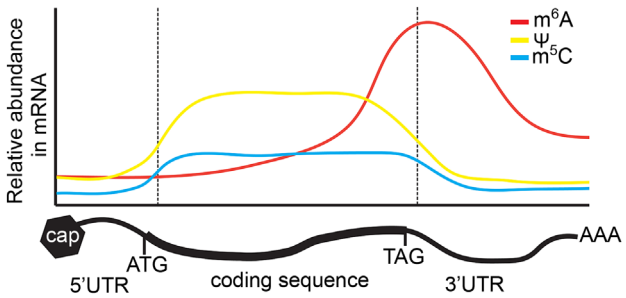


Figure 1. RNA modifications in messenger RNAs have distinct deposition patterns
Shown is a pictorial representation of relative abundance of the RNA modifications m^6A, m^5C and Ψ along mRNA transcripts. These representations are based on transcriptome-wide RNA bisulfite sequencing data for m^5C and antibody data for m^6A in animals and plants. The Ψ abundance is based on a combination of Ψ-seq and antibody enrichment data from animals. m^6A is lowly abundant along coding sequences and enriched at long last exons and at the start of 3′UTR's. While the majority of m^5C sites are detected in the coding sequence of mRNA transcripts, m^5C sites are statistically enriched in 3′UTR's. For Ψ, the modified sites are evenly distributed along the coding sequence, but are statistically underrepresented in 5′ UTRs.

novel 'GCC**A**G' consensus sequence (Deng et al. 2015). The mechanism of targeting and the functional significance of the distributions of m^6A methylation on mRNA remain to be elucidated. While the highly conserved eukaryotic consensus sequence 'RR**A**CH' is present many times in the transcriptome, the mechanism for determining which of these sites are methylated, remains unknown.

m^6A 'erasers' – reversible RNA methylation

One intriguing mechanism for regulating m^6A deposition is through the active removal of m^6A in mRNAs. Two m^6A demethylases have been characterized in mammals, namely Fat mass and obesity associated protein (FTO) and Alkylation repair homologue protein 5 (ALKBH5) (Jia et al. 2011; Zheng et al. 2013). These demethylases are part of the *Escherichia coli* ALKB dioxygenase homologs (ALKBH) family. The founding member, *E. coli* ALKB mediates oxidative demethylation of nucleic acid bases in DNA and RNA (Aas et al. 2003). Based on sequence homology, 13 ALKBH family proteins were predicted in *Arabidopsis* (Mielecki et al. 2012). These proteins showed diverse subcellular localizations, suggesting specialized functions in different cell compartments and hence potential layers of regulation for m^6A demethylation in plants (Mielecki et al. 2012). While the identity of the *Arabidopsis* m^6A demethylase(s) are still undetermined, they are expected to cause gross development

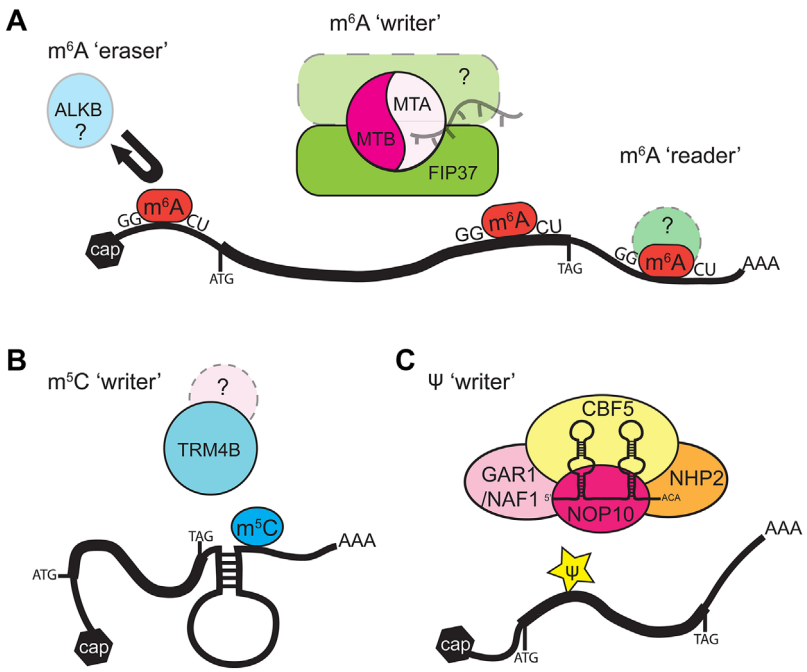


Figure 2. Distinct catalytic and targeting mechanisms of different RNA modifications in *Arabidopsis*
(**A**) The predicted *Arabidopsis* m^6A 'writer' complex is composed of a heterodimer of MTA and MTB, bound to AtFIP37 and potentially other, uncharacterized proteins. In mammals, miRNAs are able to guide the m^6A 'writer' complex, however, it is not known if this targeting mechanism is conserved in plants. Potential m^6A 'erasers' and 'readers' have been predicted in *Arabidopsis* and await further characterization. (**B**) A model for targeting of m^5C methylation by TRM4B, based on RNA structure and potentially the presence of other RNA modifications. (**C**) Proposed H/ACA snoRNP Ψ 'writer' complex in *Arabidopsis* contains a guide H/ACA box snoRNA and the proteins GAR1 (At3g03920/At5g18180) or NAF1 (At1g03530), NHP2 (At5g08180), NOP10 (At2g20490) and the Ψ synthase AtCBF5 (At3g57150).

defects, reminiscent of their animal homologs. In mice, loss of FTO leads to defects in alternative splicing and adipogenesis (Zhao et al. 2014), while loss of ALKBH5 affects mRNA processing in human cells and leads to male infertility in mice (Zheng et al. 2013). While the m^{6}A 'writer' complex and demethylase 'erasers' act in concert to dynamically regulate m^{6}A, additional RNA binding proteins or 'readers' are thought to decide the fate of m^{6}A methylated transcripts.

m^{6}A 'readers' – consequences for m^{6}A on RNAs

The presence of m^{6}A can influence RNA metabolism by regulating binding of modified RNA with proteins (m^{6}A 'readers') and can also alter local RNA structure, leading to alternate outcomes for methylated and non-methylated transcripts. Several classes of m^{6}A 'reader' proteins have been identified in animals, such as YTH domain proteins, serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) (Dominissini et al. 2012; Schwartz et al. 2013; Zhao et al. 2014; Wang et al. 2014a; Schwartz et al. 2014b; Alarcon et al. 2015a). Potential homologs of these m^{6}A 'readers' have been identified in plants, suggesting conserved functions (Lorkovic and Barta 2002).

YTH domain containing proteins have been identified as a class of RNA binding proteins that preferentially bind m^{6}A methylated RNA in mammals and yeast (Dominissini et al. 2012; Schwartz et al. 2013; Wang et al. 2014a; Schwartz et al. 2014b). The first m^{6}A 'reader' to be characterized was YTH Domain Family 2 (YTHDF2), which was shown to bind thousands of m^{6}A containing mRNAs in the cytoplasm and deliver them to processing bodies for degradation (Wang et al. 2014a). This discovery explains the negative correlation of m^{6}A with mRNA abundance in both plants and animals as m^{6}A is used as a mark for rapid turn-over of mRNAs (Li et al. 2014b; Schwartz et al. 2014b; Wang et al. 2014a; Wan et al. 2015). The sub-cellular locations of m^{6}A readers is crucial for the outcome of m^{6}A on RNAs and this is clearly demonstrated by YTHDF2 (Zhou et al. 2015). While cytoplasmic YTHDF2 leads to mRNA decay, heat shock induces YTHDF2 to re-localize to the nucleus. YTHDF2 is then able to compete with the nuclear 'eraser' FTO for binding of m^{6}A sites, leading to increased 5'UTR methylation of newly transcribed, heat stress responsive mRNAs. The increased methylation leads to increased translation initiation independent of the 5' cap, allowing selective mRNA translation under heat shock stress. In addition, the m^{6}A reader YTHDF1 has also been shown to increase translation initiation of transcripts harboring m^{6}A sites (Wang et al. 2015). This occurs in the cytoplasm, leading to competition with the mRNA degrading cytoplasmic YTHDF2. This competition is thought to allow fast responses and regulation of mRNA abundance and translation through m^{6}A methylation.

Plant proteins containing this conserved YTH domain are expected to mediate similar functions for m^{6}A in RNA. The *Arabidopsis* genome encodes 13 predicted YTH domain containing proteins, which may be responsible for 'reading' the m^{6}A code and regulating RNA metabolism (Li et al. 2014a). One such protein is the *Arabidopsis* homologue of Cleavage and Polyadenylation Specificity Factor 30 (CPSF30, At1g30460), which is required in plants and mammals for polyadenylation and 3'end formation (Thomas et al. 2012; Chan et al. 2014). Intriguingly, while the presence of the YTH domain of *Arabidopsis* AtCPSF30 is dependent on alternative splicing, the YTH domain is completely absent in yeast and mammalian CPSF30 homologs (Delaney et al. 2006; Hunt et al. 2012; Chakrabarti and Hunt 2015). Furthermore, AtCPSF30 has also been shown to be involved in oxidative stress responses (Zhang et al. 2008) and is required for programmed cell death and immunity in *Arabidopsis* (Bruggeman et al. 2014). These functions are independent of the YTH domain and raise questions about the possible roles of YTH domain-containing AtCPSF30 in regulating m^{6}A containing RNAs.

In addition to YTH domain proteins, two other classes of potential m^{6}A 'readers', namely SR proteins and hnRNPs, have been investigated. One SR protein, SR Splicing Factor 2 (SRSF2), was shown to preferentially bind mRNAs containing m^{6}A sites, leading to increased inclusion of target exons during splicing when the 'eraser' FTO is depleted (Zhao et al. 2014). SRSF2 RNA binding sites tend to overlap with m^{6}A sites, however, it is unclear if SRSF2 binds m^{6}A directly, or indirectly through interactions with other proteins. Recently, SRSF3 and SRSF10 were found to competitively bind YTHDC1, which directly binds m^{6}A, to regulate mRNA splicing (Xiao et al. 2016). While the interaction between SRSF3 and YTHDC1 promotes SRSF3 binding to RNA target sites, YTHDC1 binding of SRSF10 inhibits SRSF10 binding to RNA target sites. In combination, these events result in exon inclusion, while successful SRSF10 binding to RNA results in exon exclusion. Similarly, over expressing the predicted *Arabidopsis* SRSF2 ortholog AtSRp30 (At1g09140) demonstrated its function in regulating splicing (Lopato et al. 1999). Further studies are required to determine if m^{6}A deposition directly or indirectly affects the activities of the eighteen SR proteins in *Arabidopsis* (Lorkovic and Barta 2002). Other m^{6}A 'readers' include hnRNPs, which have diverse roles in RNA processing and export (Lorkovic et al. 2000).

Recently, m^{6}A methylation was shown to be required for the biogenesis and function of a subset of miRNAs, and this is mediated in part through the m^{6}A 'reader' HNRNPA2B1 and m^{6}A 'anti-reader' human antigen R (HuR) in animals. miRNA processing and abundance is deregulated when either the m^{6}A 'writer' METTL3 or the m^{6}A 'eraser' FTO were perturbed, demonstrating a role for m^{6}A in miRNA biogenesis (Berulava et al. 2015; Alarcon et al. 2015b). The m^{6}A mark in miRNAs is important, as nuclear HNRNPA2B1 binds a subset of m^{6}A containing pri-miRNAs and recruits the Microprocessor complex to cleave pri-miRNAs into pre-miRNAs (Alarcon et al. 2015a). In addition, m^{6}A methylation is required for efficient regulation of a sub-set of miRNA target transcripts. m^{6}A methylation can aid miRNA driven degradation of mRNAs, as m^{6}A blocks binding of the m^{6}A 'anti-reader' HuR, allowing miRNAs access to their target sites in mRNAs (Wang et al. 2014b). While m^{6}A methylation regulates the biogenesis of a subset of miRNAs, some miRNAs are also able to affect targeting of m^{6}A methylation. Artificial and endogenous miRNAs were recently shown to target m^{6}A deposition and increase m^{6}A abundance by guiding and modulating METTL3 binding to mRNAs (Chen et al. 2015b). This proposed mechanism of miRNAs guiding m^{6}A sites is supported as animal miRNA target sites are highly enriched at m^{6}A methylated regions, however, no such correlation was identified in *Arabidopsis* (Luo et al. 2014; Chen et al. 2015b). Is another class of small guide RNAs mediating m^{6}A targeting in plants? These intriguing findings raise many questions about

possible roles of m^6A in regulating miRNAs in plants, and how the plant m^6A 'writer' complex is targeted to mRNAs.

Another interesting example of an hnRNP m^6A 'reader' is HNRNPC. HNRNPC does not bind m^6A, however, it requires m^6A methylation of mRNA and lncRNA targets such as Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1) to alter local RNA structure in order to facilitate RNA binding (Liu et al. 2015). This RNA remodeling, or 'm^6A-switch', is achieved through the ability of m^6A to disrupt adenosines from forming non-Watson-Crick G:A base pairs and also destabilizes A:U base pairs (Roost et al. 2015). It remains to be seen if the nine hnRNPs in *Arabidopsis* genome also show such diverse interactions with m^6A as a dance partner (Lorkovic et al. 2000).

Additional functions for m^6A methylation in reprogramming, organ differentiation and cell division functions are conserved, as indicated by the *Arabidopsis mta* phenotype and the roles of m^6A in inducing pluripotent cells in mammals and sporulation in yeast (Clancy et al. 2002; Bodi et al. 2012; Wang et al. 2014b; Chen et al. 2015b). Furthermore, m^6A was shown to regulate the mammalian circadian clock, as loss of RNA m^6A methylation slows RNA processing resulting in delayed release of mature clock transcripts from the nucleus (Fustin et al. 2013). m^6A may also play a role in regulating the plant circadian clock, as several transcripts regulating the *Arabidopsis* clock were highly methylated (Wan et al. 2015). Functions for m^6A in splicing intron retention, polyadenylation, microRNA regulation, reprogramming and stress responses in plants warrants further investigation. Recently developed, single nucleotide resolution approaches to mapping m^6A epitranscriptomes will enable further characterization of the functions of this mark in plants (Linder et al. 2015; Chen et al. 2015a).

5-METHYLCYTOSINE (M^5C)

m^5C 'writers'

While the functions of m^5C as an epigenetic mark in DNA have been studied intensively, the role of m^5C in RNA is less well studied. The importance of m^5C has been established for tRNAs and rRNAs (Motorin et al. 2010), but functions are still being investigated for other RNAs such as mRNAs and lncRNAs. m^5C was first identified transcriptome-wide using RNA Bisulfite sequencing (bsRNA-seq) in human (HeLa) cells, uncovering over 10,000 m^5C sites (Squires et al. 2012). This prompted the development of additional techniques that enrich the direct RNA targets of specific RNA methyltransferases using RNA immunoprecipitation (Khoddami and Cairns 2013; Hussain et al. 2013a). Recently, the *Arabidopsis* m^5C landscape was mapped using bsRNA-seq in several tissue types and RNA methyltransferase mutants, identifying hundreds of m^5C sites (David et al. 2016).

Together, these studies identified two m^5C 'writers' that catalyze methylation in mRNAs and other classes of RNAs; the first RNA methyltransferase is tRNA specific methyltransferase 4 (TRM4) otherwise known as NOP2/Sun domain protein 2 (NSUN2), in yeast and animals respectively. NSUN2 plays broad roles in many organisms for mediating oxidative stress tolerance and balancing stem cell self-renewal and differentiation. This is demonstrated in *nsun2* mutant mice presenting with epidermal differentiation defects, male infertility and small size, which is thought to be due to a reduction in stem cell proliferation (Blanco et al. 2011; Hussain et al. 2013b). Furthermore, NSUN2 depletion in humans leads to mild microcephaly, short stature and neurological disorders (Abbasi-Moheb et al. 2012; Khan et al. 2012; Martinez et al. 2012; Fahiminiya et al. 2014). Loss of *nsun2* leads to increased tRNA cleavage under oxidative stress and these cleavage products are thought to cause these neuro-developmental disorders (Blanco et al. 2014). These roles are also conserved in plants, as *Arabidopsis trm4b* mutants display shorter primary roots, which is linked to a reduced capacity for cells to divide in the root meristem (David et al. 2016). Furthermore, *trm4b* mutants are also more sensitive to oxidative stress and have reduced stability of non-methylated tRNAs. However, it is difficult to tease apart the contributions of tRNA and mRNA methylations to these biological functions, as NSUN2/TRM4B methylates both these classes of RNAs.

The second m^5C 'writer' shown to target mRNAs is Transfer RNA aspartic acid methyltransferase 1 (TRDMT1) also known as DNA methyltransferase 2 (DNMT2). TRDMT1 was previously thought to methylate DNA, due to its structural similarity to DNA methyltransferases, however it is now regarded as an RNA methyltransferase (Goll et al. 2006). In plants and animals, depletion of TRDMT1 is not phenotypically evident under controlled conditions (Goll et al. 2006). However, the functions of TRDMT1 become apparent under stress conditions such as oxidative and heat stress in *Drosophila* (Schaefer et al. 2010). Stress induced cleavage of tRNAs in TRDMT1 mutants also leads to inhibition of Dicer-2 functions (Durdevic et al. 2013b). Furthermore, TRDMT1 is required for efficient immune response against viruses in *Drosophila* (Durdevic et al. 2013a). In contrast, depletion of TRDMT1 in zebrafish leads to gross morphological defects (Rai et al. 2007). NSUN2 mediates many more m^5C sites in the transcriptome than TRDMT1. Only two TRDMT1 mRNA targets were identified in human cells, type I cytokeratin *KRT18* mRNA and *KRT18* pseudogene mRNA and this methylation was not conserved in mouse (Khoddami and Cairns 2013). This minor role for TRDMT1 in mediating m^5C transcriptome-wide seems to be conserved in plants, as only tRNA targets were identified (Burgess et al. 2015; David et al. 2016).

LOCATIONS, FUNCTIONS AND TARGETING OF M^5C

In order to determine potential functions of m^5C in RNA, transcriptome-wide deposition patterns of this mark were analyzed in human cancer cells. Methylated sites are statistically enriched in ncRNAs compared to mRNAs (Squires et al. 2012). Within mRNAs, m^5C sites are observed in higher numbers than expected for untranslated regions and are relatively depleted in coding regions, when normalized for length and sequence coverage (Squires et al. 2012). Moreover, m^5C candidate sites in 3′UTRs are associated with binding regions for the Argonaute I–IV proteins, which are involved in miRNA mediated decay and translational inhibition, suggesting possible roles for m^5C in mediating miRNA activity (Squires et al. 2012). Although further experiments are required to clearly determine the m^5C and Argonaute association.

Additional functions for m^5C in increasing mRNAs half-life have been proposed, as synthetic m^5C methylated mRNAs exhibit increased stability (Warren et al. 2010). This does not seem to be the case for the majority of mRNAs, as methylation levels do not strongly correlate with gross changes in transcript abundance in mammals or plants (Hussain et al. 2013a; David et al. 2016). Furthermore, no major changes in global mRNA abundance were observed in mouse *nsun2* mutants (Tuorto et al. 2012; Hussain et al. 2013b).

In contrast to mRNAs, several functions for m^5C have been investigated in ncRNAs. Vault ncRNAs were identified as NSUN2-specific m^5C targets (Hussain et al. 2013a). Loss of m^5C in vault ncRNAs leads to processing into small RNAs which can be incorporated into Argonaute complexes to regulate genes, in a manner similar to miRNAs. In addition to roles for m^5C in small ncRNAs, functions for this modification have been elucidated for long ncRNAs. The 5′ A-region of the lncRNA X-inactive specific transcript (XIST) contains five m^5C sites, which were shown to inhibit binding of the Polycomb repressive complex 2 (PRC2) *in vitro* (Amort et al. 2013). It remains to be determined if the PRC2 complex acts globally as an m^5C 'anti-reader' in both plants and animals.

Analysis of m^5C transcriptome-wide has shown that only approximately 0.4% of cytosines are methylated in mRNA, suggesting precise targeting of m^5C to select target sites (Squires et al. 2012). In archaea, m^5C was located in a consensus motif of AU**C**GANGU in mRNAs, providing a potential targeting mechanism for archaeal m^5C 'writers' (Edelheit et al. 2013). In contrast, no such consensus target sequences were identified for m^5C sites in animals or plants (Squires et al. 2012; Hussain et al. 2013a; David et al. 2016). As a general consensus sequence has not been identified, it is hypothesized that additional factors such as local RNA structure and RNA binding proteins may regulate the site selection of TRM4 and TRDMT1 (Figure 2B).

Many questions remain unanswered such as the targeting mechanism of m^5C 'writers' and the functions of m^5C in mRNAs and other non-coding RNAs. The identification of potential m^5C 'readers' and 'erasers' using techniques such as m^5C RNA bait to immuno-precipitate m^5C binding proteins, similar to those performed for m^6A should lead to future insights into how m^5C sculpts the epitranscriptome.

PSEUDOURIDYLATION (Ψ)

Isomerization of uridine to Ψ was the first RNA modification to be discovered, and is also the most abundant (Charette and Gray 2000; Ge and Yu 2013). As discussed earlier, Ψ is common in tRNAs and rRNAs and also in spliceosomal snRNAs, however, it is an open question whether Ψ is present on *Arabidopsis* mRNAs. Recently, four research groups independently investigated Ψ transcriptome wide at single-nucleotide resolution in yeast, human and mouse cells using modified approaches to Ψ-sequencing (Carlile et al. 2014; Lovejoy et al. 2014; Schwartz et al. 2014a; Li et al. 2015). Using these transcriptome-wide approaches, they were able to confirm known Ψ sites and cognate Ψ synthases in tRNAs, rRNAs, snRNAs and snoRNAs and extend the known sites to mRNAs and lncRNAs such as XIST and MALAT1. As the components required for Ψ are conserved in *Arabidopsis*, it seems more than likely that Ψ also occurs in plant mRNAs (Lermontova et al. 2007; Yu et al. 2008; Chen et al. 2010) (Figure 2C).

Ψ synthases are targeted to specific sites in RNAs through two mechanisms (1) snoRNA guided H/ACA snoRNPs containing CBF5/Dyskerin and (2) snoRNA independent Pseudouridine synthases (PUS). Using a combination of deletion and knock down mutants for PUS proteins and CBF5/Dyskerin in yeast and human, mRNA Ψ sites were found to be dependent on Ψ synthases using both snoRNA dependent and independent mechanisms.

While the targeting mechanism for Ψ by H/ACA snoRNPs is based on the snoRNA guide, and synthetic snoRNAs have been successfully designed to target Ψ at novel sites (Karijolich and Yu 2011), the targeting mechanisms of PUS proteins to RNA are less understood. Transcriptome-wide identification of Ψ sites in several yeast PUS deletion mutants allowed analysis and confirmation of sequence consensus sites preferred by specific PUS enzymes. In particular, yeast PUS4 mediated Ψ occurs at 'GU**Ψ**C/NANNC' consensus sites, while yeast PUS7 Ψ sites generally occur at the consensus 'UG**Ψ**A/R'. Not all sites with these consensus sequences are modified, suggesting other, additional factors mediating targeting. For example, the structure, as opposed to the sequence, of the $tRNA^{Ser}$ anticodon and TΨC stem loops were required for human PUS1 targeting (Sibert and Patton 2012).

Ψ sites mediated by these enzymes were located all along mRNA transcripts, with no positional bias found in coding sequences in any of the four transcriptome-wide studies (Carlile et al. 2014; Lovejoy et al. 2014; Schwartz et al. 2014a; Li et al. 2015). However, while Ψ sites were under represented in 3'UTRs of yeast and human cervical cancer (HeLa) cells (Carlile et al. 2014), a chemical pulldown method, which enriched for Ψ sites prior to sequencing found that Ψ sites were under represented in 5′UTRs of mouse and human (H36KT) cells (Li et al. 2015). Prior enrichment of Ψ sites enabled the identification of thousands of sites transcriptome-wide, compared to other studies finding only hundreds of sites.

Dynamic regulation of Ψ sites was conserved across species, as tissue specific and stress responsive Ψ sites were identified in animals and yeast. Strong, stimuli-specific patterns of Ψ were induced for heat shock, addition of a viral mimic and oxidative stress (Schwartz et al. 2014a; Li et al. 2015). Ψ sites were also regulated by different cellular growth rates and nutrient availability (Carlile et al. 2014). Interestingly, Schwartz et al. (2014a) show that PUS7 Ψ mediates heat sensitivity in yeast, as yeast mutants have increased heat sensitivity and >200 PUS7 dependent Ψ sites are induced by heat shock. Ψ transcripts were expressed at higher levels in wild type than in PUS7 mutants during heat shock, suggesting a role for Ψ in stabilizing specific mRNAs in stress conditions.

The function of Ψ in mRNAs is unclear, however, Ψ is thought to help stabilize RNAs by promoting base stacking, pairing, and conformational stability. Ψ may also affect the translation of modified mRNAs (Davis 1995). For example, Ψ has been shown to convert nonsense codons into sense codons, thus 'rewiring' the genetic code (Karijolich and Yu 2011). However, the precise role of Ψ in translation is controversial as Ψ has been shown to both aid and inhibit translation in eukaryotic and bacterial systems, respectively, (Kariko et al. 2012; Hoernes et al. 2015). The locations and roles of this modification in *Arabidopsis* mRNAs is yet to be discovered.

OTHER MODIFICATIONS IN THE *ARABIDOPSIS* EPITRANSCRIPTOME

Of the over 100 RNA modifications discovered in RNA, high-throughput methods of detection have been limited to detecting only a small subset of these modifications. Recently, a new hybrid method has been introduced, referred to as HAMR (High-throughput Annotation of Modified Ribonucleotides), which is able to detect and predict modifications that affect Watson-Crick base pairing transcriptome-wide (Ryvkin et al. 2013). HAMR was trained using data from well characterized yeast tRNA modifications to predict the identity of several RNA modifications. Subsequently, HAMR was put to use on the *Arabidopsis* epitranscriptome, and several types of RNA modifications that perturb reverse transcription were predicted in all types of RNA classes (Vandivier et al. 2015).

Three types of RNA-seq datasets were tested and compared in this study, (i) polyadenylated, (ii) small RNAs and (iii) degrading RNA. RNA modifications were enriched in *Arabidopsis* exons and 3'UTRs of uncapped, degrading mRNA and lncRNA transcripts and the same enrichment pattern was detected in two human cell lines, suggesting broad conservation and possible regulatory functions of these RNA modifications. It remains to be determined if RNA modifications are targeted to degrading transcripts, or if the RNA modifications serve as signals to mark transcripts for degradation. In addition, RNA modifications predicted by HAMR in stable mRNAs from the polyadenylated RNA-seq data sets were enriched within introns that were annotated to be alternatively spliced in both plants and humans.

Distributions of specific RNA modifications were specific to different types of RNAs and depended on whether the transcripts were undergoing degradation. For example, degrading mRNA transcripts had much higher predicted levels of dihydrouridylation (D), N^6-isopentenyladenosylation (i^6A) and threonylcarbamoyladenosylation (t^6A) than stable mRNAs. Uncapped, degrading transcripts involved in various stress responses were enriched for HAMR-predicted modifications, suggesting possible roles in gene regulation and stress responses for these mRNA modifications in *Arabidopsis*.

FUTURE DIRECTIONS AND CONCLUSIONS

The four base constituents of RNA are modified by over 100 different RNA modifications. This additional complexity of RNA is essential for basic functions, such as gene regulation and translation. The *Arabidopsis* epitranscriptome has now been mapped for several RNA modifications, which occur in different locations across transcripts, are inducible in response to abiotic and biotic stresses and have diverse roles in plant development, ranging from subtle (m^5C) to dramatic (m^6A) effects on plant growth. While the RNA modifying 'writers' have been investigated in plants, studies on potential 'erasers' and 'readers' are lacking. The *Arabidopsis* genome encodes over 200 RNA binding proteins which serve as potential readers and effectors of outcomes for RNA modifications (Lorkovic and Barta 2002). Furthermore, potential *Arabidopsis* 'erasers' from the ALKBH family are yet to be explored for roles in plant development and mediating dynamic regulation of

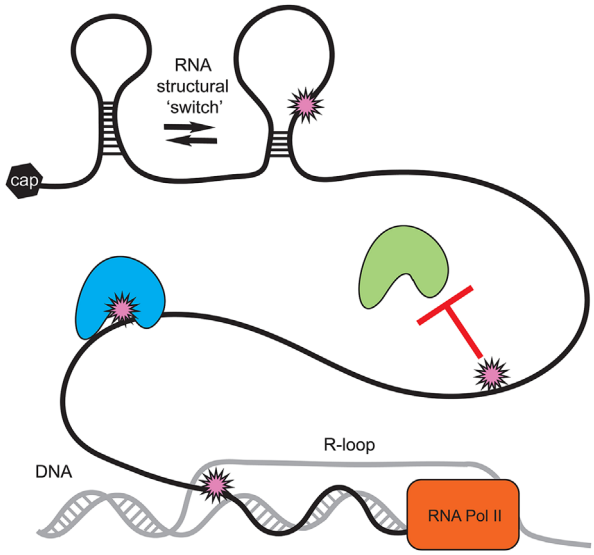


Figure 3. Potential functions of RNA modifications in mediating interactions between nucleic acids and nucleic acids and proteins

RNA modifications have diverse chemical properties and can have different effects on RNA interactions. A spikey, pink ball is used to represent a generic RNA modification. RNA modifications can regulate protein binding to RNA through remodeling of local RNA structure (e.g. 'm^6A switches'), increasing or inhibiting protein binding. In addition, RNA modifications could potentially affect other types of interactions, such as R-loops, which are RNA-DNA hybrids.

RNA modifications (Mielecki et al. 2012). The ALKBH family of dioxygenases has diverse substrate specificities and are not limited to demethylation of adenosine (Aas et al. 2003; Jia et al. 2011). Specific *Arabidopsis* ALKBH family proteins may also remove additional RNA modifications. Further research is needed to elucidate the mechanisms and functional roles of mRNA modifications such as alternative splicing, and stress responses. Using small RNA guides, it is possible to artificially induce and block m^6A and Ψ in mRNAs (Karijolich and Yu 2011; Chen et al. 2015b). This should enable the study of the specific functions of individual RNA modifications. There are many different ways that RNA modifications can affect RNA structure and interactions between RNA, RNA and proteins and even potentially RNA-DNA interactions (Figure 3). The next steps for deciphering the *Arabidopsis* epitranscriptome include Ψ-seq, mapping 2′-O-ribose methylations (Karijolich and Yu 2011; Birkedal et al. 2015), mapping N1-methyladenosine (m^1A) (Dominissini et al. 2016), single-nucleotide resolution mapping of m^6A (Ke et al. 2015), and determining potential reversibility, and the elusive targeting mechanism(s) for RNA modifications.

ACKNOWLEDGEMENTS

This research was supported by ARC grants DP110103805 and FT13100525 awarded to I.S. and an APA and a GRDC PhD top-up scholarship awarded to A.B.

REFERENCES

Aas PA, Otterlei M, Falnes PO, Vagbo CB, Skorpen F, Akbari M, Sundheim O, Bjoras M, Slupphaug G, Seeberg E, Krokan HE (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. **Nature** 421: 859–863

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** 90: 847–855

Abbasi N, Kim HB, Park NI, Kim HS, Kim YK, Park YI, Choi SB (2010) APUM23, a nucleolar Puf domain protein, is involved in pre-ribosomal RNA processing and normal growth patterning in *Arabidopsis*. **Plant J** 64: 960–976

Alarcon CR, Goodarzi H, Lee H, Liu XH, Tavazoie S, Tavazoie SF (2015a) HNRNPA2B1 is a mediator of m(6)A-dependent nuclear RNA processing events. **Cell** 162: 1299–1308

Alarcon CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF (2015b) N6-methyladenosine marks primary microRNAs for processing. **Nature** 519: 482–485

Amort T, Souliere MF, Wille A, Jia XY, Fiegl H, Worle H, Micura R, Lusser A (2013) Long non-coding RNAs as targets for cytosine methylation. **RNA Biol** 10: 1003–1008

Ansmant I, Massenet S, Grosjean H, Motorin Y, Branlant C (2000) Identification of the *Saccharomyces cerevisiae* RNA:pseudouridine synthase responsible for formation of psi(2819) in 21S mitochondrial ribosomal RNA. **Nucleic Acids Res** 28: 1941–1946

Barneche F, Steinmetz F, Echeverria M (2000) Fibrillarin genes encode both a conserved nucleolar protein and a novel small nucleolar RNA involved in ribosomal RNA methylation in *Arabidopsis thaliana*. **J Biol Chem** 275: 27212–27220

Baxter-Roshek JL, Petrov AN, Dinman JD (2007) Optimization of ribosome structure and function by rRNA base modification. **PLoS ONE** 2: e174

Berulava T, Rahmann S, Rademacher K, Klein-Hitpass L, Horsthemke B (2015) N6-adenosine methylation in MiRNAs. **PLoS ONE** 10: e0118438

Birkedal U, Christensen-Dalsgaard M, Krogh N, Sabarinathan R, Gorodkin J, Nielsen H (2015) Profiling of ribose methylations in RNA by high-throughput sequencing. **Angew Chem Int Ed Engl** 54: 451–455

Blanco S, Dietmann S, Flores JV, Hussain S, Kutter C, Humphreys P, Lukk M, Lombard P, Treps L, Popis M, Kellner S, Holter SM, Garrett L, Wurst W, Becker L, Klopstock T, Fuchs H, Gailus-Durner V, de Angelis MH, Karadottir RT, Helm M, Ule J, Gleeson JG, Odom DT, Frye M (2014) Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. **EMBO J** 33: 2020–2039

Blanco S, Frye M (2014) Role of RNA methyltransferases in tissue renewal and pathology. **Curr Opin Cell Biol** 31: 1–7

Blanco S, Kurowski A, Nichols J, Watt FM, Benitah SA, Frye M (2011) The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate. **PLoS Genet** 7: e1002403

Bodi Z, Zhong SL, Mehra S, Song J, Graham N, Li HY, May S, Fray RG (2012) Adenosine methylation in *Arabidopsis* mRNA is associated with the 3 ' end and reduced levels cause developmental defects. **Front Plant Sci** 3: 48

Bokar JA, Rath-Shambaugh ME, Ludwiczak R, Narayan P, Rottman F (1994) Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. **J Biol Chem** 269: 17697–17704

Brown JW, Echeverria M, Qu LH (2003) Plant snoRNAs: Functional evolution and new modes of gene expression. **Trends Plant Sci** 8: 42–49

Bruggeman Q, Garmier M, de Bont L, Soubigou-Taconnat L, Mazubert C, Benhamed M, Raynaud C, Bergounioux C, Delarue M (2014) The polyadenylation factor subunit CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR30: A key factor of programmed cell death and a regulator of immunity in *Arabidopsis*. **Plant Physiol** 165: 732–746

Bujnicki JM, Feder M, Radlinska M, Blumenthal RM (2002) Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. **J Mol Evol** 55: 431–444

Burgess AL, David R, Searle IR (2015) Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. **BMC Plant Biol** 15: 199

Byrne ME (2009) A role for the ribosome in development. **Trends Plant Sci** 14: 512–519

Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. **Nature** 515: 143–146

Chakrabarti M, Hunt AG (2015) CPSF30 at the interface of alternative polyadenylation and cellular signaling in plants. **Biomolecules** 5: 1151–1168

Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ (2010) A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. **PLoS Genet** 6: e1001247

Chan CTY, Pang YLJ, Deng WJ, Babu IR, Dyavaiah M, Begley TJ, Dedon PC (2012) Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. **Nat Commun** 3: 937

Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR, 3rd, Ule J, Manley JL, Shi Y (2014) CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. **Genes Dev** 28: 2370–2380

Charette M, Gray MW (2000) Pseudouridine in RNA: What, where, how, and why. **IUBMB Life** 49: 341–351

Chen K, Lu Z, Wang X, Fu Y, Luo GZ, Liu N, Han D, Dominissini D, Dai Q, Pan T, He C (2015a) High-resolution N(6) -methyladenosine (m(6) A) map using photo-crosslinking-assisted m(6) A sequencing. **Angew Chem Int Ed Engl** 54: 1587–1590

Chen P, Jager G, Zheng B (2010) Transfer RNA modifications and genes for modifying enzymes in *Arabidopsis thaliana*. **BMC Plant Biol** 10: 201

Chen T, Hao YJ, Zhang Y, Li MM, Wang M, Han W, Wu Y, Lv Y, Hao J, Wang L, Li A, Yang Y, Jin KX, Zhao X, Li Y, Ping XL, Lai WY, Wu LG, Jiang G, Wang HL, Sang L, Wang XJ, Yang YG, Zhou Q (2015b) m(6) A RNA methylation is regulated by microRNAs and promotes reprogramming to pluripotency. **Cell Stem Cell** 16: 289–301

Chen Y, Sierzputowskagracz H, Guenther R, Everett K, Agris PF (1993) 5-methylcytidine is required for cooperative binding of Mg^{2+} and a conformational transition at the anticodon stem-loop of yeast phenylalanine transfer-RNA. **Biochemistry** 32: 10249–10253

Chen Z, Zhang H, Jablonowski D, Zhou X, Ren X, Hong X, Schaffrath R, Zhu JK, Gong Z (2006) Mutations in ABO1/ELO2, a subunit of holo-Elongator, increase abscisic acid sensitivity and drought tolerance in *Arabidopsis thaliana*. **Mol Cell Biol** 26: 6902–6912

Clancy MJ, Shambaugh ME, Timpte CS, Bokar JA (2002) Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N6-methyladenosine in mRNA: A potential mechanism for the activity of the IME4 gene. **Nucleic Acids Res** 30: 4509–4518

Csepany T, Lin A, Baldick CJ, Jr., Beemon K (1990) Sequence specificity of mRNA N6-adenosine methyltransferase. **J Biol Chem** 265: 20117–20122

David R, Burgess AL, Parker BJ, Pulsford KE, Sibbritt T, Preiss T, Searle IR (2016) Transcriptome-wide mapping of RNA 5-methylcytosine in *Arabidopsis* mRNAs and ncRNAs. **Genome Res** dx.doi.org/10.6084/m9.figshare.3408193.v2

Davis DR (1995) Stabilization of RNA stacking by pseudouridine. **Nucleic Acids Res** 23: 5020–5026

Decatur WA, Fournier MJ (2002) rRNA modifications and ribosome function. **Trends Biochem Sci** 27: 344–351

Delaney KJ, Xu R, Zhang J, Li QQ, Yun KY, Falcone DL, Hunt AG (2006) Calmodulin interacts with and regulates the RNA-binding activity of an *Arabidopsis* polyadenylation factor subunit. **Plant Physiol** 140: 1507–1521

Deng X, Chen K, Luo GZ, Weng X, Ji Q, Zhou T, He C (2015) Widespread occurrence of N6-methyladenosine in bacterial mRNA. **Nucleic Acids Res** 43: 6557–6567

Desrosiers R, Friderici K, Rottman F (1974) Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. **Proc Natl Acad Sci USA** 71: 3971–3975

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R, Rechavi G (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. **Nature** 485: 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, Zheng G, Pan T, Solomon O, Eyal E, Hershkovitz V, Han D, Dore LC, Amariglio N, Rechavi G, He C (2016) The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. **Nature** 530: 441–446

Dubin DT, Taylor RH (1975) The methylation state of poly A-containing messenger RNA from cultured hamster cells. **Nucleic Acids Res** 2: 1653–1668

Durdevic Z, Hanna K, Gold B, Pollex T, Cherry S, Lyko F, Schaefer M (2013a) Efficient RNA virus control in Drosophila requires the RNA methyltransferase Dnmt2. **EMBO Rep** 14: 269–275

Durdevic Z, Mobin MB, Hanna K, Lyko F, Schaefer M (2013b) The RNA methyltransferase Dnmt2 is required for efficient dicer-2-dependent siRNA pathway activity in Drosophila. **Cell Rep** 4: 931–937

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 Genet** 9: e1003602

El Yacoubi B, Bailly M, de Crecy-Lagard V (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. **Annu Rev Genet** 46: 69–95

Fahiminiya S, Almuriekhi M, Nawaz Z, Staffa A, Lepage P, Ali R, Hashim L, Schwartzentruber J, Abu Khadija K, Zaineddin S, Gamal H, Majewski J, Ben-Omran T (2014) Whole exome sequencing unravels disease-causing genes in consanguineous families in Qatar. **Clin Genet** 86: 134–141

Faure JD, Gingerich D, Howell SH (1998) An *Arabidopsis* immunophilin, AtFKBP12, binds to AtFIP37 (FKBP interacting protein) in an interaction that is disrupted by FK506. **Plant J** 15: 783–789

Fray RG, Simpson GG (2015) The *Arabidopsis* epitranscriptome. **Curr Opin Plant Biol** 27: 17–21

Fujikura U, Horiguchi G, Ponce MR, Micol JL, Tsukaya H (2009) Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of *Arabidopsis thaliana*. **Plant J** 59: 499–508

Fustin JM, Doi M, Yamaguchi Y, Hida H, Nishimura S, Yoshida M, Isagawa T, Morioka MS, Kakeya H, Manabe I, Okamura H (2013) RNA-methylation-dependent RNA processing controls the speed of the circadian clock. **Cell** 155: 793–806

Ge JH, Yu YT (2013) RNA pseudouridylation: New insights into an old modification. **Trends Biochem Sci** 38: 210–218

Gigova A, Duggimpudi S, Pollex T, Schaefer M, Kos M (2014) A cluster of methylations in the domain IV of 25S rRNA is required for ribosome stability. **RNA** 20: 1632–1644

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang XY, Golic KG, Jacobsen SE, Bestor TH (2006) Methylation of tRNA(AsP) by the DNA methyltransferase homolog Dnmt2. **Science** 311: 395–398

Golovko A, Sitbon F, Tillberg E, Nicander B (2002) Identification of a tRNA isopentenyltransferase gene from *Arabidopsis thaliana*. **Plant Mol Biol** 49: 161–169

Hebrard C, Trap-Gentil MV, Lafon-Placette C, Delaunay A, Joseph C, Lefebvre M, Barnes S, Maury S (2013) Identification of differentially methylated regions during vernalization revealed a role for RNA methyltransferases in bolting. **J Exp Bot** 64: 651–663

Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, Dokal I (1998) X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. **Nat Genet** 19: 32–38

Henras AK, Plisson-Chastang C, O'Donohue MF, Chakraborty A, Gleizes PE (2015) An overview of pre-ribosomal RNA processing in eukaryotes. **Wiley Interdiscip Rev RNA** 6: 225–242

Hertz R, Tovy A, Kirschenbaum M, Geffen M, Nozaki T, Adir N, Ankri S (2014) The entamoeba histolytica Dnmt2 homolog (ehmeth) confers resistance to nitrosative stress. **Eukaryotic Cell** 13: 494–503

Hienzsch A, Deiml C, Reiter V, Carell T (2013) Total synthesis of the hypermodified RNA bases wybutosine and hydroxywybutosine and their quantification together with other modified RNA bases in plant materials. **Chemistry** 19: 4244–4248

Hoernes TP, Clementi N, Faserl K, Glasner H, Breuker K, Lindner H, Huttenhofer A, Erlacher MD (2015) Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. **Nucleic Acids Res** 44: 852–862

Hong B, Brockenbrough JS, Wu P, Aris JP (1997) Nop2p is required for pre-rRNA processing and 60S ribosome subunit synthesis in yeast. **Mol Cell Biol** 17: 378–388

Hongay CF, Orr-Weaver TL (2011) Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. **Proc Natl Acad Sci USA** 108: 14855–14860

Hu Z, Qin Z, Wang M, Xu C, Feng G, Liu J, Meng Z, Hu Y (2010) The *Arabidopsis* SMO2, a homologue of yeast TRM112, modulates progression of cell division during organ growth. **Plant J** 61: 600–610

Hunt AG, Xing D, Li QQ (2012) Plant polyadenylation factors: Conservation and variety in the polyadenylation complex in plants. **BMC Genomics** 13: 641

Hussain S, Sajini AA, Blanco S, Dietmann S, Lombard P, Sugimoto Y, Paramor M, Gleeson JG, Odom DT, Ule J, Frye M (2013a) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. **Cell Rep** 4: 255–261

Hussain S, Tuorto F, Menon S, Blanco S, Cox C, Flores JV, Watt S, Kudo NR, Lyko F, Frye M (2013b) The mouse cytosine-5 RNA methyltransferase NSun2 is a component of the chromatoid body and required for testis differentiation. **Mol Cell Biol** 33: 1561–1570

Iida K, Jin HL, Zhu JK (2009) Bioinformatics analysis suggests base modifications of tRNAs and miRNAs in *Arabidopsis thaliana*. **BMC Genomics** 10: 155

Jackman JE, Alfonzo JD (2013) Transfer RNA modifications: Nature's combinatorial chemistry playground. **Wires RNA** 4: 35–48

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG, He C (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. **Nat Chem Biol** 7: 885–887

Jones LH, Scott TK (1981) Transfer ribonucleic acid modification and its relationship to tumorous and nontumorous plant growth. **Plant Physiol** 67: 535–538

Karijolich J, Yu YT (2011) Converting nonsense codons into sense codons by targeted pseudouridylation. **Nature** 474: 395–399

Kariko K, Muramatsu H, Keller JM, Weissman D (2012) Increased erythropoiesis in mice injected with submicrogram quantities of pseudouridine-containing mRNA encoding erythropoietin. **Mol Ther** 20: 948–953

Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, Haripal B, Zucker-Scharff I, Moore MJ, Park CY, Vagbo CB, Kussnierczyk A, Klungland A, Darnell JE, Jr., Darnell RB (2015) A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. **Genes Dev** 29: 2037–2053

Khan MA, Rafiq MA, Noor A, Hussain S, Flores JV, Rupp V, Vincent AK, Malli R, Ali G, Khan FS, Ishak GE, Doherty D, Weksberg R, Ayub M, Windpassinger C, Ibrahim S, Frye M, Ansar M, Vincent JB (2012) Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. **Am J Hum Genet** 90: 856–863

Khoddami V, Cairns BR (2013) Identification of direct targets and modified bases of RNA cytosine methyltransferases. **Nat Biotechnol** 31: 458–464

King MY, Redman KL (2002) RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. **Biochemistry** 41: 11218–11225

Kiss T (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. **EMBO J** 20: 3617–3622

Lafontaine D, Delcour J, Glasser AL, Desgres J, Vandenhaute J (1994) The DIM1 gene responsible for the conserved m6(2)Am6(2)A dimethylation in the 3'-terminal loop of 18 S rRNA is essential in yeast. **J Mol Biol** 241: 492–497

Lafontaine D, Vandenhaute J, Tollervey D (1995) The 18S rRNA dimethylase Dim1p is required for pre-ribosomal RNA processing in yeast. **Genes Dev** 9: 2470–2481

Lawrence M, Daujat S, Schneider R (2016) Lateral thinking: How histone modifications regulate gene expression. **Trends Genet** 32: 42–56

Leihne V, Kirpekar F, Vagbo CB, van den Born E, Krokan HE, Grini PE, Meza TJ, Falnes PO (2011) Roles of Trm9-and ALKBH8-like proteins in the formation of modified wobble uridines in *Arabidopsis* tRNA. **Nucleic Acids Res** 39: 7688–7701

Lermontova I, Schubert V, Bornke F, Macas J, Schubert I (2007) *Arabidopsis* CBF5 interacts with the H/ACA snoRNP assembly factor NAF1. **Plant Mol Biol** 65: 615–626

Li D, Zhang H, Hong Y, Huang L, Li X, Zhang Y, Ouyang Z, Song F (2014a) Genome-wide identification, biochemical characterization, and expression analyses of the YTH domain-containing RNA-binding protein family in *Arabidopsis* and rice. **Plant Mol Biol Rep** 32: 1169–1186

Li X, Zhu P, Ma S, Song J, Bai J, Sun F, Yi C (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. **Nat Chem Biol** 11: 592–597

Li Y, Wang X, Li C, Hu S, Yu J, Song S (2014b) Transcriptome-wide N(6)-methyladenosine profiling of rice callus and leaf reveals the presence of tissue-specific competitors involved in selective mRNA modification. **RNA Biol** 11: 1180–1188

Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. **Nat Methods** 12: 767–772

Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X, Dai Q, Chen W, He C (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. **Nat Chem Biol** 10: 93–95

Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T (2015) N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. **Nature** 518: 560–564

Lopato S, Kalyna M, Dorner S, Kobayashi R, Krainer AR, Barta A (1999) atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. **Genes Dev** 13: 987–1001

Lorkovic ZJ, Barta A (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. **Nucleic Acids Res** 30: 623–635

Lorkovic ZJ, Wieczorek Kirk DA, Lambermon MH, Filipowicz W (2000) Pre-mRNA splicing in higher plants. **Trends Plant Sci** 5: 160–167

Lovejoy AF, Riordan DP, Brown PO (2014) Transcriptome-wide mapping of pseudouridines: Pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. **PLoS ONE** 9: e110799

Luo GZ, MacQueen A, Zheng G, Duan H, Dore LC, Lu Z, Liu J, Chen K, Jia G, Bergelson J, He C (2014) Unique features of the m6A methylome in *Arabidopsis thaliana*. **Nat Commun** 5: 5630

Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, Helm M, Bujnicki JM, Grosjean H (2013) MODOMICS: A database of RNA modification pathways–2013 update. **Nucleic Acids Res** 41: D262–267

Martinez FJ, Lee JH, Lee JE, Blanco S, Nickerson E, Gabriel S, Frye M, Al-Gazali L, Gleeson JG (2012) Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. **J Med Genet** 49: 380–385

Mehlgarten C, Jablonowski D, Wrackmeyer U, Tschitschmann S, Sondermann D, Jager G, Gong Z, Bystrom AS, Schaffrath R, Breunig KD (2010) Elongator function in tRNA wobble uridine modification is conserved between yeast and plants. **Mol Microbiol** 76: 1082–1094

Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. **Cell** 149: 1635–1646

Mielecki D, Zugaj DL, Muszewska A, Piwowarski J, Chojnacka A, Mielecki M, Nieminuszczy J, Grynberg M, Grzesiuk E (2012) Novel AlkB dioxygenases–alternative models for in silico and *in vivo* studies. **PLoS ONE** 7: e30588

Miyawaki K, Tarkowski P, Matsumoto-Kitano M, Kato T, Sato S, Tarkowska D, Tabata S, Sandberg G, Kakimoto T (2006) Roles of *Arabidopsis* ATP/ADP isopentenyltransferases and tRNA

isopentenyltransferases in cytokinin biosynthesis. **Proc Natl Acad Sci USA** 103: 16598–16603

Motorin Y, Helm M (2011) RNA nucleotide methylation. **Wiley Interdiscip Rev RNA** 2: 611–631

Motorin Y, Lyko F, Helm M (2010) 5-methylcytosine in RNA: Detection, enzymatic formation and biological functions. **Nucleic Acids Res** 38: 1415–1430

Nishimura T, Wada T, Yamamoto KT, Okada K (2005) The *Arabidopsis* STV1 protein, responsible for translation reinitiation, is required for auxin-mediated gynoecium patterning. **Plant Cell** 17: 2940–2953

Pavlopoulou A, Kossida S (2009) Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. **Genomics** 93: 350–357

Perry RP, Kelley DE (1974) Existence of methylated messenger RNA in mouse L cells. **Cell** 1: 37–42

Phizicky EM, Hopper AK (2010) tRNA biology charges to the front. **Genes Dev** 24: 1832–1860

Pih KT, Yi MJ, Liang YS, Shin BJ, Cho MJ, Hwang I, Son D (2000) Molecular cloning and targeting of a fibrillarin homolog from *Arabidopsis*. **Plant Physiol** 123: 51–58

Ping XL, Sun BF, Wang L, Xiao W, Yang X, Wang WJ, Adhikari S, Shi Y, Lv Y, Chen YS, Zhao X, Li A, Yang Y, Dahal U, Lou XM, Liu X, Huang J, Yuan WP, Zhu XF, Cheng T, Zhao YL, Wang XQ, Danielsen JMR, Liu F, Yang YG (2014) Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. **Cell Res** 24: 177–189

Preston MA, D'Silva S, Kon Y, Phizicky EM (2013) tRNAHis 5-methylcytidine levels increase in response to several growth arrest conditions in Saccharomyces cerevisiae. **RNA** 19: 243–256

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 Dev** 21: 261–266

Ramirez V, Gonzalez B, Lopez A, Castello MJ, Gil MJ, Zheng B, Chen P, Vera P (2015) Loss of a conserved tRNA anticodon modification perturbs plant immunity. **PLoS Genet** 11: e1005586

Richter U, Kuhn K, Okada S, Brennicke A, Weihe A, Borner T (2010) A mitochondrial rRNA dimethyladenosine methyltransferase in *Arabidopsis*. **Plant J** 61: 558–569

Roost C, Lynch SR, Batista PJ, Qu K, Chang HY, Kool ET (2015) Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. **J Am Chem Soc** 137: 2107–2115

Ryvkin P, Leung YY, Silverman IM, Childress M, Valladares O, Dragomir I, Gregory BD, Wang LS (2013) HAMR: High-throughput annotation of modified ribonucleotides. **RNA** 19: 1684–1692

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 Dev** 24: 1590–1595

Schwartz S, Agarwala SD, Mumbach MR, Jovanovic M, Mertins P, Shishkin A, Tabach Y, Mikkelsen TS, Satija R, Ruvkun G, Carr SA, Lander ES, Fink GR, Regev A (2013) High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. **Cell** 155: 1409–1421

Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, Leon-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, Fink G, Regev A (2014a) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. **Cell** 159: 148–162

Schwartz S, Mumbach MR, Jovanovic M, Wang T, Maciag K, Bushkin GG, Mertins P, Ter-Ovanesyan D, Habib N, Cacchiarelli D, Sanjana NE, Freinkman E, Pacold ME, Satija R, Mikkelsen TS, Hacohen N, Zhang F, Carr SA, Lander ES, Regev A (2014b) Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. **Cell Rep** 8: 284–296

Shafik A, Schumann U, Evers M, Sibbritt T, Preiss T (2016) The emerging epitranscriptomics of long noncoding RNAs. **Biochim Biophys Acta** 1859: 59–70

Sharma S, Yang J, Watzinger P, Kotter P, Entian KD (2013) Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. **Nucleic Acids Res** 41: 9062–9076

Shatkin AJ (1976) Capping of eucaryotic mRNAs. **Cell** 9: 645–653

Shugart L (1972) A possible age-related modification of phenylalanine transfer RNA from wheat tissue. **Exp Gerontol** 7: 251–262

Sibert BS, Patton JR (2012) Pseudouridine synthase 1: A site-specific synthase without strict sequence recognition requirements. **Nucleic Acids Res** 40: 2107–2118

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 Res** 40: 5023–5033

Suzuki MM, Bird A (2008) DNA methylation landscapes: Provocative insights from epigenomics. **Nat Rev Genet** 9: 465–476

Thomas PE, Wu X, Liu M, Gaffney B, Ji G, Li QQ, Hunt AG (2012) Genome-wide control of polyadenylation site choice by CPSF30 in *Arabidopsis*. **Plant Cell** 24: 4376–4388

Tokuhisa JG, Vijayan P, Feldmann KA, Browse JA (1998) Chloroplast development at low temperatures requires a homolog of DIM1, a yeast gene encoding the 18S rRNA dimethylase. **Plant Cell** 10: 699–711

Torres AG, Batlle E, Ribas de Pouplana L (2014) Role of tRNA modifications in human diseases. **Trends Mol Med** 20: 306–314

Towns WL, Begley TJ (2012) Transfer RNA methytransferases and their corresponding modifications in budding yeast and humans: Activities, predications, and potential roles in human health. **DNA Cell Biol** 31: 434–454

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. **Nat Struct Mol Biol** 19: 900–905

Vandivier LE, Campos R, Kuksa PP, Silverman IM, Wang LS, Gregory BD (2015) Chemical modifications mark alternatively spliced and uncapped messenger RNAs in *Arabidopsis*. **Plant Cell** 27: 3024–3037

Vespa L, Vachon G, Berger F, Perazza D, Faure JD, Herzog M (2004) The immunophilin-interacting protein AtFIP37 from *Arabidopsis* is essential for plant development and is involved in trichome endoreduplication. **Plant Physiol** 134: 1283–1292

Wan Y, Tang K, Zhang D, Xie S, Zhu X, Wang Z, Lang Z (2015) Transcriptome-wide high-throughput deep m(6)A-seq reveals unique differential m(6)A methylation patterns between three organs in *Arabidopsis thaliana*. **Genome Biol** 16: 272

Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, Ren B, Pan T, He C (2014a) N6-methyladenosine-dependent regulation of messenger RNA stability. **Nature** 505: 117–120

Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, Weng X, Chen K, Shi H, He C (2015) N(6)-methyladenosine modulates messenger RNA translation efficiency. **Cell** 161: 1388–1399

Wang Y, Li Y, Toth JI, Petroski MD, Zhang Z, Zhao JC (2014b) N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. **Nat Cell Biol** 16: 191–198

Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, Ebina W, Mandal PK, Smith ZD, Meissner A, Daley GQ, Brack AS, Collins JJ, Cowan C, Schlaeger TM, Rossi DJ (2010) Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. **Cell Stem Cell** 7: 618–630

Wei CM, Moss B (1977) Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. **Biochemistry** 16: 1672–1676

Wieckowski Y, Schiefelbein J (2012) Nuclear ribosome biogenesis mediated by the DIM1A rRNA dimethylase is required for organized root growth and epidermal patterning in *Arabidopsis*. **Plant Cell** 24: 2839–2856

Wu P, Brockenbrough JS, Paddy MR, Aris JP (1998) *NCL1*, a novel gene for a non-essential nuclear protein in *Saccharomyces cerevisiae*. **Gene** 220: 109–117

Xiao W, Adhikari S, Dahal U, Chen YS, Hao YJ, Sun BF, Sun HY, Li A, Ping XL, Lai WY, Wang X, Ma HL, Huang CM, Yang Y, Huang N, Jiang GB, Wang HL, Zhou Q, Wang XJ, Zhao YL, Yang YG (2016) Nuclear m(6)A reader YTHDC1 regulates mRNA splicing. **Mol Cell** 61: 507–519

Yu F, Liu X, Alsheikh M, Park S, Rodermel S (2008) Mutations in SUPPRESSOR OF VARIEGATION1, a factor required for normal chloroplast translation, suppress var2-mediated leaf variegation in *Arabidopsis*. **Plant Cell** 20: 1786–1804

Zhang J, Addepalli B, Yun KY, Hunt AG, Xu R, Rao S, Li QQ, Falcone DL (2008) A polyadenylation factor subunit implicated in regulating oxidative signaling in *Arabidopsis thaliana*. **PLoS ONE** 3: e2410

Zhao X, Yang Y, Sun BF, Shi Y, Yang X, Xiao W, Hao YJ, Ping XL, Chen YS, Wang WJ, Jin KX, Wang X, Huang CM, Fu Y, Ge XM, Song SH, Jeong HS, Yanagisawa H, Niu Y, Jia GF, Wu W, Tong WM, Okamoto A, He C, Rendtlew Danielsen JM, Wang XJ, Yang YG (2014) FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. **Cell Res** 24: 1403–1419

Zheng GQ, Dahl JA, Niu YM, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH, Lu ZK, Bosmans RPG, Dai Q, Hao YJ, Yang X, Zhao WM, Tong WM, Wang XJ, Bogdan F, Furu K, Fu Y, Jia GF, Zhao X, Liu J, Krokan HE, Klungland A, Yang YG, He C (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. **Mol Cell** 49: 18–29

Zhong SL, Li HY, Bodi Z, Button J, Vespa L, Herzog M, Fray RG (2008) MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. **Plant Cell** 20: 1278–1288

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** 526: 591–594

Zhou WB, Karcher D, Bock R (2013) Importance of adenosine-to-inosine editing adjacent to the anticodon in an *Arabidopsis* alanine tRNA under environmental stress. **Nucleic Acids Res** 41: 3362–3372

Zhou X, Hua D, Chen Z, Zhou Z, Gong Z (2009) Elongator mediates ABA responses, oxidative stress resistance and anthocyanin biosynthesis in *Arabidopsis*. **Plant J** 60: 79–90

Zorbas C, Nicolas E, Wacheul L, Huvelle E, Heurgue-Hamard V, Lafontaine DL (2015) The human 18S rRNA base methyltransferases DIMT1L and WBSCR22-TRMT112 but not rRNA modification are required for ribosome biogenesis. **Mol Biol Cell** 26: 2080–2095

Chapter 2

Conservation of tRNA and rRNA 5-methylcytosine in the kingdom *Plantae*

Alice L. Burgess[1,2], Rakesh David[1,2] and Iain R. Searle[1,2,3]

[1]School of Biological Sciences, The University of Adelaide, Adelaide, SA, 5005 Australia

[2]School of Agriculture, Food and Wine, The Waite Research Institute, The University of Adelaide, South Australia, 5005 Australia

[3]The University of Adelaide and Shanghai Jiao Tong University Joint International Centre for Agriculture and Health

BMC Plant Biology, August 2015

Statement of Authorship

Title of paper	Conservation of tRNA and rRNA 5-methylcytosine in the kingdom *Plantae*
Publication Status	Published
Publication details	Burgess A., David R. and Searle, I. (2015) Conservation of tRNA and rRNA methylation in the kingdom Plantae. *BMC Plant Biology* 15(1), 19

Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed below);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Candidate	Alice L. Burgess		
Contribution to paper	Equal first author with Rakesh David. Co-designer of experiments. Identified and characterized mutants. Performed RNA bisulfite treatments. Calculated conversion efficiencies. Analysed transfer RNA data. Wrote the manuscript and composed figures.		
Overall percentage (%)	45%		
Signature		Date	15/3/16

Name of Co-Author	Rakesh David		
Contribution to paper	Equal first author with Alice Burgess. Co-designer of experiments. Purified tRNAs, and made all RBS-seq and RNA-seq libraries. Analysed ribosomal RNA data. Performed methyl-chop PCR's for figure 2D and Figure 3D. Supervised development of work and edited the manuscript. Overall contribution 45%.		
Signature		Date	15/3/16

Name of Co-Author	Iain R. Searle		
Contribution to paper	Co-designer of experiments. Supervised development of work and edited the manuscript. Overall contribution 10%.		
Signature		Date	20/02/2016

Burgess et al. BMC Plant Biology (2015) 15:199
DOI 10.1186/s12870-015-0580-8

BMC Plant Biology

RESEARCH ARTICLE Open Access

Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae

Alice Louise Burgess[1,2†], Rakesh David[1,2†] and Iain Robert Searle[1,2,3*]

Abstract

Background: Post-transcriptional methylation of RNA cytosine residues to 5-methylcytosine (m^5C) is an important modification that regulates RNA metabolism and occurs in both eukaryotes and prokaryotes. Yet, to date, no transcriptome-wide identification of m^5C sites has been undertaken in plants. Plants provide a unique comparative system for investigating the origin and evolution of m^5C as they contain three different genomes, the nucleus, mitochondria and chloroplast. Here we use bisulfite conversion of RNA combined with high-throughput Illumina sequencing (RBS-seq) to identify single-nucleotide resolution of m^5C sites in non-coding ribosomal RNAs and transfer RNAs of all three sub-cellular transcriptomes across six diverse species that included, the single-celled algae *Nannochloropsis oculata,* the macro algae *Caulerpa taxifolia* and multi-cellular higher plants *Arabidopsis thaliana*, *Brassica rapa*, *Triticum durum* and *Ginkgo biloba*.

Results: Using the plant model *Arabidopsis thaliana*, we identified a total of 39 highly methylated m^5C sites in predicted structural positions of nuclear tRNAs and 7 m^5C sites in rRNAs from nuclear, chloroplast and mitochondrial transcriptomes. Both the nucleotide position and percent methylation of tRNAs and rRNAs m^5C sites were conserved across all species analysed, from single celled algae *N. oculata* to multicellular plants. Interestingly the mitochondrial and chloroplast encoded tRNAs were devoid of m^5C in *A. thaliana* and this is generally conserved across *Plantae*. This suggests independent evolution of organelle methylation in animals and plants, as animal mitochondrial tRNAs have m^5C sites. Here we characterize 5 members of the RNA 5-methylcytosine family in *Arabidopsis* and extend the functional characterization of TRDMT1 and NOP2A/OLI2. We demonstrate that nuclear tRNA methylation requires two evolutionarily conserved methyltransferases, TRDMT1 and TRM4B. *trdmt1 trm4b* double mutants are hypersensitive to the antibiotic hygromycin B, demonstrating the function of tRNA methylation in regulating translation. Additionally we demonstrate that nuclear large subunit 25S rRNA methylation requires the conserved RNA methyltransferase NSUN5. Our results also suggest functional redundancy of at least two of the NOP2 paralogs in *Arabidopsis*.

Conclusions: Our data demonstrates widespread occurrence and conservation of non-coding RNA methylation in the kingdom *Plantae,* suggesting important and highly conserved roles of this post-transcriptional modification.

Keywords: RNA 5-methylcytosine, Non-coding RNA, Ribosomal RNA (rRNA), Transfer RNA (tRNA), *Arabidopsis thaliana*, TRDMT1, DNMT2, TRM4, NOP2, NSUN5

* Correspondence: Iain.Searle@adelaide.edu.au
†Equal contributors
[1]School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia
[2]School of Agriculture, Food and Wine, The Waite Research Institute, The University of Adelaide, Adelaide, South Australia 5005, Australia
Full list of author information is available at the end of the article

Background

5-methylcytosine (m^5C) is a modification that occurs both on DNA and RNA. In DNA, m^5C has been extensively studied due to its ease of detection and functional roles of DNA methylation in eukaryotes have been demonstrated for transcriptional silencing of transposons and transgenes, genomic imprinting and X chromosome inactivation (reviewed in [1]). While DNA appears to be devoid of other modifications [1], RNA has over 100 different modifications that have been identified in different species across all three domains of life [2–4]. Transfer RNAs (tRNAs) are heavily decorated with modifications that have been shown to stabilize secondary structure, affect codon identification and tRNA aminoacylation [5–8]. Of these modifications, m^5C sites in tRNAs are commonly identified in the variable region and anticodon loop. In response to oxidative stress, m^5C has been demonstrated to be dynamically modulated in yeast [9, 10] and m^5C plays an important role in regulating tRNA stability and translation in mice under controlled conditions [11]. Furthermore, m^5C is required for tRNA stability under heat stress and oxidative stress conditions in fruit flies [12]. In ribosomal RNAs (rRNA), m^5C sites are thought to play a role in translation, rRNA processing and structure [13–15].

In eukaryotes, transfer RNA m^5C methylation is catalysed by two RNA methyltransferases (RMTases); the first class of RMTase is known as tRNA specific methytransferase 4 (TRM4) or NOP2/Sun domain protein 2 (NSUN2), in yeast and animals respectively [11, 16, 17]. *NSUN2* mutations in humans are linked to inherited intellectual disability and this is thought to be mediated by increased cleavage of tRNAs by the ribonuclease angiogenin [18–22]. In mice, *nsun2* mutants are smaller and have reduced male fertility and have revealed a role in stem cell self-renewal and differentiation [23, 24]. Using phylogenetic analysis, two putative TRM4/NSUN2 paralogs, *TRM4A* and *TRM4B,* were identified in the *Arabidopsis* genome [25, 26], however these genes have not been characterized in plants. The second class of eukaryotic RMTase; Transfer RNA aspartic acid methyltransferase 1 (TRDMT1), also known as DNA methyltransferase 2 (DNMT2), has been shown to methylate tRNAs in *Drosophila*, *Arabidopsis* and *Homo sapiens*. In plants, only one m^5C site in $tRNA^{Asp(GTC)}$ at position C38 has been shown to be methylated by TRDMT1 [27]. While *Drosophila*, and *Arabidopsis trdmt1* mutants appear wild type under standard laboratory conditions, zebrafish deficient in TRDMT1 have reduced body size and impaired differentiation of specific tissues [27, 28]. In nuclear encoded eukaryotic tRNAs, m^5C methylation has been commonly reported at six cytosine positions; C34, C38, C48, C49, C50 and C72 [2, 3, 18, 29–31]. Methylation has also been discovered on mitochondrial encoded tRNAs in humans and cows on several tRNAs at positions C48, C49 and C72 [29, 32]. However, the methylation status of chloroplast encoded tRNAs and rRNAs has not been previously reported.

Like tRNAs, ribosomal RNAs are highly conserved and have important roles in translation. The ribosome consists of two subunits, the large subunit (LSU) and the small subunit (SSU). The LSU is composed of three rRNA species in eukaryotes, and generally two rRNA species in prokaryotes, while the SSU contains only one rRNA species in both prokaryotes and eukaryotes [33–35]. The rRNA sequences are conserved, although the names of rRNA species are often not. Whereas rRNA methylation has not been investigated in plants, the location and enzymatic requirements of a few m^5C sites in select organisms has been determined. For example, the human nuclear LSU rRNAs (28S and 5S) are methylated. The 28S rRNA contains two sites at C3782 and C4447 while 5S rRNA is methylated at C92 [30, 31, 36]. The orthologous yeast LSU 25S rRNA contains two sites at C2278 and C2870 [13, 15] and *E. coli* LSU 23S rRNA at C1962 [37] and SSU 16S rRNA at C967 [38] and C1407 [39]. Hamster mitochondrial SSU 13S rRNA also contains one m^5C site [40], similarly mouse mitochondrial SSU 12S rRNA is methylated at position C911 [41]. Two RMTases that have been identified to methylate ribosomal RNA in eukaryotes are NOP2 (nucleolar protein 2) and RCM1 (rRNA cytosine methyltransferase 1). NOP2 methylates position C2870 and RCM1 methylates position C2278 in the LSU 25S rRNA in *Saccharomyces* [13, 15]. Yeast NOP2 is required for correct rRNA biosynthesis and processing [14] and *nop2* mutants are lethal. In contrast, yeast *rcm1* mutants are viable, however they are hypersensitive to anisomycin and this is thought to be due to structural changes being induced by methylation of rRNA [15]. While there is only one copy of the RCM1 homolog, referred to here as NSUN5 in *Arabidopsis*, there are three paralogs of *NOP2* in the *Arabidopsis* genome, *OLI2 (NOP2A)*, *NOP2B* and *NOP2C* [26]. One of these, *NOP2A*/*OLI2* was identified in a forward genetic screen for genes involved in compensation of cell size [42]. The methylation activity or m^5C sites mediated by the three *Arabidopsis* NOP2 paralogs and NSUN5 are unknown. Another RMTase, which is related to the bacterial Fmu rRNA MTase was recently identified in *Arabidopsis* [43]. *Arabidopsis rnmt* (RNA methyltransferase) mutants had reduced global RNA methylation, indicating that it may methylate highly abundant rRNA transcripts.

Unlike animals, plant cells contain three evolutionary distinct genomes; nuclear, mitochondrial and chloroplast, thus providing a unique model for investigating m^5C catalysis and biological function. The mitochondria is a striking example of how a prokaryotic translational machinery has adapted to input from eukaryotic translational machinery as nuclear, eukaryotic tRNAs are required to be imported

into the mitochondria, as the mitochondria no longer has a full complement of tRNAs [44, 45]. tRNA sequences present in plants are dynamic, as there are multiple copies of each tRNA isodecoder and these can be lost within a genome or transferred from the chloroplast and mitochondrial genomes to the nucleus [46]. This gives rise to incidents where a nuclear encoded tRNA has an organelle-like sequence. It is unknown whether these "transferred" tRNAs are expressed after integration into a new genome as a systematic analysis of tRNA expression in plants is yet to be undertaken [47–49].

In this study, we describe single nucleotide resolution of post-transcriptionally modified cytosine residues in plant rRNA and tRNAs by combining RNA bisulfite conversion with second generation Illumina sequencing (RBS-seq). We report the identification of novel modified cytosines in *A. thaliana* nuclear transcribed tRNAs and that these sites are dependent on RMTases TRDMT1 and the previously undescribed *Arabidopsis* TRM4B. Additionally, we show these modified sites in nuclear tRNAs are conserved through evolution from the single celled algae *Nannochloropsis oculata* to multicellular higher plants. Interestingly, no m^5C sites were detected in *Arabidopsis* chloroplast or mitochondrial tRNAs, which is in contrast to animal mitochondrial tRNAs. The function of tRNA methylation in regulating translation is demonstrated, as *trdmt1 trm4b* double mutants are hypersensitive to the antibiotic hygromycin B. Furthermore, we identify novel modified cytosines in nuclear, mitochondrial and chloroplast rRNAs. In *Arabidopsis* nuclear LSU 25S rRNA, m^5C at C2268 was dependent on NSUN5, but methylation at C2860 was not found to be dependent on any particular NOP2 ortholog in *Arabidopsis*. Furthermore, RMTases responsible for methylation of tRNAs were not required for rRNA methylation, and vice versa indicating functional specialization of the RMTase family. These data represent the first high-resolution description of tRNA and rRNA modifications in the *plantae* kingdom and creates a platform to begin understanding the function, significance and evolution of non-coding RNA methylation.

Results

Detection and enrichment of transcribed tRNAs in *Arabidopsis thaliana*

To identify transcribed tRNAs in *A. thaliana* we implemented a two-step approach. First, a tRNA isodecoder consensus list was constructed to facilitate expression analysis and second, a tRNA enrichment protocol combined with Illumina deep-sequencing was developed similar to those recently described [50]. The tRNA isodecoder consensus approach was undertaken as there are over 640 predicted tRNA genes in *A. thaliana,* originating from the nuclear, mitochondrial and chloroplast genomes often with multiple identical isodecoder sequences that makes assigning IIlumina sequences to individual transcribed tRNA loci challenging. Using this consensus approach, the predicted *A. thaliana* tRNAs were resolved into 100 reference consensus sequences (Additional file 1: Table S1).

To identify transcribed tRNAs, we initially used total RNA to construct an Illumina library, deep-sequenced the library and aligned the sequenced reads to our tRNA consensus list. Only 0.0007 % of sequence reads aligned to tRNAs using this traditional approach. Therefore, we developed a method for tRNA enrichment prior to Illumina sequencing similar to those recently described (see Methods). Briefly, after separation of total RNA on a polyacrylamide gel, a region corresponding to the tRNAs was excised, RNA purified and then either bisulfite treated or directly used as template in library construction. Using this enrichment method, a nearly 20,000-fold increase in the sequence reads aligning to tRNAs was observed, when compared to using total RNA (Fig. 1a). Expression of 56 out of 100 isodecoder consensus sequences from all three genomes, nuclear, chloroplast and mitochondrial was observed using our RBS-seq data. Of these, seven tRNA sequences were ambiguously aligning with two or more genomes (Fig. 1b). A wide-range of tRNA transcript abundances were observed from our RNA-seq data, with chloroplast and mitochondrial derived tRNAs having the highest abundance (Fig. 1c). This is most likely a reflection of the high copy number of plastid and mitochondrial organelles per mesophyll cell.

RBS-seq analysis to identify 5-methylcytosine (m^5C) sites in tRNAs of *A. thaliana*

To identify m^5C sites in tRNAs at single-nucleotide resolution, we performed bisulfite (BS) conversion on enriched tRNAs from wild type *Arabidopsis* that were combined with an *in vitro* transcribed Renilla Luciferase (R-Luc) mRNA BS conversion control lacking m^5C. Complete BS conversion of R-Luc control results in no cytosines and serves as an important internal control. After BS conversion, Illumina libraries were constructed, deep-sequenced and aligned to *in silico* BS converted, cytosine to thymine, endogenous *Arabidopsis* tRNA consensus sequences and the R-Luc control. For a BS converted sample to pass our quality control standards, the R-Luc control required a minimum of 98 % conversion across the 178 cytosines present in the R-Luc mRNA BS conversion control (Additional file 1: Figure S1A). After passing R-Luc quality control, we then determined the global endogenous cytosine abundance. In all stranded RBS-seq libraries, global endogenous cytosine abundance was less than ~1 % compared to ~22 % for non-BS treated RNA-seq samples (Additional file 1: Figure S1B, S1C). Together these results demonstrated that bisulfite conversion of RNA cytosines was highly efficient using our method.

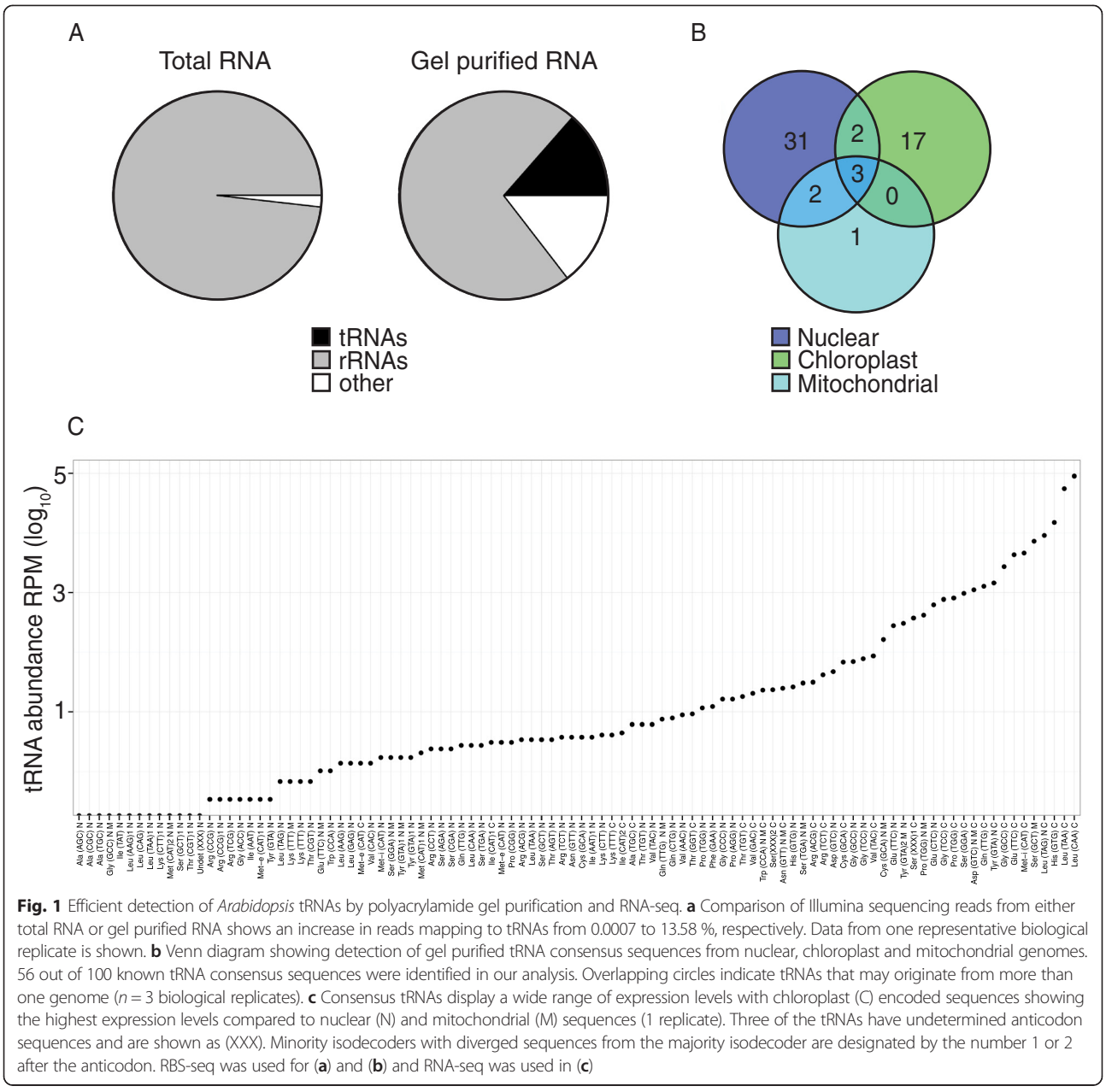


Fig. 1 Efficient detection of *Arabidopsis* tRNAs by polyacrylamide gel purification and RNA-seq. **a** Comparison of Illumina sequencing reads from either total RNA or gel purified RNA shows an increase in reads mapping to tRNAs from 0.0007 to 13.58 %, respectively. Data from one representative biological replicate is shown. **b** Venn diagram showing detection of gel purified tRNA consensus sequences from nuclear, chloroplast and mitochondrial genomes. 56 out of 100 known tRNA consensus sequences were identified in our analysis. Overlapping circles indicate tRNAs that may originate from more than one genome ($n = 3$ biological replicates). **c** Consensus tRNAs display a wide range of expression levels with chloroplast (C) encoded sequences showing the highest expression levels compared to nuclear (N) and mitochondrial (M) sequences (1 replicate). Three of the tRNAs have undetermined anticodon sequences and are shown as (XXX). Minority isodecoders with diverged sequences from the majority isodecoder are designated by the number 1 or 2 after the anticodon. RBS-seq was used for (**a**) and (**b**) and RNA-seq was used in (**c**)

To identify m^5C sites in nuclear, chloroplast and mitochondrial *Arabidopsis* tRNAs, we aligned the Illumina RBS-seq reads against an *in silico* converted tRNA consensus list. *In silico* conversion involved converting all cytosines to thymines. 5-methylcytosine sites were then identified as cytosines that resist bisulfite conversion. These sites are to be noted as candidate m^5C sites, as other types of modified cytosine can also be resistant to bisulfite conversion [29, 51]. We applied a threshold of at least 5 reads aligning to an individual tRNA consensus and a minimum of 20 % methylation. Using these parameters, we identified 24 methylated tRNAs and 32 non-methylated tRNAs out of a total of 56 (Fig. 2a, Additional file 1: Table S2). Interestingly, only nuclear encoded tRNAs were found to contain m^5C sites, whereas non-methylated tRNAs were encoded by all three genomes.

Cytosine methylation of *Arabidopsis* nuclear tRNAs ranged from 23 to 100 %, and were consistent between the three biological replicates. 39 m^5C sites were identified at 5 structural positions and are illustrated on the representative tRNA secondary structure at positions C38, C48, C49, C50 and C72 (Fig. 2b). Methylation at these sites is consistent with observations in other non-plant species [2, 3, 18, 29–31]. Next we examined the pattern of methylation in individual tRNA isodecoders. Seventeen tRNAs were identified with methylation at

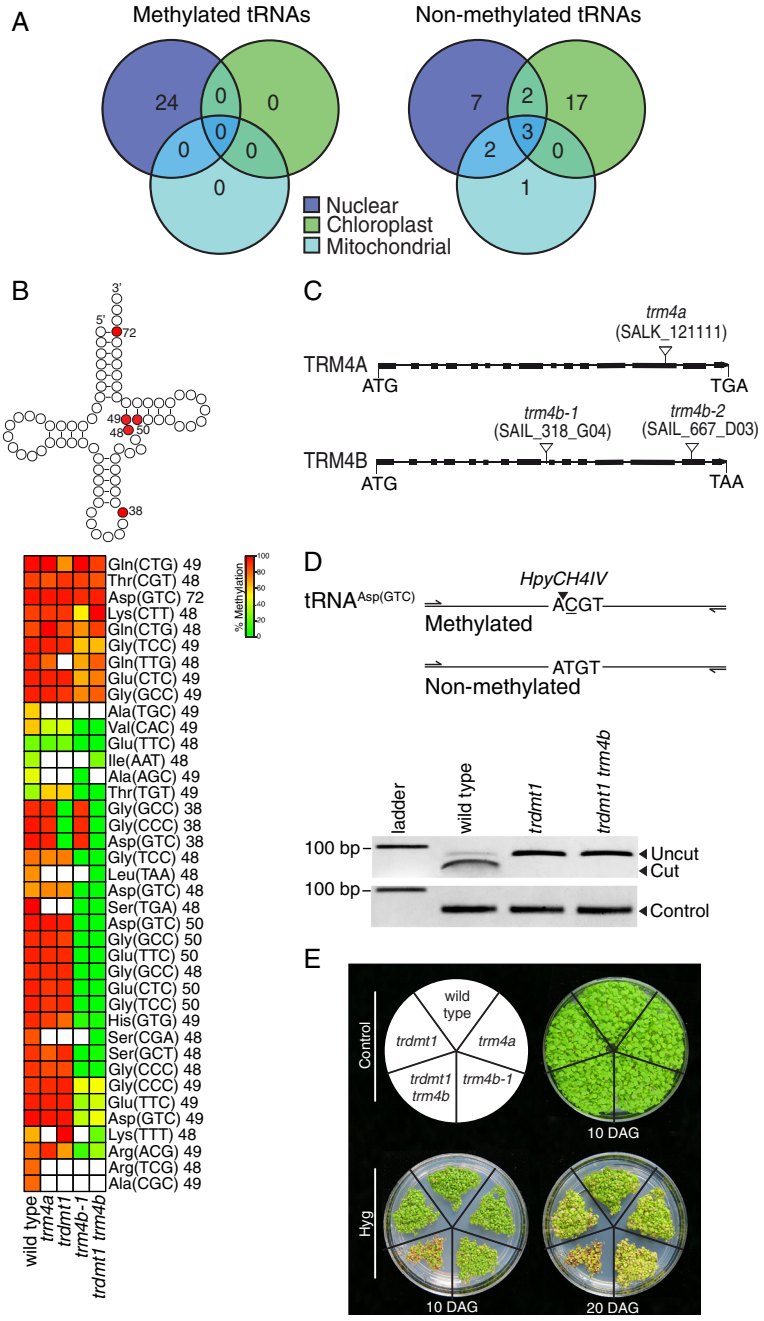


Fig. 2 TRDMT1 and TRM4B methylate *Arabidopsis* nuclear encoded transfer RNAs. **a** Genomic origins of methylated and non-methylated tRNAs. Methylated tRNAs were only detected from the nuclear genome (3 biological replicates). **b** Above: clover-leaf representative secondary structure of tRNA indicating in red, the five cytosine positions methylated in wild type. Below: Heatmap showing percentage methylation of all cytosines detected in nuclear tRNAs of wild type, and mutants *trdmt1*, *trm4a*, *trm4b-1* and *trdmt1 trm4b* using RBS-seq. Cytosine positions are indicated next to tRNA isodecoders. White boxes represent cytosine positions with coverage less than five reads. (wild type 3 biological replicates, mutants $n = 1$). **c** Genomic structure of *trm4a* and *trm4b* mutants showing T-DNA insertions (triangles) in exons (filled boxes). **d** Analysis of RNA methylation by TRDMT1 at position C38 on BS treated $tRNA^{Asp(GTC)}$ template. Above: Restriction maps of PCR amplified products showing the expected digest patterns of methylated and non-methylated template. Below: Cleavage of PCR amplified product by HpyCH4IV confirms C38 methylation in wild type as opposed to non-methylated C38 in *trdmt1* results in loss of HpyCH4IV restriction site. Loading control is undigested PCR product. **e** Hygromycin B stress assay. *Trdmt1 trm4b* double mutants and to a lesser extent, *trm4b-1* mutants display increased sensitivity to hygromycin B (Hyg) at 10 and 20 days after germination (DAG) compared to controls

only 1 structural position, while the other remaining 7 tRNAs contained 2–5 m^5C sites per tRNA. The most frequently methylated sites corresponded to structural positions C48, C49 and C50, indicating that methylation in this region may be important for tRNA structure or stability. $tRNA^{Asp(GTC)}$ was the most highly methylated tRNA and was the only tRNA containing methylation at all 5 structural positions. The structure of $tRNA^{Asp(GTC)}$ may require these additional m^5C sites for greater stability or resistance to cleavage.

Identification of TRM4B and TRDMT1 dependent m^5C sites in nuclear tRNAs

To confirm the m^5C sites in *Arabidopsis* nuclear tRNAs and determine the RMTases required for methylation, we identified mutants for the predicted *Arabidopsis* homologs of RMTases TRM4 and TRDMT1 and then performed RBS-seq on libraries enriched for tRNAs.

Two TRM4 paralogs were identified in the *Arabidopsis* genome [25] and we refer to them as TRM4A and TRM4B. T-DNA mutations in *TRM4A* or *TRM4B* were identified and the homozygous mutants characterized by semi-quantitative RT-PCR to demonstrate null expression (Fig. 2c and Additional file 1: Figure S2C) and show mutants are most likely complete loss of function. Mutants *trm4a*, and the two isolated T-DNA mutants for *TRM4B*; *trm4b-1 and trm4b-2* were grown on soil and appeared phenotypically similar to wild type like the previously characterized RMTase mutant *trdmt1* [27] (Additional file 1: Figure S2A). To test for divergent functions of TRM4A and TR4MB, the m^5C single-nucleotide profile of tRNAs was determined in the mutants (Fig. 2b). In *trm4a*, the m^5C profile was the same as wild type, showing that TRM4A is not required for methylation of any of the detected tRNAs. In contrast for *trm4b-1* and *trm4b-2,* a total of 18 sites had no detectable methylation and 7 sites had reduced methylation when compared to wild type (Fig. 2b and Additional file 1: Figure S3A). The sites that lost methylation or had reduced methylation corresponded to structural positions C48, C49 and C50 which is consistent with animal studies [2, 3, 18, 29–31].

Further investigation of the functional motifs of TRM4A and TRM4B by sequence alignment demonstrated that TRM4A is missing motif I (Additional file 1: Figure S4A). Motif I is essential for methyltransferase activity and is required for AdoMet binding and catalysis [52]. Loss of motif I in TRM4A most likely explains why no reduction in tRNA m^5C levels was observed in *trm4a*. However we cannot exclude the possibility that TRM4A has other functional roles. As TRM4B contains all of the predicted motifs required for RMTase activity and there is a reduction in m^5C tRNA methylation in the *trrm4b* mutants, this demonstrated that TRM4B is the functional homolog of TRM4/NSUN2 in *Arabidopsis thaliana*.

TRDMT1 was previously reported to methylate three tRNAs, $tRNA^{Asp(GTC)}$, $tRNA^{Gly(GCC)}$ and $tRNA^{Val(AAC)}$ at structural position C38, in animals [11, 12, 27, 30] and $tRNA^{Asp(GTC)}$ in *Arabidopsis* [27]. RBS-seq analysis of wild type *Arabidopsis* and *trdmt1* not only confirmed that TRDMT1 is required for position C38 methylation of $tRNA^{Asp(GTC)}$ but is also required for C38 methylation of $tRNA^{Gly(CCC)}$ and $tRNA^{Gly(GCC)}$ in plants as these sites had no detectable methylation in *trdmt1*. In contrast to animals, position C38 of $tRNA^{Val(AAC)}$ is not methylated in *Arabidopsis* (Additional file 1: Table S2). All other detected tRNAs were not methylated at position C38.

Nine m^5C sites in nuclear tRNAs did not show a reduction of methylation in *trm4a-1, trm4b-1* or *trdmt1* single mutants when compared to wild type. These sites occur at structural positions C47, C48, C49 and C72 and are shown clustered together at the top of the heatmap (Fig. 2b). To exclude the possibility of functional redundancy of TRM4B and TRDMT1, we constructed a *trdmt1 trm4b* double mutant and then performed RBS-seq. All 9 sites were methylated in the double mutant and therefore we concluded that no functional redundancy of TRM4B and TRDMT1 for methylation of specific cytosine residues occurs in *Arabidopsis*. We cannot rule out the possibility that these 9 sites are cytosines with other RNA modifications that, like m^5C, are also resistant to bisulfite conversion and therefore are independent of TRM4A, TRM4B, or TRDMT1 methylation.

To further demonstrate the reproducibility of our tRNA methylation data, we developed a rapid PCR-digestion assay to investigate individual m^5C sites derived from BS treated RNA. Position C38 of $tRNA^{Asp(GTC)}$ coincides with the restriction enzyme digestion site, ACGT, of HpyCH4IV. Methylation of C38 protects the site from BS conversion maintaining the HpyCH4IV site in methylated $tRNA^{Asp(GTC)}$ derived PCR products. Therefore HpyCH4IV only cleaves $tRNA^{Asp(GTC)}$ PCR products when position C38 is methylated. Methylation of $tRNA^{Asp(GTC)}$ at position C38 by TRDMT1 was confirmed using the digestion assay on wild type and *trdmt1* BS treated RNA (Fig. 2d). As expected, C38 of $tRNA^{Asp(GTC)}$ is not methylated in *trdmt1* or *trdmt1 trm4b* double mutants and is not cleaved by HpyCH4IV after BS treatment. The rapid digestion assay confirmed our RBS-seq data.

To test the role of tRNA m^5C sites in regulating translation, the antibiotic hygromycin B, hereafter described as hygromycin, was used to perturb translation. Hygromycin alters the conformation of the A-site in the ribosome, which increases binding of tRNAs to the A-site, inhibits translocation and reduces translational fidelity [53]. The tRNA RMTases TRDMT1 and TRM4B mutants are expected to be more sensitive to hygromycin, as the loss of

methylation is predicted to weaken the structural integrity of select tRNAs and increase the ability of hygromycin to bind and 'lock' tRNAs in the A-site, stopping translocation. Therefore we tested this expectation by growing wild type and mutants on control and hygromycin containing plates. Both *trm4b* and t*rdmt1 trm4b* double mutants displayed increased sensitivity to hygromycin at 10 and 20 days after germination (DAG) when compared to the controls (Fig. 2e). The sensitivity of *trm4b* mutants to hygromycin is more apparent at 20 days DAG than at 10 DAG. As a number of tRNAs lose methylation in *trm4b* and t*rdmt1 trm4b* mutants (Fig. 2b) and previous reports that loss of methylation affects tRNA structure, we attribute the hygromycin sensitivity of the mutants to a modified tRNA structure and the increased interaction between these tRNAs and the A-site of the ribosome reducing translation.

Identification of m^5C sites in *Arabidopsis* nuclear, chloroplast and mitochondrial ribosomal RNAs

To identify m^5C sites in rRNAs from *A. thaliana*, we first constructed a list of rRNA sequences to represent all rRNAs from nuclear, mitochondrial and chloroplast genomes (Additional file 1: Table S3). Then we *in silico* bisulfite converted all cytosines to thymines before aligning the RBS-seq data. RBS-seq transcriptome libraries from total RNA were sequenced and efficient bisulfite conversion of cytosine residues was determined as previously described (Additional file 1: Figure S1A, S1B and Methods).

We identified a total of 7 m^5C sites in the nuclear LSU 25S rRNA, chloroplast SSU 16S, LSU 23S and mitochondrial SSU 18S and LSU 26S rRNAs (Fig. 3a, b). This pattern is in contrast to tRNA methylation, which was only detected on nuclear tRNAs (Fig. 2a). Each methylated rRNA contained one m^5C site except for the nuclear LSU 25S and chloroplast LSU 23S rRNAs that each contained two m^5C sites (Fig. 3b). Of the 7 m^5C sites, 6 were highly methylated in all three biological replicates and the average wild type methylation levels ranged from 66 to 82 %. In contrast, one m^5C site, C960 in mitochondrial SSU 18S rRNA, was lowly methylated, with an average of 28 % methylation (Fig. 3b). There were 6 rRNA species that were not methylated (Fig. 3a and Additional file 1: Table S3).

NSUN5 is required for m^5C at position C2268 in nuclear LSU 25S rRNA

Two positions, C2268 and C2860, in nuclear LSU 25S rRNA were highly methylated in our RBS-seq datasets and both sites occur in the conserved domain IV of the large rRNA subunit in helices 70 and 89, respectively. Recently, for the orthologous positions C2278 and C2870 in the yeast nuclear LSU 25S rRNA, the RMTases RCM1 and NOP2 were shown to be required for methylation, respectively [13, 15]. Therefore, we predicted that the *Arabidopsis* homolog of RCM1, described here as NSUN5, and NOP2 paralogs described here as NOP2A/OLI2, NOP2B and NOP2C would be required for m^5C at these sites [25, 42]. To test these predictions we performed RBS-seq on *nsun5*, *nop2a*, *nop2b* and *nop2c* mutants (Fig. 3c, Additional file 1: Figure S2B, S2C).

To test if NSUN5 is required for m^5C at position C2268 of nuclear LSU 25S rRNA we performed RBS-seq on total RNA from *nsun5-1* and wild type (Fig. 3b). Methylation was reduced from 66 % in wild type to 2 % in *nusn5-1* at position C2268 and methylation was not reduced at any other rRNA m^5C sites. Similar results were obtained for a second, independent allele, *nsun5-2* (Additional file 1: Figure S3D). Methylation of C2268 was reduced to 29 % in *nsun5-2*. The low level of background methylation in *nsun5-2* may be due to low levels of NSUN5 expression in this mutant. While no transcripts were detected spanning the T-DNA insertion site, (Additional file 1: Figure S2C) spurious splicing may be occurring at low frequency to produce a small amount of functional, truncated protein. To confirm reduced methylation at position C2268 in nuclear 25S rRNA in *nsun5* mutants, we developed a restriction enzyme digestion of PCR products using a dCAPs (derived cleaved amplified polymorphic sequences) primer derived from BS treated 25S rRNA. Cytosine methylation of C2268 retains the HinfI restriction site and the enzyme cleaves the PCR products in wild type (Fig. 3d and Additional file 1: Figure S3E). A reduction of C2268 methylation in *nsun5-1* and *nusn5-2* was observed by reduced cleavage of PCR products. Together these results demonstrate that C2268 25S rRNA is methylated by NSUN5 in *Arabidopsis*.

Next we tested if NOP2A, NOP2B or NOP2C were required for methylation at C2860 of nuclear LSU 25S rRNA by RBS-seq from the mutants (Fig. 3b and Additional file 1: Figure S3D). All mutants, *nop2a*, *nop2b* and *nop2c* had wild type levels of methylation at C2860 25S rRNA, suggesting these RMTases do not methylate this site or are functionally redundant. To address this question, we attempted to identify *nop2a nop2b* double mutants, however these double mutants could not be identified from a segregating population. This suggests that NOP2A and NOP2B may act redundantly and are essential for plant viability. Sequence alignment of NOP2A, NOP2B and NOP2C revealed that NOP2B is missing motif IV, which is predicted to be involved in release of methylated RNA from the enzyme [54, 55] and NOP2C has an altered motif N1, which is involved in RNA binding, but is not essential for RMTase activity, as TRM4 homologs do not contain this motif [56] (Additional file 1: Figure S4B). Further research is required to uncover the RMTase(s) responsible for this m^5C site and the redundancy of NOP2 paralogs in *Arabidopsis*. We

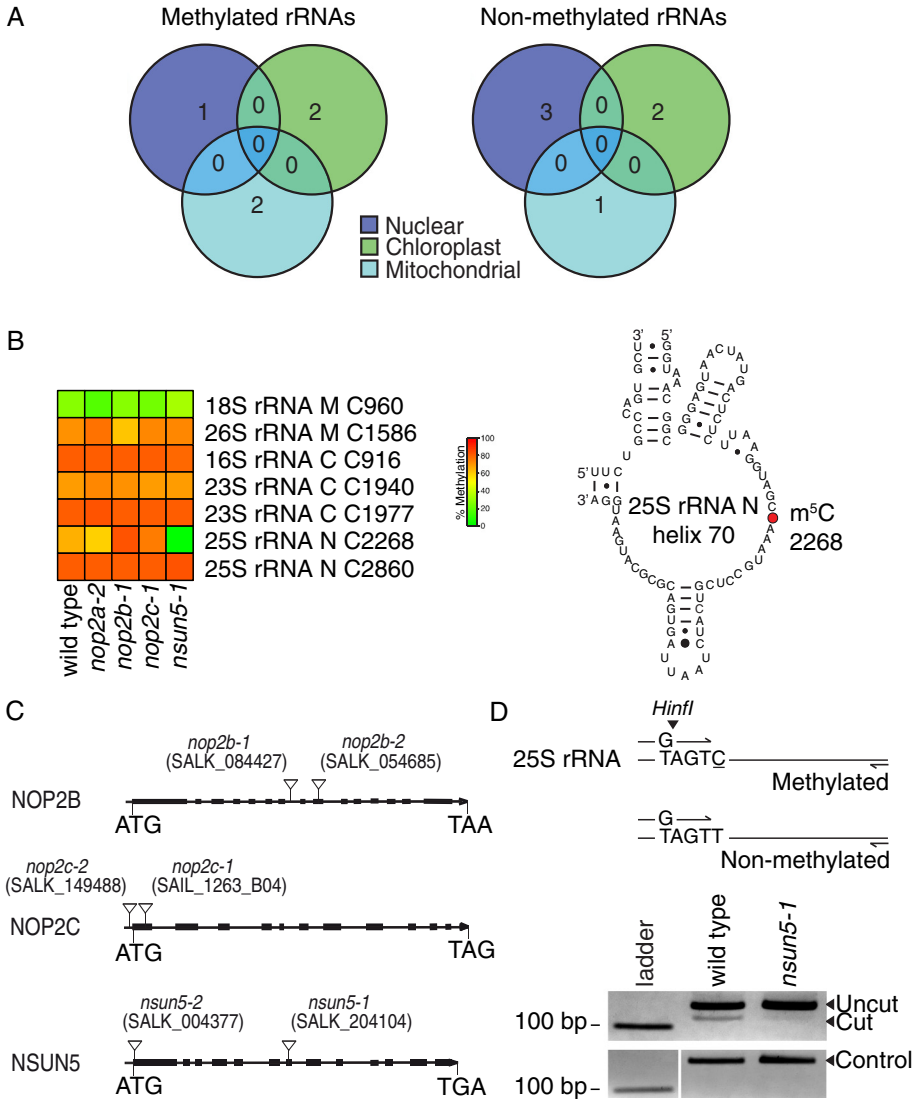


Fig. 3 NSUN5 methylates C2268 in *Arabidopsis* nuclear LSU 25S rRNA. **a** Genomic origins of methylated and non-methylated rRNA species. Methylated rRNAs were detected from all three genomes (3 biological replicates). **b** Left: Heatmap showing percentage methylation of cytosines in nuclear (N), chloroplast (C) and mitochondrial (M) rRNA sequences in wild type and mutants *nop2a-2*, *nsun5-1*, *nop2b-1* and *nop2c-1*. Cytosine positions are indicated next to rRNAs (3 biological replicates). Right: Partial secondary structure of 25S nuclear LSU rRNA helix 70 (domain IV) showing the cytosine position 2268 in red, which is methylated by NSUN5. **c** Genomic structure of *nop2b*, *nop2c* and *nsun5* mutants showing T-DNA insertions (triangles) in exons (filled boxes). **d** Analysis of RNA methylation by NSUN5 at position C2268 on BS treated nuclear LSU 25S rRNA template. Above: Restriction maps of dCAPS amplified products showing the expected digest patterns of methylated and non-methylated template. The 25S_rRNA_F dCAPS primer contains a G mismatch at position four to generate a HinfI restriction site when C2268 is methylated. Below: Cleavage of PCR amplified product by HinfI confirms C2268 methylation in wild type as opposed to non-methylated C2268 in *nsun5-1* results in loss of HinfI restriction site. Loading control is undigested PCR product

also tested if the tRNA RMTases TRM4A, TRM4B and TRDMT1 methylate the remaining 6 m^5C sites in rRNAs by RBS-seq from the mutants, *trm4a*, *trm4b-1*, *trdmt1*, *trdmt1 trm4b* and wild type (Additional file 1: Figure S3C). As expected, no reduction in rRNA methylation levels for the 7 m^5C sites was observed in the mutants. Similarly, to demonstrate NOP2A and NSUN5 are rRNA specific and do not methylate tRNAs, we performed RBS-seq from both *nop2a-2* and *nsun5-2*. No reductions in m^5C tRNA sites were observed (Additional file 1: Figure S3B).

tRNA and rRNA m^5C sites are conserved from single-celled algae to multicellular plants

To test if methylated sites in nuclear tRNAs and organelle rRNAs are conserved through evolution, we constructed tRNA enriched RBS-seq libraries from six organisms; the single-celled algae, *N. oculata,* the multicellular

macro algae *C. taxifolia,* and four vascular plants, the monocotyledonous plant *T. durum, the* dicotyledonous plants *A. thaliana* and *B. rapa* and the evolutionarily distinct *ginkgophyte plant G. biloba*. First, to identify transcribed tRNAs in non-*Arabidopisis* species, we mapped RNA-seq and RBS-seq to both our *Arabidopsis* tRNA isodecoder consensus sequences (Additional file 1: Table S1) and unique tRNA sequences from the closest relative with annotated tRNAs from the PlantRNA Database [49]. Similarly to construct species-specific rRNA references, we performed RNA-seq from total RNA from the five organisms and aligned the reads to either *Arabidopsis* rRNA references, species-specific rRNA references, or an *Arabidopsis*-rRNA guided assembled reference (Additional file 1: Table S3). These species-specific rRNA references were then utilized to align and annotate subsequent RBS-seq reads.

To test for conservation of m^5C of tRNAs we performed RBS-seq on tRNA enriched libraries from *N. oculata, C. taxifolia*, *T. durum*, *B. rapa*, *A. thaliana*, and *G. biloba* and detected 35, 30, 51, 48, 56 and 34 tRNA isodecoders respectively (Fig. 4, Additional file 1: Table S2 and Table S4). Of these tRNAs, 30 were nuclear tRNAs, which are for the greater part methylated across all six species and the remaining 8 were putative chloroplast or mitochondrial tRNAs methylated in only one of the two species, *T. durum* or *N. oculata*. As these tRNAs were only methylated in one of the six species this may reflect chloroplast or mitochondrial tRNAs recently integrated into the nuclear genome of *T. durum* or *N. oculata*. Together these data demonstrate that methylation of chloroplast or mitochondrial encoded tRNAs is rare in the Kingdom *Plantae* and m^5C methylation of tRNAs is generally restricted to nuclear-encoded tRNAs.

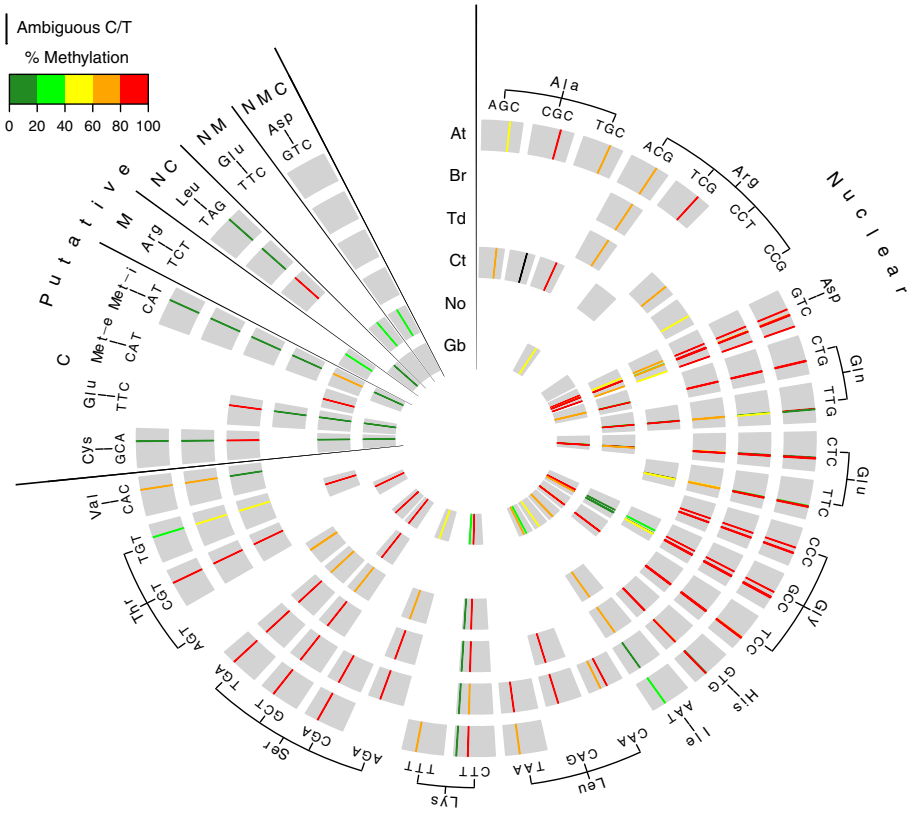


Fig. 4 Conservation of tRNA methylation in Kingdom *Plantae*. **a** Concentric circles from outer to inner represent *Arabidopsis thaliana* (At), *Brassica rapa* (Br), *Triticum durum* (Td), *Nannochloropsis oculata* (No), *Caulerpa taxifolia* (Ct) and *Ginkgo biloba* (Gb) tRNAs, respectively. The circles are split into two major sections for nuclear encoded tRNAs and tRNAs with putative genomic origins (nuclear-N, chloroplast-C, mitochondrial-M). In each circle, individual tRNA consensus sequences are indicated as thick grey arcs and are organized alphabetically by amino acid, and then by anticodon. Specific tRNAs sequences for each species were aligned based on structural positions corresponding to the 72 bp representative tRNA structure. Cytosines that are methylated in at least one of the 6 species analysed are shown as a color-coded percentage methylation bar. The percentage methylation scheme used, Green = lowly methylated (0–40 %), red = highly methylated (80–100 %). Absence of a methylation bar indicates that the corresponding position in the tRNA does not contain a cytosine in that species. A black bar at position 49 in $tRNA^{Ala(CGC)}$ in Ct represents an ambiguous nucleotide which may be a C or T at this position. tRNAs that were not detected in the RBS-seq are not shown (*Arabidopsis thaliana*- 3 biological replicates and all other plant species 1 replicate)

Detailed analysis of the 30 conserved nuclear tRNA isodecoders identified a total of 51 methylated positions. These 51 sites were divided into three classes, class one contained 35 highly conserved sites across all six species, class two contained 5 highly conserved sites in five species and the other species contained a single-nucleotide polymorphism (SNP) and the third class contained 11 sites which are methylated in at least one species and not methylated in the other species. Class two that contained SNPs, were either transitions (C > T) or transversions (C > G) at the methylated positions. An example of a transversion occurs in $tRNA^{Asp(GTC)}$. At position C50 in $tRNA^{Asp(GTC)}$ in *C. taxifolia* had a transversion from C to G, abolishing an otherwise highly conserved m^5C site. The G transversion was confirmed by using RNA-seq. An example of class three, m^5C site reduction in one species, was position C48 of $tRNA^{Glu(CTC)}$. While *T.durum* and *G.biloba* had low levels of methylation (22–40 %) at C48, three other species were not methylated at this site, despite the presence of a cytosine residue in non-BS converted RNA.

Within class three, containing conserved cytosine residues methylated in at least one species, a noteworthy example was $tRNA^{Gln(TTG)}$ which contained methylated positions in all species however sites were not conserved. For example, in *T. durum* and *N. oculata* positions C48 and C49 were both methylated however in the other tested species only C48 or C49 was methylated, but not both sites despite the presence of cytosines at these positions. This site variability was also identified by Blanco et al. [18], as mice are methylated at one site in $tRNA^{Gln(TTG)}$, while humans are methylated at two sites. A clearer understanding of the other ribonucleotide modifications near these tRNA positions may provide further insight into these observations.

We also identified two additional m^5C structural positions, C34 and C68 in $tRNA^{Leu(CAA)}$ and $tRNA^{Lys(CTT)}$ of *B. rapa* and *G. biloba,* respectively, that were not methylated in other species. $tRNA^{Leu(CAA)}$ position C34 methylation was only detected in *B. rapa* and *G. biloba* at 89 and 20 %, respectively. The variation of methylation at this position may be due to environmental factors, as methylation at this site in yeast was previously shown to be altered under oxidative stress conditions [10]. It is predicted that $tRNA^{Leu(CAA)}$ position C34 is methylated in *Arabidopsis* but we did not detect *Arabidopsis* $tRNA^{Leu(CAA)}$ in our datasets. For $tRNA^{Lys(CTT)}$ position C68, *G. biloba* had 25 % methylation while *A. thaliana*, *B. rapa* and *T. durum* had very low methylation (below our 20 % methylation threshold). Similarly, methylation at nearby structural positions C67 and C69 in other tRNAs has also been reported in humans [30].

Conservation of rRNAs m^5C sites was tested amongst all six organisms, *N. oculata, C. taxifolia*, *T. durum*, *B. rapa*, *A. thaliana*, and *G. biloba,* by RBS-seq from total RNA. A total of 8 highly conserved m^5C sites in nuclear, chloroplast and mitochondrial structural positions of LSU and SSU rRNAs were identified (Fig. 5 and Additional file 1: Table S3 and Table S5). Interestingly, methylation of LSU 25S rRNA cytosines C2268 and C2860, which are predicted to be dependent upon homologs of NSUN5 and NOP2A/NOP2B/NOP2C, respectively are conserved in all six species [13, 15]. Six of these 8 m^5C sites were highly conserved in methylation percentage and position across all tested species except C916 in SSU 16S chloroplast rRNA for which the methylation across species ranged from 31 to 87 %. The remaining two highly conserved sites, mitochondrial C960 in SSU 18S rRNA and C1549 in LSU 26S rRNA were highly methylated in four of the six tested species. A further eight m^5C sites, were species specific of which 6, C1703 and C1713-1717, occurred in a 15 bp region on *T. durum* nuclear SSU 18S rRNA and the other two methylated sites, C1566 mitochondrial SSU 18S rRNA and C1887 chloroplast LSU 23S rRNA occurred only in *N. oculata.* The five clustered m^5C sites in 18S rRNA maybe attributed to BS non-conversion events as a result of strong secondary structure of the rRNA. The remaining species-specific sites in *N. oculata* may reflect species-specific factors regulating translation by ribosomes.

Discussion

Here, we show that the post-transcriptional modification 5-methylcytosine is only detected on nuclear-encoded tRNAs of plants however methylation of rRNAs occurs in transcripts from all three organelles. Strong conservation of tRNA and rRNA methylated sites were observed in species ranging from single-celled algae to multicellular plants. Furthermore, in *Arabidopsis thaliana,* the evolutionarily conserved RNA methyltransferases TRM4B and TRDMT1 were found to be required for tRNA methylation at multiple nucleotide sites, while NSUN5 specifically methylates nuclear LSU 25S rRNA at position C2268.

Our study detected 39 candidate sites for 5-methylcytosine in *Arabidopsis* nuclear tRNAs and an additional 20 m^5C sites were detected across diverse plant species and all sites except one are new discoveries in plants. The majority of m^5C sites were found at positions within tRNA secondary structure known to have 5-methylcytosine in animals [2, 3, 18, 29–31], broadly supporting existing expectations of the role of m^5C in modulating tRNA function [2]. An emerging facet of tRNA biology in both plants and animals is their processing into smaller regulatory RNAs [57–61], and TRDMT1- mediated addition of m^5C has been demonstrated to protect tRNAs against heat and oxidative stress-mediated cleavage in *Drosophila* [12]. Likewise, methylation by TRM4/NSUN2 in humans and mouse has been demonstrated to protect tRNAs from oxidative stress

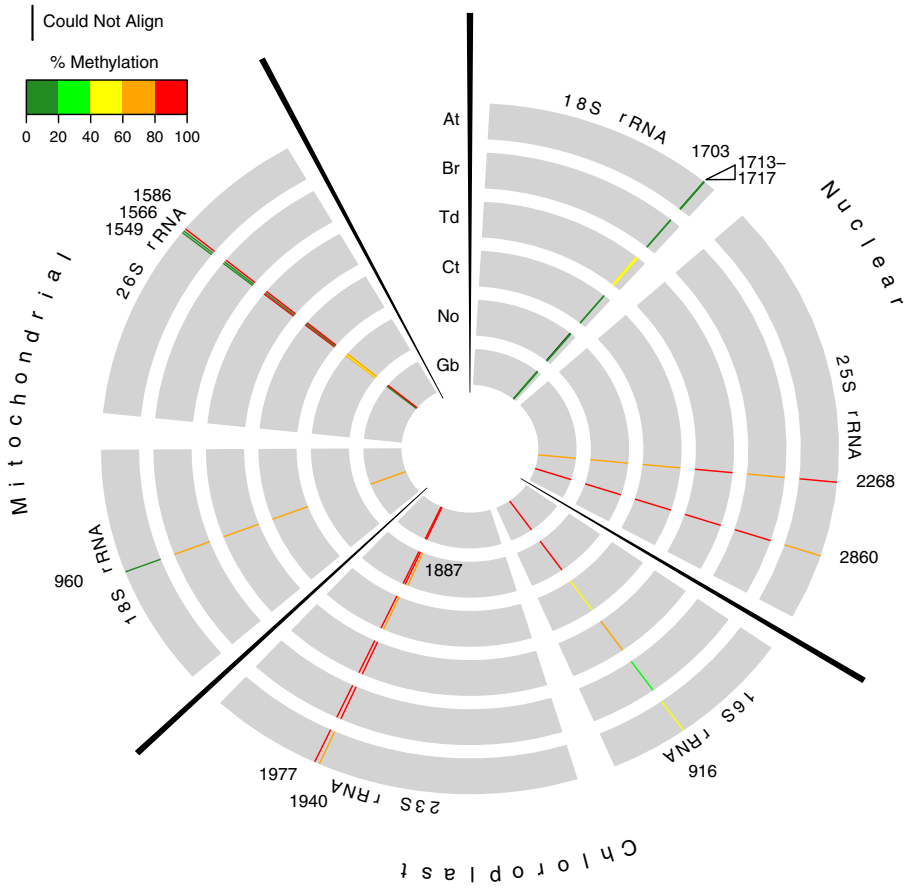


Fig. 5 Conservation of rRNA methylation in Kingdom Plantae. **a** Concentric circles from outer to inner represent *Arabidopsis thaliana* (At), *Brassica rapa* (Br), *Triticum durum* (Td), *Nannochloropsis oculata* (No), *Caulerpa taxifolia* (Ct) and *Ginkgo biloba* (Gb) rRNAs, respectively. The circles are split into three sections for nuclear, chloroplast and mitochondrial encoded rRNAs. In each circle, individual rRNA sequences are represented as thick grey bars. The nucleotide positions for each rRNA species are based on alignment with the corresponding *Arabidopsis* consensus sequences. Cytosines that are methylated in at least one of the 6 species analysed are shown as a color-coded percentage methylation bar. In the percentage methylation scheme used, Green = lowly methylated (0–40 %), red = highly methylated (80–100 %). Absence of a methylation bar indicates that the corresponding position in the rRNA does not contain a cytosine in that species. Open triangle indicates where consecutive cytosines are methylated in *Triticum durum*. The black methylation bar at position 1703 in *Nannochloropsis oculata* SSU 18S rRNA shows where the sequence could not be aligned to the *Arabidopsis* reference sequence (At 3 replicates and other species 1 replicate)

induced cleavage [18]. Together, this data provides a wealth of m^5C sites mediated by TRDMT1 and TRM4B that can now be interrogated for a role of this phenomenon in plants. Future experiments will determine if increased cleaved tRNA fragments are observed in the RMTase mutants and testing these mutants under various environmental conditions may highlight additional roles for these genes in modulating stress responses.

Detection of m^5C sites on only nuclear tRNAs in *Arabidopsis* is consistent with the mitochondrial and chloroplast genomes being derived from alpha-proteobacteria [62] and cyanobacterial ancestors [63], respectively. Complementary to our data, m^5C sites were not detected on tRNAs from bacterium *Escherichia coli* and *Bacillus subtilis* [3]. In contrast, six mitochondrial tRNAs in bovine [32] and five mitochondrial tRNAs in human [29, 64] contain methylation at positions C48, C49 and C72. These data suggest that methylation of mitochondrial tRNAs may have evolved independently in animals since the last common ancestor between plants and animals. A lack of methylation of mitochondrial and chloroplast tRNAs was generally conserved across diverse plant species. Three notable exceptions were chloroplast-like $tRNA^{Cys(GCA)}$, $tRNA^{Glu(TTC)}$ and $tRNA^{Leu(TAG)}$ in *T. durum* that we detected methylation however no methylation was observed in the homologues of other plants species. One interpretation of these observations is that the three chloroplast-like tRNAs of *T. durum* represent recent DNA integration events into the nucleus. After nuclear integration and transcription, these tRNAs are methylated by RMTases, presumably TRM4B, TRDMT1 or RCMT9.

Many mitochondrial genes and tRNAs are often incorporated into nuclear genomes and accordingly lost from the organelle genomes over time [46]. As a result, to obtain the full complement of tRNAs required for translation of mitochondrial encoded proteins, requires the import of tRNAs from the nucleus. Nine tRNA isoacceptors are predicted to be imported from the nucleus to the mitochondrion in *Arabidopsis* [44] and several of these tRNAs, such as $tRNA^{Gly(CCC)}$ were found to be methylated in our data. We speculate that m^5C methylation by TRM4B and/or TRDMT1 of these mitochondrial-imported nuclear tRNAs occurs in the cytoplasm before import into the mitochondria. Methylation of nuclear, eukaryotic tRNAs that are imported into the mitochondria implies they are not inherently incompatible with the prokaryotic mitochondrial translation machinery.

Nine putative cytosine RMTase enzymes including TRDMT1/DNMT2 are encoded in the *Arabidopsis* genome of which we demonstrate TRM4B and TRDMT1 methylate tRNAs and not rRNAs [25, 26]. Duplication of TRM4 resulted in two paralogs, TRM4A and TRM4B in *Arabidopsis.* TRM4B retains methyltransferase activity on tRNAs while TRM4A contains a deletion of motif I which is required for target cytosine binding [52]. We cannot rule out that TRM4A contains other regulatory functions, for example regulating m^5C stability or modulating accessibility of m^5C sites to RNA binding proteins. TRM4B methylation of tRNAs at positions C48, C49 and C50 is consistent with the fact that animal and yeast homologues also methylate these tRNA structural positions [2, 3, 16, 18, 29–31]. NSUN2/TRM4 in human has been found to methylate tRNA(s) at position C72 [30]. In contrast, we identified only one C72 methylated position in $tRNA^{Asp(GTC)}$, which was independent of TRM4B. TRDMT1 has been previously described as a tRNA C38 specific RMTase, in animals and methylates $tRNA^{Asp(GTC)}$, $tRNA^{Gly(GCC)}$ and $tRNA^{Val(AAC)}$ [11, 12, 30]. We confirmed this C38 specific observation in *Arabidopsis* by not only detecting the previously described $tRNA^{Asp(GTC)}$, but also two new tRNAs, $tRNA^{Gly(GCC)}$ and $tRNA^{Gly(CCC)}$. The importance of these methylated sites in tRNAs is illustrated in other organisms, as loss of TRM4 and TRDMT1 results in reduced abundance of mature tRNAs and translational efficiency in mice [11]. In addition, tRNA m^5C sites mediated by TRM4 in yeast are required for tolerance to the antibiotic paromomycin, which is an aminoglycoside, like hygromycin B [65]. Similarly our data of *trm4b* mutants increased sensitivity to hygromycin B when compared to wild type suggests TRM4B methylated tRNAs have a role in translational efficiency. Interestingly the loss of both TRDMT1 and TRM4B resulted in a severe sensitivity to hygromycin B. There are only three tRNAs which we found to be methylated by both TRDMT1 and TRM4B suggesting that the m^5C mediated structure of these three tRNAs is important for translation. Translation is tightly regulated in order for organisms to adapt quickly to environmental stresses, such as oxidative and heat stress [10, 66]. Alterations in translation in *Arabidopsis trdmt1 trm4b* mutants may affect translational regulation under stress conditions and reduce plant fitness.

Our identification of m^5C sites at tRNA structural positions C48, C49, C50 and C72 independent of either TRM4B or TRDMT1 is similar to recent observations in mouse using RBS-seq and raises the possibility that another RMTase methylates these sites [18]. This is in contrast to humans where all tRNA m^5C sites are dependent on either TRM4 or TRDMT1 [18]. We propose that the additional tRNA RMTase in plants is the closest TRM4 homologue RCMT9 (At5g66180). This hypothesis could be tested by identification of *rcmt9* mutants and performing RBS-seq on enriched tRNAs as described here. An alternative hypothesis is that the TRM4 and TRDMT1 independent m^5C sites are not m^5C sites but other cytosine modifications that are resistant to BS conversion [29, 51]. This could be tested by performing mass spectrometry on purified tRNAs from plant *trm4b trdmt1* double mutants and determining the presence or absence of m^5C. In *nsun2 trdmt1* double mutant mice, RNA m^5C levels were reduced by at least 90 % compared to wild type mice [11]. It is unclear whether this indicates that additional m^5C sites in tRNAs remain, or if the detected m^5C was derived from contaminating rRNA. We favour the first hypothesis that these TRM4B and TRDMT1 independent sites are *bona fide* m^5C sites as they are in tRNA structural positions which normally contain m^5C and that these sites are methylated by Arabidopsis RCMT9, which shares sequence homology with TRM4 homologues [25].

Our approach not only detected methylated sites in RNAs but also provides a quantitative measure of the percentage of transcripts having this modification. This allowed us to undertake a quantitative comparative analysis of methylation in more than 200 individual sites in tRNAs and 50 individual sites in rRNAs amongst diverse plants. Interestingly both the percent methylation and specific sites in tRNAs and rRNAs were broadly conserved across the six plant species. This strong conservation strongly supports the functional importance for these m^5C sites in roles such as regulating the structure and stability of rRNAs and tRNAs [2].

Our study detected 7 candidate sites for 5-methylcytosine in *Arabidopsis* nuclear, chloroplast and mitochondrial LSU and SSU rRNAs, all of them novel in plants. Many of these high-confidence sites were found at positions within rRNA regions known to have 5-methylcytosine in animals and bacteria [2, 12, 14, 33–36]. Of note, we did not detect LSU 5S rRNA methylation in any plant species analyzed, while in contrast, this rRNA species is methylated by TRM4/NSUN2 in HeLa cells [30, 31]. It is intriguing that while plant chloroplast and mitochondrial tRNAs are devoid of

m^5C, organelle rRNAs contain m^5C. It is unknown whether the rRNAs are methylated inside the chloroplast and mitochondria, or if they are exported to allow addition of modifications from nuclear derived modifying enzymes before they are imported back into the organelles. *Arabidopsis* RMTases NOP2B and RNMT/FMU are both predicted to localize in chloroplasts [26]. This suggests that these RMTases methylate the rRNA inside the organelles. Further study is required on the location of the RMTases, to confirm these findings and to assess where catalysis of m^5C occurs.

There are five RMTases present in the *Arabidopsis* genome, which are predicted to methylate rRNA, based on sequence similarity to rRNA RMTases in other organisms. In this study, we investigated the rRNA m^5C sites requiring the RCM1 homolog, NSUN5 and the NOP2 paralogs NOP2A, NOP2B and NOP2C. Here we showed that *Arabidopsis* NSUN5 is required for methylation of C2278 in nuclear LSU 25S rRNA. Interestingly, the yeast NOP2 ortholog in *Arabidopsis,* NOP2A was not found to be required for m^5C at C2860 of nuclear LSU 25S rRNA. We hypothesise that this may be due to functional redundancy with the other NOP2 paralogs in the *Arabidopsis* genome, NOP2B and NOP2C. It is uncertain if the paralogs NOP2B and NOP2C are functional RMTases. NOP2B lacks motif IV, which is involved in release of target RNA after methylation by motif VI [54, 55]. In yeast, a point mutation in motif IV of the conserved residue Cys^{424} in NOP2 leads to accumulation of mutant *nop2* protein-RNA complexes and cell death [54, 67]. It is possible that NOP2B is utilizing a Cys residue contained in a highly diverged, nonconforming motif IV to evade cell death. An alternative possibility is that although the m^5C site is conserved, that another RMTase in the genome is responsible for methylation at this site. The most promising candidate for this possibility is the *Arabidopsis* homolog of bacterial Fmu, RNMT, which is predicted to methylate rRNA [25, 43].

Conservation of the enzymes and methylation sites in rRNA across species suggests conservation of function. The possible functions include, regulation of protein synthesis, stability and maturation of rRNA and translational fidelity. It remains to be seen if the phenotype of *nop2a* can be linked to any specific m^5C sites, or alterations in rRNA processing. The rRNA m^5C sites and the mutants identified in this work provide a platform to launch studies into the roles of specific rRNA m^5C sites under different environmental conditions.

Conclusions

This comprehensive characterization of the tRNA and rRNA methylation profiles in plants uncovered nuclear specific methylation of tRNAs, while rRNAs are methylated from all three genomes. The method of enriching for tRNAs combined with RBS-seq on wild type and mutants allowed us to identify m^5C sites dependent on NSUN2/TRM4 and to extend the known target range of TRDMT1 to an additional two tRNAs at position C38 in plants. We also determined that NSUN5 is required for m^5C at C2278 in nuclear LSU 25S rRNA and uncovered functional redundancy among the *Arabidopsis* NOP2 paralogs, NOP2A, NOP2B and NOP2C, as loss of one of these three enzymes is insufficient to remove any rRNA m^5C sites, while loss of both NOP2A and NOP2B appears lethal. *Arabidopsis* RMTase enzymes are encoded in the nuclear genome. This suggests movement of either the nuclear encoded RMTase enzymes to organelles, or transport and re-import of organelle rRNAs. We favour the former hypothesis as the *Arabidopsis* Fmu-like RNMT and NOP2B are predicted to be located in organelles [26], which suggests that they methylate rRNA inside the mitochondria and chloroplast, and that NOP2B acts redundantly with NOP2A and NOP2C. While several mitochondrial tRNAs are methylated in vertebrates [29, 32, 64], our data suggests that like bacterial tRNAs, plant mitochondrial and chloroplast tRNAs are not methylated [3]. This suggests that vertebrates gained the ability to methylate mitochondrial tRNAs during evolution. We discovered high levels of conservation of tRNA and rRNA methylation across diverse plant species, including the sites shown to be methylated by TRDMT1, TRM4B and NSUN5 in *Arabidopsis*, indicating that these enzymes are most likely responsible for methylation at these sites in other *Plantae*. The conservation of RMTases and m^5C sites strongly suggests important, conserved functions, which deserve investigation. We investigated the function of tRNA m^5C sites using the antibiotic hygromycin B. Loss of both TRDMT1 and TRM4B resulted in increased sensitivity of mutants to the antibiotic hygromycin B, suggesting that tRNA m^5C sites affect tRNA structure in a combinatorial manner. Our data provides the foundation and characterization of *Arabidopsis* genetic mutants needed to further probe the functions of RNA m^5C in plants.

Methods

Plant material and growth conditions

A. thaliana (Columbia ecotype) and *B.rapa* plants were grown in Phoenix Biosystem controlled environment rooms at 21°C under metal halide lights that provided a level of PAR (photosynthetic active radiation) of 110 μmol of photos/m^2/s. Plants were grown under long day photoperiod conditions of 16 h light and 8 h darkness on soil (Debco Seedling raising mix) or ½ MS media supplemented with 1 % sucrose. *G. biloba* was grown at the Botanic Gardens of Adelaide (34.9181° S, 138.6107° E, Australia). *C. taxifolia* was grown in artificial seawater at 25°C under fluorescent lights that provided PAR of 80

μmol of photos/m^2/s. *N. oculata* was grown in artificial seawater supplemented with nitrogen and phosphorous under fluorescent lights that provided PAR of 40 μmol of photos/m^2/s to log phase. For *Arabidopsis* hygromycin B assays, seeds were sown on control and hygromycin B [15 ug/mL] supplemented ½ MS media with 1 % sucrose. Hygromycin B was purchased from A.G. Scientific, Inc.. Plants were photographed with a Canon PowerShot G15 camera at 10 and 20 days after germination (DAG).

Mutant alleles described are; *nop2a-2 (oli2-2)* (SALK_129648), *nop2b-1* (SALK_084427), *nop2b-2* (SALK_054685), nop2c-1 (SAIL_1263_B04), *nop2c-2* (SALK_149488), *nsun5-1* (SALK_204104), *nsun5-2* (SALK_004377), *trdmt1* (SALK_136635), *trm4a* (SALK_121111), *trm4b-1* (SAIL_318_G04) and *trm4b-2* (SAIL_667_D03). The *trdmt1 trm4b* double mutants were generated using the *trdmt1* and the *trm4b-1* mutant alleles. The *nop2a nop2b* double mutants were generated using the *nop2a-2 (oli2-2)* and the *nop2b-1* mutant alleles. Primers used to identify homozygous T-DNA mutants are provided (Additional file 1: Table S6).

Nucleotide sequence data for the following genes are available from The Arabidopsis Information Resource (TAIR) database under the following accession numbers: NOP2A/OLI2 (At5g55920), NOP2B (At4g26600), NOP2C (At1g06560), NSUN5 (At5g26180), TRDMT1 (At5g25480), TRM4A (At4g40000), TRM4B (At2g22400).

RNA isolation and bisulfite conversion of RNA

Total RNA was isolated from either 10 day old *A. thaliana* seedlings, or *A. thaliana* floral buds, *T. durum* flag leaf, *B. rapa* and *G. biloba* shoot apexes, *C. taxifolia* fronds or log phase growth *N. oculata* cells using the Spectrum Plant total RNA kit (SIGMA-ALDRICH) and contaminating DNA removed using DNase (SIGMA-ALDRICH). To enrich for tRNAs, 10μg of total RNA was separated on a 10 % polyacrylamide gel, the region containing 65–95 nts was removed and RNA was purified. Either total RNA or purified tRNAs were used for library construction using Illumina's TruSeq RNA sample kit v2, as per the manufacturer's instructions. As bisulfite treated RNA is sheared, bisulfite treated samples were quickly processed after addition of the fragmentation buffer. For bisulfite conversion, 200 pg of control *in vitro* transcribed Renilla luciferase (R-Luc) RNA was added to either 2 μg of total RNA or 200 ng of purified tRNAs and converted with sodium metabisulfite (SIGMA-ALDRICH) as previously described [29, 51]. Bisulfite treated total RNA or purified tRNAs were used as template for Illumina library construction as described above. Illumina sequencing was performed on a MiSeq platform at ACRF, Adelaide. For a description of the sequenced libraries, see (Additional file 1: Table S7).

Sequence read mapping and methylation analysis

Sequences were trimmed for adapters and filtered for low quality reads using CLC Genomics Workbench (Qiagen). In order to identify tRNAs and to reduce mapping ambiguity, we collapsed identical and highly similar tRNA isodecoder sequences from The Arabidopsis Information Resource [68] to create a reference consensus list of 100 tRNA isodecoders (Additional file 1: Table S1). For tRNAs from other species, both the Arabidopsis reference consensus list and unique tRNA sequences obtained from the closest relative available in the PlantRNA Database [49], were used. rRNA sequences were obtained from the NCBI Refseq database where available. For species without available rRNA sequences, RNA-seq reads were aligned to the *Arabidopsis* RNA reference sequence for the corresponding rRNA subunit and a consensus was derived in order to obtain species-specific SNPs (Additional file 1: Table S3). RBS-seq reads of tRNAs and rRNAs were mapped to *in silico* converted reference sequences, while RNA-seq reads were mapped to unconverted sequences. CLC Genomics Workbench (Qiagen) was used to align sequence reads to the corresponding tRNA and rRNA reference sequences. In order to compare structural positions in tRNAs with different sequences and lengths, the representative structure model was used [69]. For rRNAs, the reference sequences of all plant species analysed were aligned to the *Arabidopsis* references and the numbering was based on the nucleotide position of the corresponding *Arabidopsis* rRNA reference sequence.

In order to identify methylated cytosines, non-conversion of a cytosine in read sequences was taken to indicate the presence of m^5C. Renilla Luciferase *in vitro* transcribed mRNA lacking m^5C was used to ensure that conversion efficiency was greater than 98 %. To ensure robust detection of methylated sites in tRNAs and rRNAs, a minimum of 5 reads coverage was required and a minimum methylation level of 20 % (≤80 % conversion rate). Percentage methylation at specific positions was calculated as the number of mapped cytosines divided by the combined total number of mapped cytosines and mapped thymines. Heatmaps showing percentage methylation were created in R [70], using the R package "Pretty heatmaps" [71].

Methyl-chop PCR methylation assay

cDNA derived from the bisulfite treated RNA was used to generate PCR products from tRNA Asp(GTC) and helix 70 of 25S nuclear rRNA using the primers forward Asp tRNA_At_FWD and reverse Asp tRNA_At_REV and primers forward 25S_rRNA_F and reverse 25S_rRNA_R, respectively. The 25S_rRNA_F dCAPS primer contains a G mismatch at position four from the 3′ end to generate a HinfI restriction site. PCR products were digested with HpyCH4IV or HinfI restriction enzymes (New England Biolabs), respectively. The $tRNA^{Asp(GTC)}$ 72bp PCR product

is digested by *Hpy*CH4IV resulting in two digestion products of 35bp and 37bp if C38 is methylated, and is undigested if C38 is non-methylated and thus converted to T38 by bisulfite treatment, causing the HpyCH4IV restriction enzyme site to be lost. The 25S rRNA PCR product is 155bp in length. When C2268 is methylated, the restriction enzyme HinfI cleaves the PCR product into two fragments of 29bp and 126bp. Non-methylated 25S rRNA has a T at position 2268 after bisulfite treatment, and this eliminates the HinfI restriction enzyme site, leaving the product at 155bp and undigested. Undigested PCR products were used as loading controls. Primer sequences used are provided (Additional file 1: Table S6).

Semi-quantitative PCR

Semi-quantitative PCR was performed using an Invitrogen SuperScript III kit as per the manufacturer's recommendations from 2 μg of total RNA and oligo-dT primed cDNA synthesis. Semi-quantitative PCR detection of *TRM4A, TRM4B, NSUN5, NOP2B* and *NOP2C* mRNA was performed using the primers provided (Additional file 1: Table S6). For the RNA input control, amplification of the housekeeping gene PDF2A was used with primers forward PDF2_RT-PCR_F and reverse PDF2_RT-PCR_R. Quantitative PCR was performed to test NOP2C mRNA abundance using Roche LightCycler480 and SYBER green.

Availability of supporting data

The data sets supporting the results of this article are available in NCBI's GEO database repository, and are accessible through GEO Series accession numbers GSE68444, GSE68445, GSE68447 and GSE68448.

Additional file

Additional file 1: Figure S1. Efficient bisulfite conversion of non-methylated cytosine residues. **Figure S2**. Characterization of *Arabidopsis thaliana* T-DNA mutants. **Figure S3**. Specificity of tRNA and rRNA MTases. **Figure S4**. Multiple sequence alignment of methyltransferase motifs from *Arabidopsis thaliana* RMTases. The amino acid sequences of two subfamilies of RMTases in Arabidopsis; (A) TRM4A and TRM4B; (B) NOP2A/OLI2, NOP2B and NOP2C were aligned using Clustal Omega [72]. **Table S1**. Unique tRNA isodecoder consensus sequences used for tRNA expression and methylation analysis. **Table S2**. Annotation table of tRNA isodecoder sequences detected in diverse plant species. **Table S3**. Ribosomal RNA sequences used for methylation analysis in diverse plant species and number of m^5C sites. **Table S4**. Methylation % of tRNAs from *A. thaliana*, *B. rapa*, *T. durum*, *C. taxifolia*, *N. occulata* and *G. biloba*. **Table S5**. Methylation % of rRNAs from *A. thaliana*, *B. rapa*, *T. durum*, *C. taxifolia*, *N. occulata* and *G. biloba*. **Table S6**. Primer sequences used in this study [73]. **Table S7**. Read coverage of libraries sequenced. (ZIP 5681 kb)

Abbreviations

rRNA: Ribosomal RNA; tRNA: Transfer RNA; m^5C: 5-methylcytosine; RMTase: RNA methyltransferase; TRM4: Transfer RNA methyltransferase 4; TRDMT1: Transfer RNA Asp methyltransferase 1; RNA-seq: Illumina RNA sequencing; RBS-seq: Illumina RNA BiSulfite sequencing; NOP2: Nucleolar Protein 2; OLI2: Oligocellula 2; NSUN2: NOP2/Sun domain protein 2; NSUN5: NOP2/Sun domain protein 5; BS: Bisulfite; R-Luc: Renilla luciferase; RCMT9: RNA cytosine methyltransferase 9; AdoMet: S-adenosyl-L-methionine; dCAPS: Derived cleaved amplified polymorphic sequences; PAR: Photosynthetic active radiation.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Experiments were designed by AB, RD and IS. Experiments and analysis were equally conducted by AB and RD. The manuscript was equally prepared and edited by AB, RD and IS. All authors read and approved the final manuscript.

Author's information

AB and RD are joint first authors.

Acknowledgements

We thank Jian Qin (Flinders University, Australia) for *N. oculata* cells, Carlos Gurgel (The University of Adelaide, Australia) for *C. taxifolia* rhizomes and the Genomics and Bioinformatics core facilities (IMVS ACRF, Adelaide) for Illumina sequencing. We thank Joy Raison and Simon Baxter (The University of Adelaide, Australia) for assistance in generating the circle plots and critical reading of the manuscript respectively. This research was supported by ARC grants DP0988846 and DP110103805 awarded to I.S. and an APA and a GRDC PhD top-up scholarship awarded to AB.

Author details

[1]School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia. [2]School of Agriculture, Food and Wine, The Waite Research Institute, The University of Adelaide, Adelaide, South Australia 5005, Australia. [3]The University of Adelaide and Shanghai Jiao Tong University Joint International Centre for Agriculture and Health, Adelaide, Australia.

Received: 23 May 2015 Accepted: 24 July 2015
Published online: 14 August 2015

References

1. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9:465–76.
2. Motorin Y, Lyko F, Helm M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. Nucleic Acids Res. 2010;38:1415–30.
3. Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. PLoS Genet. 2013;9, e1003602.
4. Jackman JE, Alfonzo JD. Transfer RNA modifications: nature's combinatorial chemistry playground. Wires RNA. 2013;4:35–48.
5. Motorin Y, Helm M. tRNA Stabilization by modified nucleotides. Biochemistry. 2010;49:4934–44.
6. Alexandrov A, Chernyakov I, Gu WF, Hiley SL, Hughes TR, Grayhack EJ, et al. Rapid tRNA decay can result from lack of nonessential modifications. Mol Cell. 2006;21:87–96.
7. Agris PF. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. EMBO Rep. 2008;9:629–35.
8. Helm M. Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res. 2006;34:721–33.
9. Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ. A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet. 2010;6, e1001247.
10. Chan CTY, Pang YLJ, Deng WJ, Babu IR, Dyavaiah M, Begley TJ, et al. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun. 2012;3.
11. Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, et al. RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nat Struct Mol Biol. 2012;19:900–5.
12. Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M, et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev. 2010;24:1590–5.
13. Gigova A, Duggimpudi S, Pollex T, Schaefer M, Kos M. A cluster of methylations in the domain IV of 25S rRNA is required for ribosome stability. RNA. 2014;20:1632–44.

14. Hong B, Brockenbrough JS, Wu P, Aris JP. Nop2p is required for pre-rRNA processing and 60S ribosome subunit synthesis in yeast. Mol Cell Biol. 1997;17:378–88.
15. Sharma S, Yang J, Watzinger P, Kotter P, Entian KD. Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. Nucleic Acids Res. 2013;41:9062–76.
16. Motorin Y, Grosjean H. Multisite-specific tRNA : m(5)C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: Identification of the gene and substrate specificity of the enzyme. RNA. 1999;5:1105–18.
17. Auxilien S, Guerineau V, Szweykowska-Kulinska Z, Golinelli-Pimpaneau B. The Human tRNA m(5)C methyltransferase Misu is multisite-specific. RNA Biol. 2012;9:1331–8.
18. Blanco S, Dietmann S, Flores JV, Hussain S, Kutter C, Humphreys P, et al. Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. EMBO J. 2014;33:2020–39.
19. Khan MA, Rafiq MA, Noor A, Hussain S, Flores JV, Rupp V, et al. Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. Am J Hum Genet. 2012;90:856–63.
20. Abbasi-Moheb L, Mertel S, Gonsior M, Nouri-Vahid L, Kahrizi K, Cirak S, et al. Mutations in NSUN2 cause autosomal-recessive intellectual disability. Am J Hum Genet. 2012;90:847–55.
21. Martinez FJ, Lee JH, Lee JE, Blanco S, Nickerson E, Gabriel S, et al. Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. Journal of Medical Genetics. 2012;49:380–5.
22. Fahiminiya S, Almuriekhi M, Nawaz Z, Staffa A, Lepage P, Ali R, et al. Whole exome sequencing unravels disease-causing genes in consanguineous families in Qatar. Clin Genet. 2014;86:134–41.
23. Blanco S, Kurowski A, Nichols J, Watt FM, Benitah SA, Frye M. The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate. PLoS Genet. 2011;7, e1002403.
24. Hussain S, Tuorto F, Menon S, Blanco S, Cox C, Flores JV, et al. The mouse cytosine-5 RNA methyltransferase NSun2 is a component of the chromatoid body and required for testis differentiation. Mol Cell Biol. 2013;33:1561–70.
25. Pavlopoulou A, Kossida S. Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. Genomics. 2009;93:350–7.
26. Chen P, Jager G, Zheng B. Transfer RNA modifications and genes for modifying enzymes in Arabidopsis thaliana. BMC Plant Biol. 2010;10:201.
27. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang XY, et al. Methylation of tRNA(AsP) by the DNA methyltransferase homolog Dnmt2. Science. 2006;311:395–8.
28. Rai K, Chidester S, Zavala CV, Manos EJ, James SR, Karpf AR, et al. Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. Genes Dev. 2007;21:261–6.
29. Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res. 2012;40:5023–33.
30. Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat Biotechnol. 2013;31:458–64.
31. Hussain S, Sajini AA, Blanco S, Dietmann S, Lombard P, Sugimoto Y, et al. NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep. 2013;4:255–61.
32. Suzuki T, Suzuki T. A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs. Nucleic Acids Res. 2014;42:7346–57.
33. Sugiura M, Hirose T, Sugita M. Evolution and mechanism of translation in chloroplasts. Annu Rev Genet. 1998;32:437–59.
34. Granneman S, Baserga SJ. Ribosome biogenesis: of knobs and RNA processing. Exp Cell Res. 2004;296:43–50.
35. Lindahl L, Zengel JM. Ribosomal genes in Escherichia coli. Annu Rev Genet. 1986;20:297–326.
36. Maden BE. Locations of methyl groups in 28 S rRNA of Xenopus laevis and man. Clustering in the conserved core of molecule. J Mol Biol. 1988;201:289–314.
37. Purta E, O'Connor M, Bujnicki JM, Douthwaite S. YccW is the m5C methyltransferase specific for 23S rRNA nucleotide 1962. J Mol Biol. 2008;383:641–51.
38. Gu XR, Gustafsson C, Ku J, Yu M, Santi DV. Identification of the 16S rRNA m5C967 methyltransferase from Escherichia coli. Biochemistry. 1999;38:4053–7.
39. Andersen NM, Douthwaite S. YebU is a m5C methyltransferase specific for 16 S rRNA nucleotide 1407. J Mol Biol. 2006;359:777–86.
40. Baer RJ, Dubin DT. Methylated regions of hamster mitochondrial ribosomal RNA: structural and functional correlates. Nucleic Acids Res. 1981;9:323–37.
41. Metodiev MD, Spahr H, Loguercio Polosa P, Meharg C, Becker C, Altmueller J, et al. NSUN4 is a dual function mitochondrial protein required for both methylation of 12S rRNA and coordination of mitoribosomal assembly. PLoS Genet. 2014;10, e1004110.
42. Fujikura U, Horiguchi G, Ponce MR, Micol JL, Tsukaya H. Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of Arabidopsis thaliana. Plant J. 2009;59:499–508.
43. Hebrard C, Trap-Gentil MV, Lafon-Placette C, Delaunay A, Joseph C, Lefebvre M, et al. Identification of differentially methylated regions during vernalization revealed a role for RNA methyltransferases in bolting. J Exp Bot. 2013;64:651–63.
44. Duchene AM, Pujol C, Marechal-Drouard L. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. Curr Genet. 2009;55:1–18.
45. Marechaldrouard L, Weil JH, Dietrich A. Transfer-Rnas and Transfer-Rna Genes in Plants. Annu Rev Plant Phys. 1993;44:13–32.
46. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29:380–95.
47. Michaud M, Cognat V, Duchene AM, Marechal-Drouard L. A global picture of tRNA genes in plant genomes. Plant J. 2011;66:80–93.
48. Yukawa Y, Mizutani T, Akama K, Sugiura M. A survey of expressed tRNA genes in the chromosome I of Arabidopsis using an RNA polymerase III-dependent in vitro transcription system. Gene. 2007;392:7–13.
49. Cognat V, Pawlak G, Duchene AM, Daujat M, Gigant A, Salinas T, et al. PlantRNA, a database for tRNAs of photosynthetic eukaryotes. Nucleic Acids Res. 2013;41:D273–279.
50. Zhou W, Karcher D, Fischer A, Maximova E, Walther D, Bock R. Multiple RNA processing defects and impaired chloroplast function in plants deficient in the organellar protein-only RNase P enzyme. PLoS One. 2015;10, e0120533.
51. Schaefer M, Pollex T, Hanna K, Lyko F. RNA cytosine methylation analysis by bisulfite sequencing. Nucleic Acids Res. 2009;37, e12.
52. Bujnicki JM, Feder M, Ayres CL, Redman KL. Sequence-structure-function studies of tRNA : m(5)C methyltransferase Trm4p and its relationship to DNA : m(5)C and RNA : m(5)U methyltransferases. Nucleic Acids Res. 2004;32:2453–63.
53. Borovinskaya MA, Shoji S, Fredrick K, Cate JHD. Structural basis for hygromycin B inhibition of protein biosynthesis. RNA. 2008;14:1590–9.
54. King MY, Redman KL. RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. Biochemistry. 2002;41:11218–25.
55. Liu YQ. Santi DV: m(5)C RNA and m(5)C DNA methyl transferases use different cysteine residues as catalysts. Proc Natl Acad Sci. 2000;97:8263–5.
56. Foster PG, Nunes CR, Greene P, Moustakas D, Stroud RM. The first structure of an RNA m(5)C methyltransferase, Fmu, provides insight into catalytic mechanism and specific binding of RNA substrate. Structure. 2003;11:1609–20.
57. Zhang SD, Sun L, Kragler F. The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation. Plant Physiol. 2009;150:378–87.
58. Loss-Morais G, Waterhouse PM, Margis R. Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets. Biol Direct. 2013;8.
59. Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA. Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA. 2010;16:673–95.
60. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Suzuki H, Hayashizaki Y, et al. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA Biol. 2011;8:158–77.
61. Sobala A, Hutvagner G. Small RNAs derived from the 5 end of tRNA can inhibit protein translation in human cells. RNA Biol. 2013;10:553–63.
62. Gray MW, Burger G, Lang BF. Mitochondrial evolution. Science. 1999;283:1476–81.
63. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, et al. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol. 2005;15:1325–30.
64. Suzuki T, Nagao A, Suzuki T. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu Rev Genet. 2011;45:299–329.
65. Wu P, Brockenbrough JS, Paddy MR, Aris JP. NCL1, a novel gene for a non-essential nuclear protein in Saccharomyces cerevisiae. Gene. 1998;220:109–17.
66. Merret R, Nagarajan VK, Carpentier MC, Park S, Favory JJ, Descombin J, et al. Heat-induced ribosome pausing triggers mRNA co-translational decay in Arabidopsis thaliana. Nucleic Acids Res. 2015;43:4121–32.
67. King M, Ton D, Redman KL. A conserved motif in the yeast nucleolar protein Nop2p contains an essential cysteine residue. Biochem J. 1999;337(Pt 1):29–35.
68. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012;40:D1202–1210.

69. Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 1998;26:148–53.
70. R Core Team. R: A language and environment for statistical computing. R version 3.1.0 edition. Vienna, Austria: R Foundation for Statistical Computing; 2014.
71. Kolde R. pheatmap: Pretty Heatmaps. R package version 0.7.7 edition; 2013.
72. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 2010;38:W695–699.
73. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen HM, Shinn P, et al. Genome-wide Insertional mutagenesis of Arabidopsis thaliana. Science. 2003;5633:653–7.

Chapter 3

Transcriptome-wide mapping of RNA 5-methylcytosine in *Arabidopsis* mRNAs and ncRNAs

Rakesh David[1], Alice L. Burgess[1], Brian J. Parker[2], Kalinya E. Pulsford[1], Tennille Sibbritt[3], Thomas Preiss[3] and Iain R. Searle[1]

[1]School of Biological Sciences, The University of Adelaide, Adelaide, SA, 5005, Australia, The University of Adelaide and Shanghai Jiao Tong University Joint International Centre for Agriculture and Health

[2]Department of Biology, New York University, New York, NY, 1003-6688, United States of America

[3]The John Curtin School of Medical Research, The Australian National University, Canberra, ACT, 0200, Australia

In preparation for Nature Communications 2016

Statement of Authorship

Title of paper	Transcriptome-wide mapping of RNA 5-methylcytosine in *Arabidopsis* mRNAs and ncRNAs
Publication Status	Unpublished and Unsubmitted work written in manuscript style
Publication details	To be submitted to *Nature Communications*

Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed below);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Candidate	Alice L. Burgess		
Contribution to paper	Equal first author with Rakesh David. Co-designer of experiments. Performed RNA bisulfite treatments. Calculated conversion efficiencies. Analysed the Fluidigm Access Array data. Crossed GUS reporter into *trm4b* mutant. Performed preliminary experiments for short roots and identified oxidative stress sensitivity of *trm4b* (not shown). Wrote the abstract, introduction and co-wrote the results and discussion. Composed figures 1, 2, and 6. Composed supplemental figures .		
Overall percentage (%)	35		
Signature		Date	21/2/2016

Name of Co-Author	Rakesh David
Contribution to paper	Equal first author with Alice Burgess. Co-designer of experiments. Made all RBS-seq and RNA-seq libraries. Generated TRM4B over expression constructs and root complementation experiments. Identified *trm4b* short root phenotype, performed GUS reporter experiments, confocal microscopy and root length experiments under controlled and stress conditions. Wrote the Materials and Methods and co-

	wrote the results and discussion. Supervised development of work and edited the manuscript. Composed figures 3, 4 and 5. Composed supplemental figures. Overall contribution 35%.		
Signature		Date	21/02/2016

Name of Co-Author	Brian J. Parker		
Contribution to paper	Bioinformatic analysis of RBS-seq and RNA-seq data. Wrote Materials and Methods section on bioinformatic analysis. Overall contribution 13%.		
Signature		Date	23/2/2016

Name of Co-Author	Kalinya E. Pulsford		
Contribution to paper	Made constructs and performed sensor fragment experiments. Overall contribution 2%.		
Signature		Date	22/02/16.

Name of Co-Author	Tennille Sibbritt		
Contribution to paper	Contributed to method development. Overall contribution 1%.		
Signature		Date	22/02/2016

Name of Co-Author	Thomas Preiss		
Contribution to paper	Supervised development of work. Overall contribution 4%.		
Signature			10/3/2016

Name of Co-Author	Iain R. Searle		
Contribution to paper	Co-designer of all experiments. Supervised development of work and edited the manuscript. Overall contribution 10%.		
Signature		Date	20/02/2016

Transcriptome-wide mapping of RNA 5-methylcytosine in *Arabidopsis* mRNAs and ncRNAs

David, R[1†]., Burgess, A. L[1†]., Parker, B. J[2]., Pulsford, K. E.[1], Sibbritt, T[3]., Priess, T[3]., Searle, I.R[1*].

[†]Equal contributors

[1]School of Biological Sciences, The University of Adelaide, South Australia, 5005 Australia, The University of Adelaide and Shanghai Jiao Tong University Joint International Centre for Agriculture and Health
[2]Department of Biology, New York University, New York, NY, 1003-6688, United States of America
[3] The John Curtin School of Medical Research, The Australian National University, Canberra, Australian Capital Territory 0200, Australia

* Corresponding author: Iain R Searle **Iain.Searle@adelaide.edu.au**

Corresponding author
Iain Robert Searle
School of Biological Sciences
The University of Adelaide
Adelaide, 5005
Australia
Email: iain.searle@adelaide.edu.au

Keywords: RNA 5-methylcytosine, m^5C, non-coding RNA, mRNA, *Arabidopsis thaliana,* epigenetic, NSUN2, TRM4

Running Title: RNA methylation in plants

Abstract

Post-transcriptional methylation of RNA cytosine residues to 5-methylcytosine (m^5C) is an important modification with diverse roles such as regulating stress responses, stem cell proliferation and RNA metabolism. Here, we use RNA bisulfite sequencing (RBS-seq) for transcriptome-wide quantitative mapping of m^5C in the model plant, *Arabidopsis thaliana*. We discover and validate more than a thousand m^5C sites in *Arabidopsis* mRNAs, long non-coding RNAs and other non-coding RNAs across three tissue types; siliques, seedling shoots and seedling roots. Quantitative differences in methylated sites between these three tissues suggest tissue-specific regulation of m^5C. Perturbing the RNA m^5C methyltransferase TRM4B resulted in loss of m^5C sites on mRNAs and non-coding RNAs and reduced the stability of $tRNA^{Asp(GTC)}$. We also demonstrate the importance of m^5C in plant development as *trm4b* mutants have reduced cell division in the root apical meristem and as a result have a shorter primary root. In addition, *trm4b* mutants show increased sensitivity to oxidative stress. Finally, we demonstrate that a 50 nt sequence flanking m^5C C3349 in *MAG5* mRNA is sufficient to confer methylation of a transgene reporter in *Nicotiana benthamiana*. This provides the first insights into the targeting mechanism of TRM4B and suggests conservation of the targeting mechanism.

Introduction

DNA 5-methylcytosine (m^5C) is a well-characterized epigenetic modification with many functional roles such as transcriptional silencing and genomic imprinting in eukaryotes[1]. While DNA appears to have only a small number of other modifications, for example 5-hydroxymethylcytosine (hm^5C) and N^6-methyladenosine (m^6A)[2], RNA has over 100 different modifications that have been identified in organisms across all three domains of life [3, 4, 5, 6].

The development of high-throughput techniques to identify RNA modification sites transcriptome-wide has enabled deeper investigations into the function of m^5C in diverse classes of RNA[7]. Although m^5C has been described to occur on transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) in many organisms[4, 8], it has only more recently been described transcriptome-wide on mRNA and long non-coding RNAs (lncRNAs) from humans[9, 10, 11]. tRNAs are heavily modified with modifications that have been shown to affect tRNA aminoacylation, codon identification and to stabilize tRNA secondary structure[12, 13, 14, 15]. Of these modifications, m^5C sites in tRNAs commonly occur in the variable region and anticodon loop. m^5C sites in tRNAs are required for tRNA stability and efficient translation[16, 17, 18], and are induced under oxidative stress conditions[19, 20]. Similarly, m^5C sites play important roles in rRNA processing, structure and translation[21, 22, 23]. Additional roles for rRNA m^5C sites in organismal lifespan have also been demonstrated[24]. Several functions have been investigated for m^5C in other non-coding (ncRNAs), such as vault ncRNAs and lncRNAs. While m^5C is required for stability of vault ncRNAs[10], m^5C sites on the lncRNA, X-inactive specific transcript (*XIST)* prevent binding of the Polycomb repressive complex 2 (PRC2) complex *in vitro*[25]. In contrast to ncRNAs, the roles of m^5C in mRNAs requires further investigation. Within human HeLa cell mRNAs, m^5C sites are enriched in untranslated regions and relatively depleted in coding regions[9]. Potential functions such as mediating mRNA stability[26] and RNA induced silencing complex (RISC) activity[9] have been proposed.

In eukaryotes, there are two classes of RNA methyltransferases (RMTases), which catalyse m^5C of both mRNAs and other types of ncRNAs. The first class of eukaryotic RMTase; Transfer RNA aspartic acid methyltransferase 1 (TRDMT1),

also known as DNA methyltransferase 2 (DNMT2), has been shown to methylate tRNAs in animals, plants and fission yeast[8, 27, 28, 29]. Thus far, only two m^5C sites in mRNAs have been shown to require TRDMT1 in human cells[11], suggesting a minor role for TRDMT1 in mRNA methylation. The second class of RMTase is known as tRNA specific methyltransferase 4 (TRM4) or NOP2/Sun domain protein 2 (NSUN2), in yeast and animals respectively[16, 30, 31]. TRM4/NSUN2 depletion in mice results in male infertility, reduced growth and epidermal differentiation defects, revealing a role for NSUN2 in stem cell self-renewal and differentiation[32, 33]. In addition, *NSUN2* mutations in humans and are linked to inherited intellectual disability and reduced growth[34, 35, 36, 37]. Similarly to humans, *Drosophila nsun2* mutants also display short-term-memory deficits[35]. The underlying cause of these neurological symptoms is thought to be mediated by increased cleavage of non-methylated tRNAs by the ribonuclease angiogenin during oxidative stress conditions[38]. Conserved and plant specific functions for TRM4 homologs in plants requires further investigation. The *Arabidopsis* genome encodes two putative TRM4/NSUN2 paralogs, *TRM4A* and *TRM4B*[39, 40]. Previously, we showed that *trm4a* mutants had a similar tRNA methylation pattern to wild type *Arabidopsis* and suggested TR4M4A was catalytically inactive[8]. In contrast, *trm4b* mutants lost methylation at tRNA positions C48, C49 and C50 in *Arabidopsis*.

To further investigate the functions of m^5C and to facilitate future studies of m^5C in plants, we describe transcriptome–wide single nucleotide resolution of post-transcriptionally modified cytosine residues in *Arabidopsis thaliana* by combining RNA bisulfite conversion with second generation Illumina sequencing (RBS-seq). We report over a thousand m^5C sites in *Arabidopsis* transcribed mRNAs, lncRNAs and other ncRNAs. Within mRNAs, the majority of m^5C sites were identified in the coding sequence, however, when normalized for read coverage and sequence length, m^5C sites were observed at higher frequencies than expected in 3'UTR's. We also show that the global methylation level of m^5C sites varies amongst silique, shoot and root tissues and that many m^5C sites are dependent on TRM4B. To test the determinants for targeting of TRM4B, we performed a LOGO motif analysis and were unable to identify a consensus sequence to target TRM4B methylation. We subsequently identified a 50 nt sequence which was able to confer TRM4B dependent methylation of a transgene *in vivo,* suggesting that additional factors, such as RNA structure may

be involved. To determine the role of TRM4B in plant development, we investigated primary root growth. We show *trm4b* mutants have a short root phenotype as a consequence of reduced cell division when compared to wild type. Furthermore, *trm4b* mutants are more sensitive to oxidative stress and have reduced stability of non-methylated tRNAs. These data represent the first genome-wide high-resolution view of m^5C in plants and a link to biological function.

Results

Transcriptome-wide detection of m^5C sites in *Arabidopsis* RNA

To identify transcriptome-wide m^5C sites in *Arabidopsis thaliana* RNA at single-nucleotide resolution, we performed bisulfite conversion of ribosomal depleted RNA from wild type plants, combined with a non-methylated *in vitro* transcribed Renilla Luciferase (R-Luc) mRNA bisulfite conversion control followed by stranded Illumina RNA-sequencing, RBS-seq. We obtained 255, 262 and 116 million paired-end (100 nt) Illumina reads from *Arabidopsis* siliques, shoots and roots, respectively and aligned the sequences to an *in silico* bisulfite converted *Arabidopsis* transcriptome and R-Luc mRNA control (see Methods, Supplementary Table 1). Analysis of our R-Luc controls demonstrated almost complete C to T conversion for all samples. For example, wild type seedling shoots showed >99.8% conversion of the R-Luc control (Supplementary Fig. 1a) and global endogenous cytosine abundance was less than 0.5% compared to approximately 22% for non-bisulfite treated RNA-seq libraries (Supplementary Fig. 1b). We were able to confirm previously identified m^5C sites in rRNAs (mitochondrial 26S rRNA C1586) and tRNAs ($tRNA^{Glu(TTC)}$), demonstrating the robustness of our bisulfite conversion protocol[8] (Supplementary Tables 2, 3). In addition, we PCR amplified, cloned and conventional Sanger sequenced the highly structured $tRNA^{Asp(GTC)}$ and all sequences demonstrated complete bisulfite conversion at known non-methylated sites and no evidence of over conversion of known m^5C sites[8] (Supplementary Fig. 1c). Together these data demonstrate that our bisulfite conversion was highly efficient, with a low false-positive rate, allowing detection of lowly methylated m^5C sites transcriptome-wide.

Discovery of novel m^5C sites in *Arabidopsis* siliques, shoots and roots

To determine tissue-specific methylation profiles, we next examined all reads aligning to the *Arabidopsis* transcriptome to identify hundreds of novel m^5C sites in three different tissue types. In *Arabidopsis* siliques, seedling shoots and roots, 128, 201 and 859 m^5C sites were identified with methylation levels ranging from 1% to 92%, respectively (FDR≤0.3) (Fig. 1a, Supplementary Tables 2, 3). The majority of these sites were tissue-specific and only 15 sites were commonly methylated between all three tissue types. The two tissues showing the greatest number of conserved methylated sites were siliques and shoots, sharing 48 common m^5C sites. When global methylation levels of m^5C sites within siliques, shoots and roots were compared, roots had lower average methylation of m^5C sites when compared to siliques or shoots (p-value ≤0.0001) (Supplementary Fig. 2).

We then asked the question whether the percentage methylation at specific sites in mRNA transcripts expressed in shoots, siliques and roots varied and might indicate functional significance. We compared the m^5C level at common sites and discovered that the level of methylation at specific m^5C sites varied in shoot, silique and root tissues (Supplementary Tables 2, 3). For example, C3349 in At5g47480 (*MAG5*) had 55% methylation in shoots, 33% in shoots and 26% methylation in roots (Fig. 1b).

Annotation of m^5C-containing transcripts from the three tissue types revealed many diverse RNA types, including mRNAs, pseudogenes, lncRNAs, natural antisense transcripts, snoRNAs, snRNAs and other ncRNAs (Fig. 1c). While methylation of specific tRNAs and other types of ncRNAs is associated with increased stability and abundance[10, 16, 41], no link between mRNA abundance and m^5C levels in mammals has been observed so far[10, 16, 33]. To examine if a correlation between m^5C methylation and gene expression exists in *Arabidopsis*, we plotted percentage of m^5C methylation (≥ 2%) identified in the RBS-seq root tissue against RNA-seq gene expression data generated from the same tissue type (Fig. 1d). We observed a small negative correlation of mRNA abundance with higher levels of methylation in *Arabidopsis* roots (r_s = -0.32, p-value≤0.0001) indicating m^5C deposition in these transcripts may affect its stability or turnover rate within the cell.

To further investigate the role of m^5C in mRNAs, we next searched for any bias in m^5C site distribution within mRNAs. This revealed a significant enrichment of m^5C

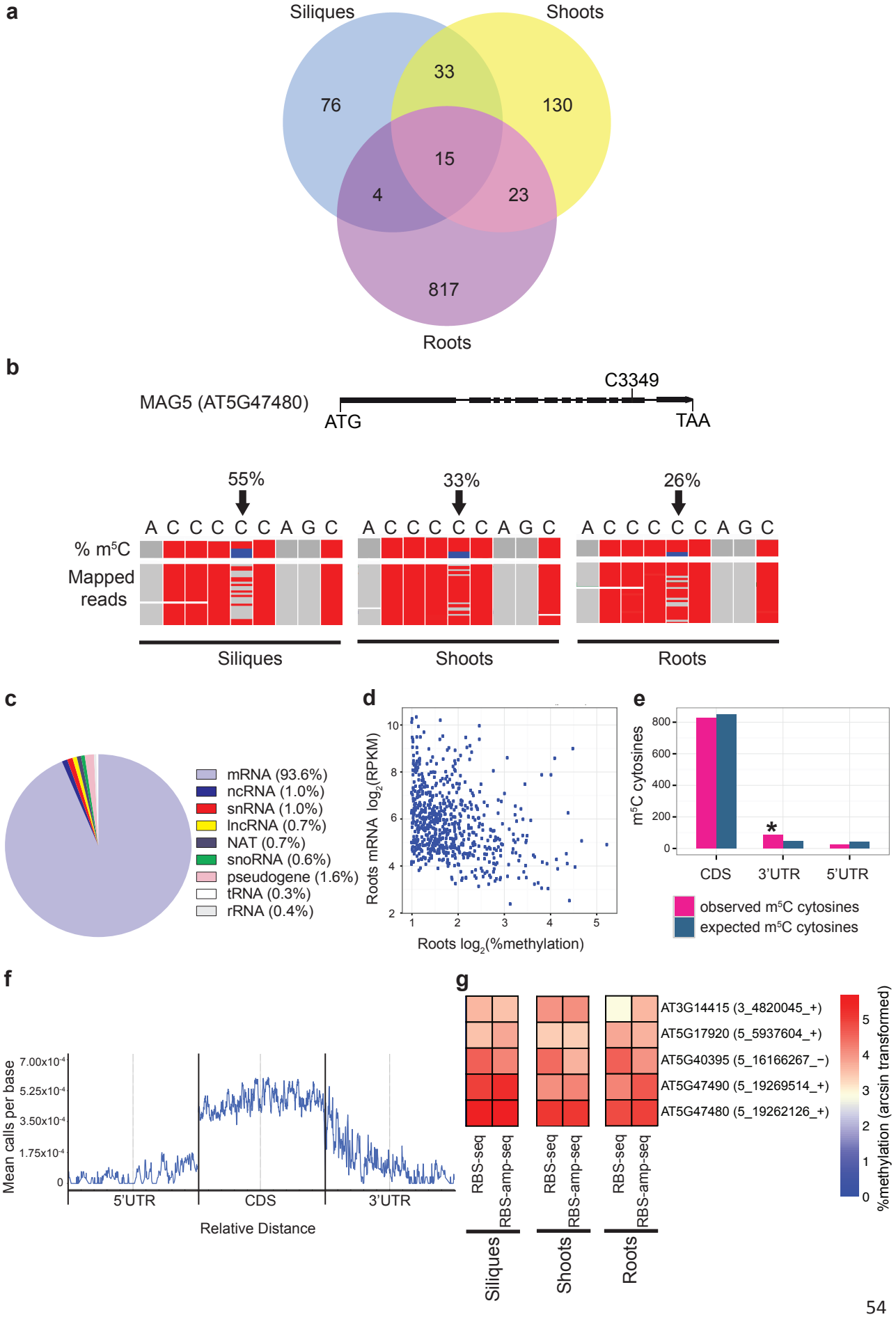

a
Siliques
Shoots
Roots
76
33
130
15
4
23
817
b
MAG5 (AT5G47480)
ATG
TAA
C3349
55%
33%
26%
A C C C C C A G C
% m5C
Mapped reads
Siliques
Shoots
Roots
c
mRNA (93.6%)
ncRNA (1.0%)
snRNA (1.0%)
lncRNA (0.7%)
NAT (0.7%)
snoRNA (0.6%)
pseudogene (1.6%)
tRNA (0.3%)
rRNA (0.4%)
d
Roots mRNA log2(RPKM)
Roots log2(%methylation)
e
m5C cytosines
CDS
3'UTR
5'UTR
observed m5C cytosines
expected m5C cytosines
f
Mean calls per base
7.00x10-4
5.25x10-4
3.50x10-4
1.75x10-4
0
5'UTR
CDS
3'UTR
Relative Distance
g
AT3G14415 (3_4820045_+)
AT5G17920 (5_5937604_+)
AT5G40395 (5_16166267_−)
AT5G47490 (5_19269514_+)
AT5G47480 (5_19262126_+)
RBS-seq
RBS-amp-seq
Siliques
Shoots
Roots
%methylation (arcsin transformed)

Figure 1: Transcriptome-wide mapping of m^5C using RBS-seq in *Arabidopsis thaliana.* (**a**) Venn diagram showing m^5C sites in *Arabidopsis* wild type (Col-0) siliques at 1 days after pollination (DAP), seedling shoots at 10 days after germination (DAG) and seedling roots at 6 DAG using RBS-seq (n=2-3). Methylated sites are called as significant at (FDR ≤0.3 and ≥1% methylation). (**b**) Gene schematic and genome browser view of a differentially methylated gene, *MAG5* (AT5G47480) in three different tissue types. In the *MAG5* gene schematic, the methylated cytosine C3349 is indicated. Boxes represent exons, lines represent introns. In the genome browser view, black arrows indicate m^5C position C3349 (chr5:19,262,126). Top: m^5C methylation percentage (proportion of blue in column represents methylation percentage or numbers of non-converted cytosines) and bottom: a subset of RBS-sequencing reads mapped to *MAG5* locus (chr5:19,262,122-19,262,130) from the separate tissue data sets, siliques, shoots and roots. Each row represents one sequence read and each column a nucleotide. Grey boxes represent nucleotides matching with the *MAG5* reference sequence, red boxes indicate mismatching nucleotides and/or non-methylated, bisulfite converted cytosines. Sequencing gaps are shown in white. (**c**) Distribution of m^5C sites within different RNA types from siliques, shoots and roots. (**d**) Increased methylation correlates with lower mRNA abundance in *Arabidopsis* roots (r_s = -0.32, *P-value ≤0.0001, Spearman's correlation). (**e**) Histogram showing relative enrichment of observed m^5C sites versus expected number of m^5C sites in 3'UTR compared to CDS and 5'UTR across all three tissue types (*P-value ≤0.0001, binomial test). (**f**) Metagene profile of m^5C site abundance along 5'UTR, CDS and 3'UTRs, normalized for transcript length (relative distance) across all three tissue types (FDR ≤0.3 and ≥2% methylation). (**g**) Methylation of candidate m^5C sites using RBS-amplicon-seq were analyzed to independently confirm RBS-seq results. Heat map depicts log arcsine transformed methylation percentages of five m^5C sites in *Arabidopsis* wild type siliques, shoots and roots using RBS-seq and RBS-amp-seq.

sites within 3'UTR's (p-value ≤0.0001), when normalized for read coverage and sequence length (Fig. 1e). When examining the abundance of m^5C sites that occur along the features of mRNAs, the greatest number of m^5C sites were identified within the coding sequence (Fig. 1f). While m^5C sites were evenly distributed within 5'UTRs and coding sequences, m^5C sites were most abundant in the first quarter of 3'UTRs.

Validation of m^5C candidate sites by RBS amplicon sequencing (RBS-amp-seq)

To validate our global analysis of m^5C sites in silique, shoot and root transcriptomes and to test the range of sensitivity of our RBS-seq approach, we independently tested five candidate m^5C sites using RNA Bisulfite amplicon sequencing (RBS-amp-seq). RBS-amp-seq involves next-generation sequencing of targeted PCR amplicons derived from bisulfite treated total RNA (see Methods). All five selected m^5C sites had similar, reproducible methylation levels in both the transcriptome-wide and targeted amplicon sequencing experiments ranging from approximately 3% to 55% methylation (Fig. 1g, Supplementary Table 4). Methylation percentages as low as 3.2±0.6% for *ATMS1* (AT5G17920) mRNA obtained using RBS-seq in wild type seedling shoots were able to be reproduced and independently validated using RBS-amp-seq (3.3±0.8%) (Fig. 1g, Supplementary Table 4). Similarly, our targeted amplicon sequencing validated sites that showed differential methylation levels amongst the tissues using RBS-seq, such as *MAG5* C3349 which shows higher methylation levels in siliques than in seedling shoot and roots (Supplementary Table 4). Together this data demonstrated that a majority of our novel m^5C sites identified in our transcriptome-wide RBS-seq data are independently verifiable.

TRM4B is required for methylation of m^5C sites in shoot, silique and root transcriptomes

To gain insight into the enzymes mediating m^5C methylation, we performed transcriptome-wide RBS-seq on rRNA depleted RNA from *trm4b* mutant silique seedling shoot and root tissues and from *trdmt1* mutant seedling shoots (Fig. 2a, Supplementary Table 3). In siliques and seedling root tissues, 40 and 17 sites had no or significantly reduced methylation in *trm4b* mutants when compared to wild type, respectively (FDR≤0.3). Similarly for seedling shoots of *trm4b*, 69 sites had completely lost or significantly reduced methylation when compared to wild type, whereas no m^5C sites were differentially methylated in *trdmt1* mutants (FDR≤0.3)

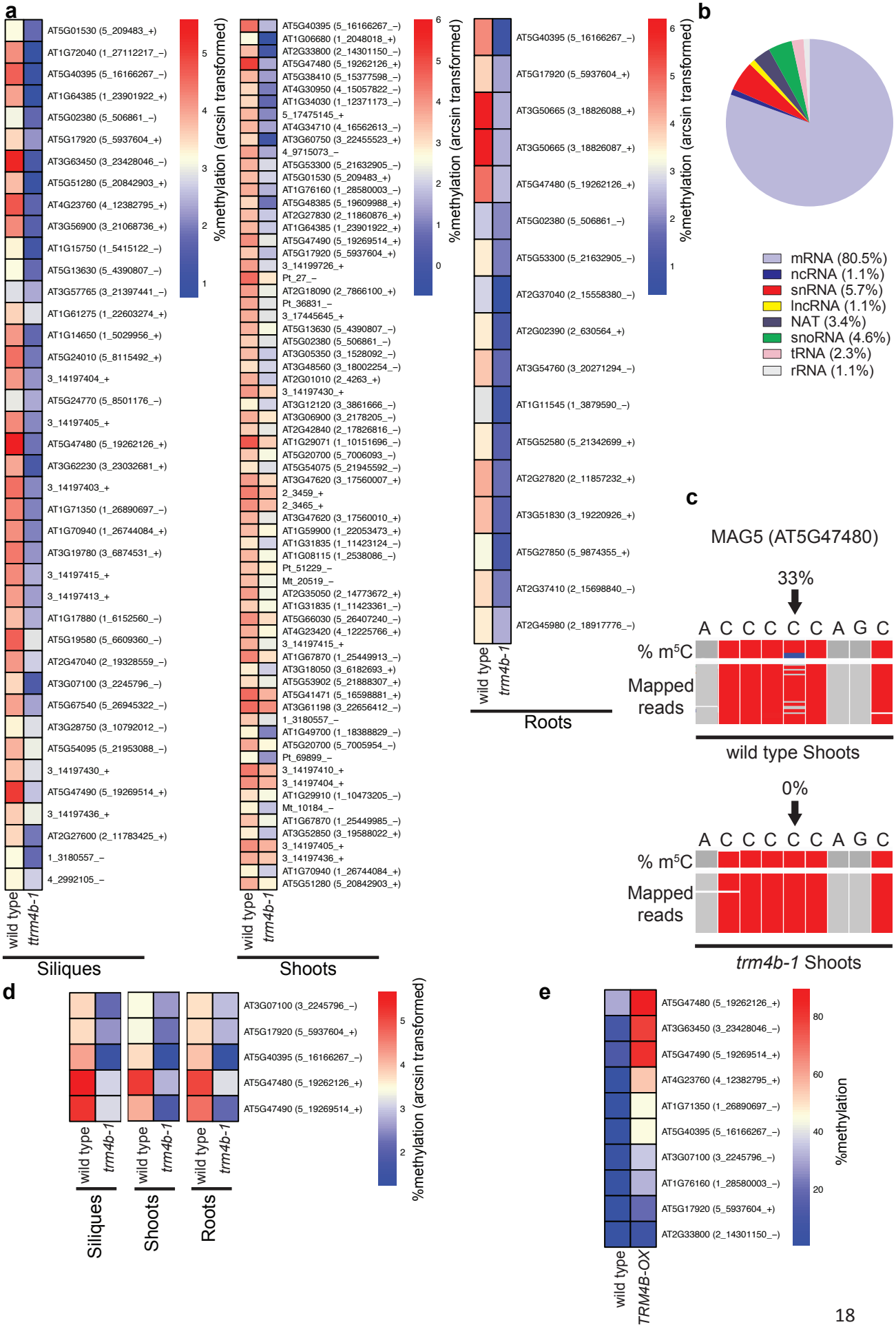

a
AT5G01530 (5_209483_+)
AT1G72040 (1_27112217_–)
AT5G40395 (5_16166267_–)
AT1G64385 (1_23901922_+)
AT5G02380 (5_506861_–)
AT5G17920 (5_5937604_+)
AT3G63450 (3_23428046_–)
AT5G51280 (5_20842903_+)
AT4G23760 (4_12382795_+)
AT3G56900 (3_21068736_+)
AT1G15750 (1_5415122_–)
AT5G13630 (5_4390807_–)
AT3G57765 (3_21397441_–)
AT1G61275 (1_22603274_+)
AT1G14650 (1_5029956_+)
AT5G24010 (5_8115492_+)
3_14197404_+
AT5G24770 (5_8501176_–)
3_14197405_+
AT5G47480 (5_19262126_+)
AT3G62230 (3_23032681_+)
3_14197403_+
AT1G71350 (1_26890697_–)
AT1G70940 (1_26744084_+)
AT3G19780 (3_6874531_+)
3_14197415_+
3_14197413_+
AT1G17880 (1_6152560_–)
AT5G19580 (5_6609360_–)
AT2G47040 (2_19328559_–)
AT3G07100 (3_2245796_–)
AT5G67540 (5_26945322_–)
AT3G28750 (3_10792012_–)
AT5G54095 (5_21953088_–)
3_14197430_+
AT5G47490 (5_19269514_+)
3_14197436_+
AT2G27600 (2_11783425_+)
1_3180557_–
4_2992105_–
wild type
ttrm4b-1
Siliques
%methylation (arcsin transformed)
AT5G40395 (5_16166267_–)
AT1G06680 (1_2048018_+)
AT2G33800 (2_14301150_–)
AT5G47480 (5_19262126_+)
AT5G38410 (5_15377598_–)
AT4G30950 (4_15057822_–)
AT1G34030 (1_12371173_–)
5_17475145_+
AT4G34710 (4_16562613_–)
AT3G60750 (3_22455523_+)
4_9715073_–
AT5G53300 (5_21632905_–)
AT5G01530 (5_209483_+)
AT1G76160 (1_28580003_–)
AT5G48385 (5_19609988_+)
AT2G27830 (2_11860876_+)
AT1G64385 (1_23901922_+)
AT5G47490 (5_19269514_+)
AT5G17920 (5_5937604_+)
3_14199726_+
Pt_27_–
AT2G18090 (2_7866100_+)
Pt_36831_–
3_17445645_+
AT5G13630 (5_4390807_–)
AT5G02380 (5_506861_–)
AT3G05350 (3_1528092_–)
AT3G48560 (3_18002254_–)
AT2G01010 (2_4263_+)
3_14197430_+
AT3G12120 (3_3861666_–)
AT3G06900 (3_2178205_–)
AT2G42840 (2_17826816_–)
AT1G29071 (1_10151696_–)
AT5G20700 (5_7006093_–)
AT5G54075 (5_21945592_–)
AT3G47620 (3_17560007_+)
2_3459_+
2_3465_+
AT3G47620 (3_17560010_+)
AT1G59900 (1_22053473_+)
AT1G31835 (1_11423124_–)
AT1G08115 (1_2538086_–)
Pt_51229_–
Mt_20519_–
AT2G35050 (2_14773672_+)
AT1G31835 (1_11423361_–)
AT5G66030 (5_26407240_–)
AT4G23420 (4_12225766_+)
3_14197415_+
AT1G67870 (1_25449913_–)
AT3G18050 (3_6182693_+)
AT5G53902 (5_21888307_+)
AT5G41471 (5_16598881_+)
AT3G61198 (3_22656412_–)
1_3180557_–
AT1G49700 (1_18388829_–)
AT5G20700 (5_7005954_–)
Pt_69899_–
3_14197410_+
3_14197404_+
AT1G29910 (1_10473205_–)
Mt_10184_–
AT1G67870 (1_25449985_–)
AT3G52850 (3_19588022_+)
3_14197405_+
3_14197436_+
AT1G70940 (1_26744084_+)
AT5G51280 (5_20842903_+)
wild type
trm4b-1
Shoots
%methylation (arcsin transformed)
AT5G40395 (5_16166267_–)
AT5G17920 (5_5937604_+)
AT3G50665 (3_18826088_+)
AT3G50665 (3_18826087_+)
AT5G47480 (5_19262126_+)
AT5G02380 (5_506861_–)
AT5G53300 (5_21632905_–)
AT2G37040 (2_15558380_–)
AT2G02390 (2_630564_+)
AT3G54760 (3_20271294_–)
AT1G11545 (1_3879590_–)
AT5G52580 (5_21342699_+)
AT2G27820 (2_11857232_+)
AT3G51830 (3_19220926_+)
AT5G27850 (5_9874355_+)
AT2G37410 (2_15698840_–)
AT2G45980 (2_18917776_–)
wild type
trm4b-1
Roots
%methylation (arcsin transformed)
b
mRNA (80.5%)
ncRNA (1.1%)
snRNA (5.7%)
lncRNA (1.1%)
NAT (3.4%)
snoRNA (4.6%)
tRNA (2.3%)
rRNA (1.1%)
c
MAG5 (AT5G47480)
33%
A C C C C C A G C
% m5C
Mapped reads
wild type Shoots
0%
A C C C C C A G C
% m5C
Mapped reads
trm4b-1 Shoots
d
AT3G07100 (3_2245796_–)
AT5G17920 (5_5937604_+)
AT5G40395 (5_16166267_–)
AT5G47480 (5_19269514_+)
AT5G47490 (5_19269514_+)
wild type
trm4b-1
Siliques
Shoots
Roots
%methylation (arcsin transformed)
e
AT5G47480 (5_19262126_+)
AT3G63450 (3_23428046_–)
AT5G47490 (5_19269514_+)
AT4G23760 (4_12382795_+)
AT1G71350 (1_26890697_–)
AT5G40395 (5_16166267_–)
AT3G07100 (3_2245796_–)
AT1G76160 (1_28580003_–)
AT5G17920 (5_5937604_+)
AT2G33800 (2_14301150_–)
wild type
TRM4B-OX
%methylation

Figure 2: TRM4B dependent m^{5}C sites in mRNAs and ncRNAs of siliques, shoots and roots in *Arabidopsis.* (**a**) Heat map showing log arcsine transformed methylation percentages of TRM4B dependent m^{5}C sites in *Arabidopsis* wild type (Col-0) and *trm4b-1* siliques at 1 DAP, seedling shoots at 10 DAG and seedling roots at 6 DAG using RBS-seq (n=2-3). Differentially methylated sites are called as significant at (FDR ≤0.3 and ≥1% methylation). (**b**) Distribution of TRM4B dependent m^{5}C sites in different RNA classes across all three tissue types. (**c**) *MAG5* (AT5G47480) requires TRM4B for methylation at cytosine position chr5:19,262,126 (black arrows). Top: wild type and bottom: *trm4b-1* shoots 10 DAG. Integrative Genomics Viewer (IGV) browser snapshots of percentage (%) m^{5}C: (proportion of blue in column represents methylation percentage or numbers of non-converted cytosines) and mapped reads: a subset of the RBS-sequencing reads mapped to *MAG5* locus (chr5:19,262,122-19,262,130). Each row represents one sequence read and each column a nucleotide. Grey boxes represent nucleotides matching with the *MAG5* reference sequence, red boxes indicate mismatching nucleotides and/or non-methylated, bisulfite converted cytosines. Sequencing gaps are shown in white. (**d**) Methylation of candidate m^{5}C target sites of TRM4B in wild type and *trm4b-1* mutants using RBS-amp-seq was analyzed to independently confirm RBS-seq results. Heat map depicts log arcsine transformed methylation percentages of five m^{5}C sites in *Arabidopsis* wild type and *trm4b-1* siliques, shoots and roots using RBS-seq (n=2-3) and RBS-amp-seq (n=3). (**e**) Methylation of TRM4B dependent m^{5}C sites in 3 week old leaves of wild type and transgenic plants over expressing *TRM4B* (*TRM4B-OX)* using RBS-amp-seq. Heat map depicts log arcsine transformed methylation percentages of ten m^{5}C sites, all of which show increased methylation in *TRM4B-OX* (wild type n=3, TRM4B-OX n=1).

(Fig. 2a, Supplementary Tables 3, 5). Furthermore, we used RBS-amp-seq to investigate the requirement for the proposed catalytically inactive TRM4A at 15 m^5C sites and found no change in methylation levels in *trm4a* mutants compared to wild type (Supplementary Fig. 3, Supplementary Table 6). We looked closer at TRM4B dependent sites and discovered that TRM4B mediates many m^5C sites on diverse classes of RNAs such as mRNAs, tRNAs, snoRNAs, snRNAs, lncRNAs and natural antisense transcripts across all three tissue types examined (Fig. 2b). Next we asked the question if sites dependent on TRM4B were located within specific regions of mRNA transcripts. Most TRM4B dependent sites were located in the coding sequence, which is consistent with the global distribution of m^5C sites within mRNAs (Fig. 1f and Supplementary Table 3). One such TRM4B dependent m^5C site, which is methylated in all three tissues types occurs at cytosine position C3349 in the coding sequence of *MAG5* (Fig. 2c). While 33% of *MAG5* transcripts are methylated at C3349 in wild type seedling shoots, methylation is reduced to 0% in *trm4b*.

To independently confirm TRM4B dependent sites, including *MAG5* C3349, identified using our bioinformatics pipeline and statistical cut-off (FDR≤0.3 and ≥1% methylation) we performed RBS-amp-seq on wild type and *trm4b* mutants from siliques, seedling shoots and roots (Fig.2d, Supplementary Table 7). All five m^5C sites tested showed reproducible methylation in wild type and loss or severely reduced methylation in *trm4b* mutants across all three tissue types.

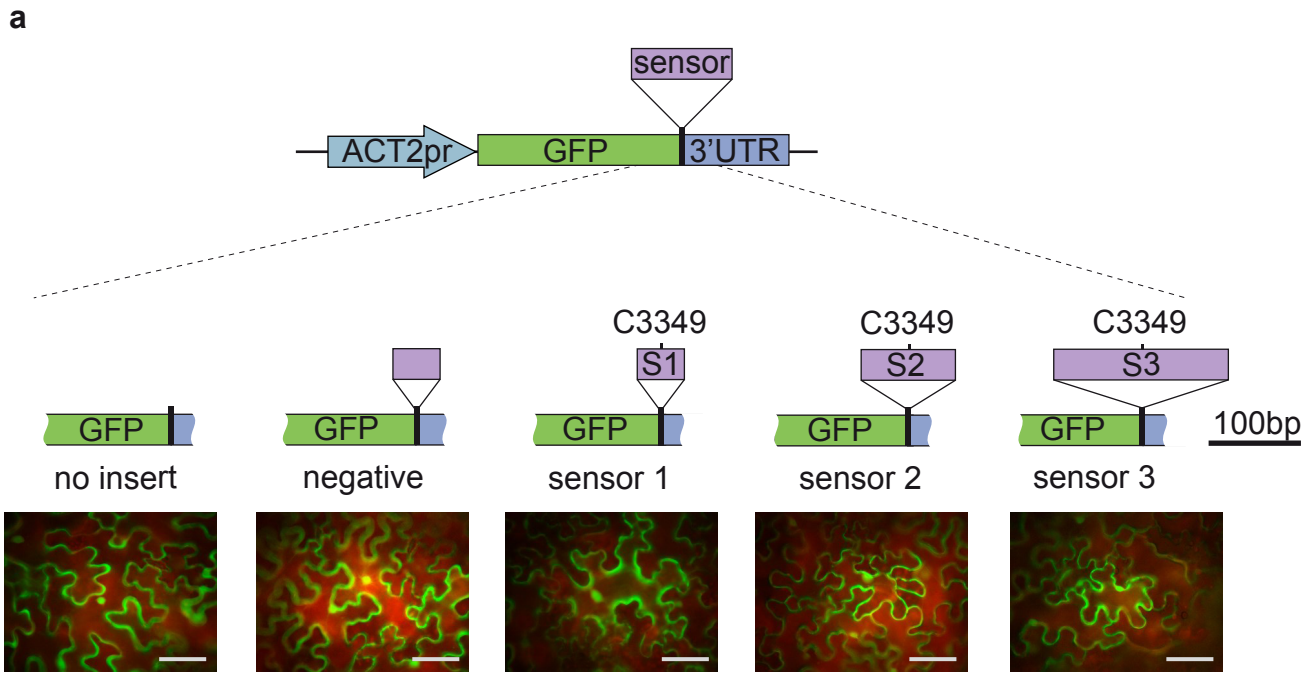
TRM4B over-expression specifically increases m^5C methylation

To investigate if TRM4B expression may contribute to altered methylation rates between different tissues, we compared methylation percentages of TRM4B dependent sites in plants over-expressing *TRM4B* mRNA to wild type (Fig. 2e, Supplementary Table 8). We over-expressed *TRM4B* (Supplementary Fig. 4a) and performed RBS-amp-seq on 10 TRM4B dependent m^5C sites. At all ten sites, methylation levels were increased in TRM4B over-expression plants when compared to wild type plants. We observed between 3 to 124 fold increases in methylation percentages specifically at TRM4B dependent m^5C sites. No additional, spurious cytosines were methylated when *TRM4B* expression was increased, suggesting controlled, specific targeting of TRM4B (Supplementary Fig. 4b).

After identifying and validating TRM4B dependent m^5C sites transcriptome-wide, we investigated how TRM4B is targeted to these specific cytosines in RNA. To test the sequence determinants for targeting of TRM4B, we performed a LOGO motif analysis using a 50 nucleotide (nt) region flanking both sides of TRM4B dependent m^5C sites identified from siliques, seedling shoots and roots (Fig. 2a). Using this approach, we were unable to identify a conserved consensus sequence for targeting TRM4B methylation (Supplementary Fig. 5). This is similar to results obtained from human cancer cell lines[9, 10], and led us to further investigate the sequence requirements for targeting of TRM4B in plants.

TRM4B requires minimum sequence length flanking the m^5C site for effective methylation

As the LOGO motif analysis did not reveal any consensus sequences, we next investigated the sequence requirements for TRM4B to confer methylation in mRNA. We developed a rapid, transient expression assay in *Nicotiana benthamiana* to express m^5C sensor constructs which contained sequences of varying lengths flanking the m^5C site (C3349) in the *MAG5* transcript. The *MAG5* m^5C site was chosen as it was highly methylated across a range of tissues, the methylation percentage varied across tissues and was dependent on TRM4B (Fig. 1b and 2a, d). In addition, a search for TRM4B homologs in the *N. benthamiana* genome revealed two TRM4B-like genes, suggesting that the core components required for m^5C methylation machinery are present (data not shown). Three sensor fragments (51, 101 and 189 nt) containing the *MAG5* C3349 m^5C site and a negative control sensor fragment containing a region of *MAG5* which is not methylated endogenously in *Arabidopsis*, was cloned downstream of a GFP reporter. Transgene expression was confirmed by GFP fluorescence observed in the infiltrated *N. benthamiana* leaves (Fig. 3a) and RNA was purified for RBS-amp-seq. For each construct, three replicate leaves were infiltrated and RBS-amp-seq was performed separately. The results show differential methylation of *MAG5* C3349 site for each of the four sensor constructs tested (Fig. 3b). As expected, efficient BS conversion was observed for all cytosines flanking the methylated site as no methylation was detected using a 0.9% threshold for non-conversion and no methylation was detected in the negative control sensor construct (Fig. 3b). Sensor 2 (S2), which contains a 101 nt *MAG5*


a
sensor
ACT2pr
GFP
3'UTR
C3349
S1
C3349
S2
C3349
S3
GFP
GFP
GFP
GFP
GFP
100bp
no insert
negative
sensor 1
sensor 2
sensor 3

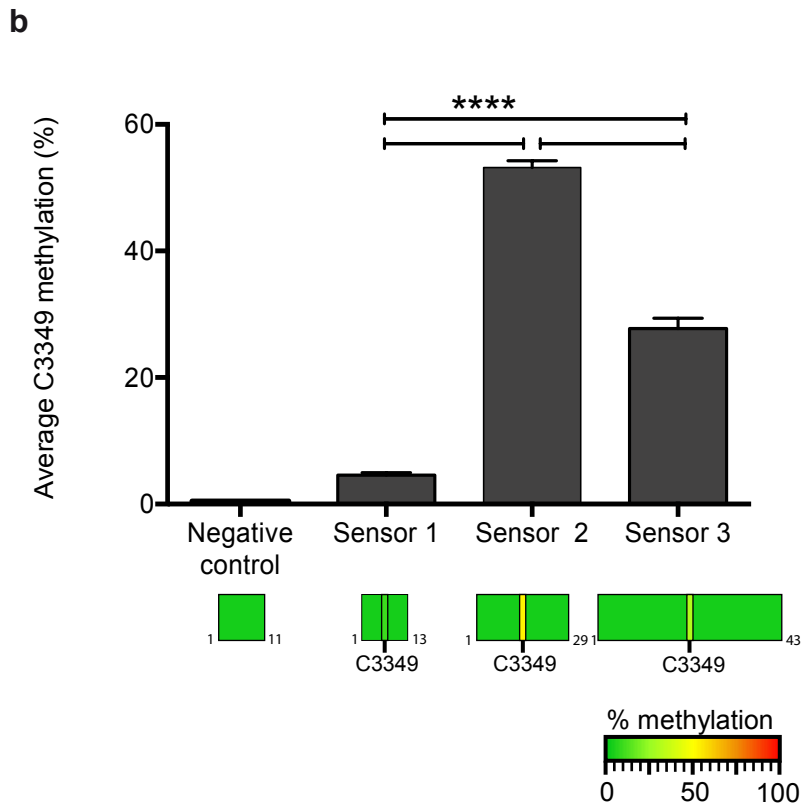

b
****
Average C3349 methylation (%)
60
40
20
0
Negative control
Sensor 1
Sensor 2
Sensor 3
1
11
1
13
C3349
1
29
1
43
C3349
C3349
% methylation
0
50
100

Figure 3: *In vivo* TRM4B dependent substrate recognition. (**a**) Schematic representation of m^5C methylation sensor constructs. Three different sized sequences, flanking the *MAG5* C3349 methylated site were cloned in between the GFP coding region and the beginning of the 3'UTR sequence and are referred to as sensor 1 (S1, 51 nt), sensor 2 (S2, 101 nt) and sensor 3 (S3, 189 nt). The C3349 methylated site is located in the center of all three sensor fragments. A fourth region, derived from *MAG5* exon one was used as 'negative' control in addition to a 'no insert' empty vector control. Representative images of *N. benthamiana* leaves transiently expressing GFP reporter for each of the sensor constructs are shown. Scale bar, 100 μm. (**b**) Cytosine methylation in sensor fragments identified by RBS amplicon seq. (above) Average *MAG5* C3349 methylation percentage in sensor 1, 2 and 3. No non-converted cytosines were identified in the negative control construct. RNA was extracted from three replicate infiltrated leaves of *N. benthamiana* and BS treated, then cDNA synthesis performed and sensor regions were PCR amplified and Illumina sequenced. Error bars indicate s.e (*P<0.001, one way ANOVA; n=3). (below) Heat maps showing the cytosine methylation status within sensor fragments. No no-converted cytosines were identified in the flanking sequences.

insert, had the highest methylation of 53.2±1% at C3349. In contrast, the smallest sensor (S1), with 51 nt, was the least methylated at 4.6±0.4%. This indicates that shorter sequences of up to 25 nt flanking the m^5C site may not carry all the substrate information required for effective methylation by TRM4B. Unexpectedly, sensor 3 (S3) with the longest *MAG5* fragment sequence of 189 nt had only half the methylation level, (27.7±2%) compared to sensor 2. While the level of methylation of *N. benthamiana* MAG5 is not known, sensor 2 cytosine methylation of 53% is similar to that observed in *Arabidopsis* silique tissue (55%). Together these results demonstrate that longer sequences of 50 nt and 94 nt flanking the methylated C3349 site are able to confer moderate methylation, while 25 nt flanking sequence severely decreased methylation, suggesting that additional factors such as RNA structure may be involved.

TRM4B is required for cell proliferation in the *Arabidopsis* root apical meristem

In yeast and animals, TRM4/NSUN2 has been shown to have broad functional roles in mediating stem cell self-renewal and oxidative stress tolerance[16, 20, 32, 33, 38]. Loss of NSUN2 in mice results in male infertility and reduced size, which is thought to be due to a reduced proliferative capacity of stem cells[16, 33]. In plants, stem cell niches are located in the shoot and root apical meristems and are responsible for postembryonic growth of shoot and root systems[42, 43]. While shoot growth during early seedling development appeared unaffected in *Arabidopsis trm4b* mutants[8], primary roots were observed to be significantly shorter compared to wild type in both *trm4b* T-DNA mutant alleles analyzed (Fig. 4a). Monitoring of primary root lengths over a seven-day period revealed decreased elongation rates in the *trm4b* mutant, indicating TRM4B positively regulates root growth at early stages of seedling growth (Fig. 4d). Furthermore, complementation of the *trm4b-1* mutant with a *TRM4B* genomic DNA construct confirmed that the loss of TRM4B causes the observed short-root phenotype (Supplementary Fig. 6). Root growth of *trdmt1*, the second RMTase mutant analyzed in this study, was indistinguishable from wild type plants. Meanwhile, *trm4b-1 trdmt1* double mutants showed short roots similar to those of *trmb-1* single mutants, indicating the phenotype is solely linked to TRM4B loss of function (Supplementary Fig. 7).

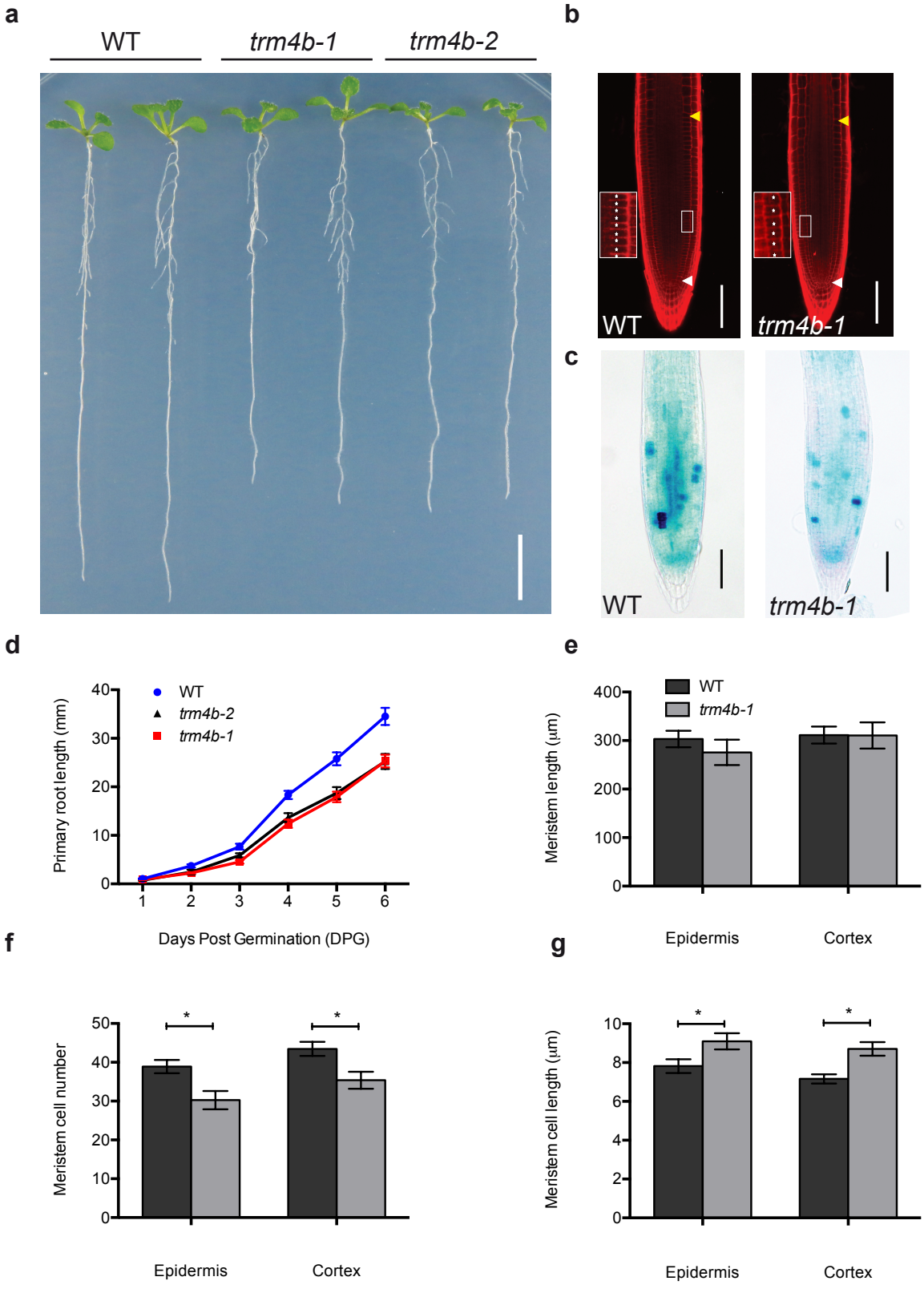

a
WT
trm4b-1
trm4b-2
b
WT
trm4b-1
c
WT
trm4b-1
d
WT
trm4b-2
trm4b-1
Primary root length (mm)
Days Post Germination (DPG)
e
WT
trm4b-1
Meristem length (µm)
Epidermis
Cortex
f
Meristem cell number
Epidermis
Cortex
g
Meristem cell length (µm)
Epidermis
Cortex

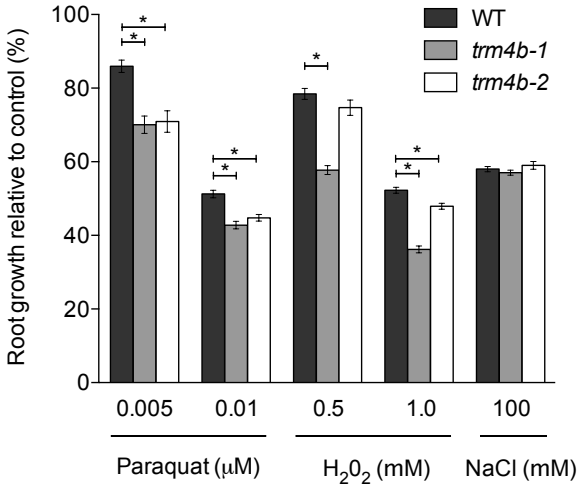
Figure 4: TRM4B is required for primary root growth elongation in young seedlings. (**a**) Root elongation of 7 DAG wild type (Col-0) and two TRM4B T-DNA mutants, *trm4b-1* and *trm4b-2*, grown on half-strength MS medium with 1% sucrose. Scale bar, 1 cm. (**b**) Confocal images of 7 DAG wild type and *trm4b-1* root tips stained with propidium iodide. Scale bar, 100 μm. White arrowheads indicate the position of the QC and yellow arrowheads indicate the first rapidly elongated cell in the cortical layer marking the end of the meristem. Insets (50 μm length) are enlarged images of the boxed areas showing cortical cells in the meristematic zone marked with asterisks. (**c**) Activity of CYCB1;1:GUS reporter in the root meristems of wild type and *trm4b-1* plants, detected by GUS staining. Scale bar, 100 μm. (**d**) Primary root growth measurements over time of wild type and *trm4b-1* and *trm4b-2* mutants. (**e**) Root apical meristem size, (**f**) cell number and (**g**) cell length of wild type and *trm4b-1* mutant plants. Error bars in (d-h) indicate the s.e. (*P<0.05, Students t-test; n≥8 seedlings)

At the cellular level, primary root growth is determined by cell production rates in the meristematic zone and the final cell lengths achieved in the differentiation zone[43]. To further understand the role of TRM4B in coordinating root growth, we examined cellular anatomy in the different root growth zones in *trm4b* mutants and wild type plants. The cellular organization within the root apical meristem (RAM) and overall size was unaffected in the *trm4b-1* mutant (Fig. 4b). The RAM size, as defined as the length from the quiescent center to the first significantly elongated cell, was measured in the epidermal (outer) and cortical (inner) cell files. Interestingly, quantification of meristematic cell numbers revealed *trm4b-1* to have significantly fewer epidermal and cortical cells, reduced by 21% and 17% respectively, compared to wild type. The reduced cell numbers and an unchanged meristem size observed in the *trm4b-1* mutant correlated with an increased average cell length (Fig. 4e-g). However, the final cell lengths achieved in the differentiation zone were similar in the mutant and wild type indicating the larger meristem cells in *trm4b-1* as a possible compensatory mechanism to maintain overall meristem size (Supplementary Fig. 8). Thus, the short-root phenotype observed for *trm4b-1* is most likely linked to the reduced capacity for the cells to divide in the meristem. Supporting evidence was also obtained from CYCB1;1-GUS reporter line crossed into the wild type and *trmb-1* mutant background. CYCB1;1 is expressed during the G2-M phase of cell cycle, allowing visualization of mitotic activity[44]. Based on the intensity of GUS staining in the RAM, a marked reduction of mitotic activity was evident in the *trm4b-1* roots compared to wild type (Fig 4c). Together, this suggests TRM4B positively regulates root growth by controlling cell production numbers in the meristem.

trm4b mutants display increased sensitivity to oxidative stress

Loss of TRM4/NSUN2 in yeast and mice has been shown to result in heightened sensitivity to oxidative stress[20, 38]. We next asked if TRM4B in *Arabidopsis* is also required for oxidative stress response by challenging the mutants with paraquat and H_2O_2 in a root growth sensitivity assay. The *trm4b* mutant displayed increased sensitivity to oxidative stress damage compared to wild type seedlings across a range of paraquat and H_2O_2 concentrations (Fig. 5a). In contrast *trm4b* mutant plants showed a similar level of sensitivity as wild type when grown on salt (Fig. 5a),

a


Root growth relative to control (%)
100
80
60
40
20
0
WT
trm4b-1
trm4b-2
*
0.005
0.01
0.5
1.0
100
Paraquat (μM)
H202 (mM)
NaCl (mM)

b

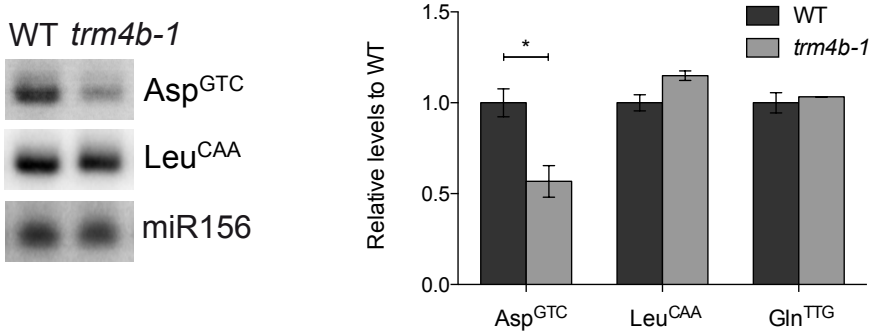

WT trm4b-1
AspGTC
LeuCAA
miR156
Relative levels to WT
1.5
1.0
0.5
0.0
WT
trm4b-1
*
AspGTC
LeuCAA
GlnTTG

Figure 5: Loss of TRM4B leads to oxidative stress sensitivity and decreased tRNA stability. (**a**) Relative root lengths of seedlings on paraquat, H_2O_2 and NaCl compared to untreated controls. Seedlings of wild type, *trmb4-1* and *trmb-2* were grown on half-strength MS media for four days before being transferred to control media or media containing paraquat, H_2O_2 or NaCl at indicated concentrations. Root growth was measured for three subsequent days after transfer. Error bars represent s.e. (*P<0.05, one-way ANOVA comparison to wild type; n≥15-30 seedlings). (**b**) Left: representative northern blot analysis of $tRNAs^{Asp(GTC)}$ and $tRNA^{Leu(CAA)}$ for wild type and *trm4b-1*. The loading control miR156 was used for the analysis. Normalized intensities are given beneath each lane. Right: quantitative analysis of TRM4B substrate $tRNA^{Asp(GTC)}$ and controls; $tRNA^{Leu(CAA)}$ and $tRNA^{Gln(TTG)}$ from two to three independent northern blots, with two replicate RNA samples run for each blot. The tRNA signal intensities were normalized to miR156 loading control and then compared to wild type. Error bars represent s.d. (*P<0.05, Students t-test; n=3-4 replicate RNA samples).

indicating TRM4B is unlikely to be part of the general stress pathway and specifically acts in mediating oxidative responses in *Arabidopsis*.

In humans and mouse, methylation by TRM4/NSUN2 has been demonstrated to protect tRNAs from oxidative stress-induced cleavage[38] and in yeast, TRM4-dependent m^5C sites in tRNAs are dynamically modulated in response to oxidative stress[19, 20]. As we have recently identified TRM4B to methylate a number of nuclear tRNAs in *Arabidopsis*[8], we investigated the steady-state levels of $tRNA^{Asp(GTC)}$ as a representative TRM4B substrate. Three m^5C sites (C48, C49 and C50) in the variable region of $tRNA^{Asp(GTC)}$ were identified as being TRM4B dependent[8]. Northern blotting analysis revealed a 50% reduction of $tRNA^{Asp(GTC)}$ abundance in the *trm4b-1* mutant compared to wild type indicating that TRM4B mediated methylation is required to stabilize $tRNA^{Asp(GTC)}$ substrate (Fig. 5b). In contrast, both $tRNA^{Leu(CAA)}$ and $tRNA^{Gln(TTG)}$, which are not dependent on TRM4B methylation, display similar abundance levels in mutant and wild type (Fig. 5b). We also observed a modest reduction in steady-state levels for the $tRNA^{His(GTG)}$, which contains one TRM4B dependent m^5C site, indicating additional m^5C sites like that observed for $tRNA^{Asp(GTC)}$ may increase tRNA stability or resistance to cleavage (Supplementary Fig. 9). These findings together with earlier reports of TRM4 and TRDMT1 function in other species suggest reduced tRNA stability as a possible cause for *trm4b* plants exhibiting hypersensitivity to oxidative stress[16, 17, 38].

Genes in the oxidative response pathway are constitutively activated in *trm4b*

To further investigate the molecular basis of *trm4b* short-root phenotype and hypersensitivity to oxidative stress, we performed RNA sequencing of the mutant and wild type in paraquat stressed and non-stressed control conditions from whole root tissue. As expected, a large number of genes in the oxidative response pathway were differentially expressed between wild type control and wild type paraquat treated seedlings (Supplementary Fig. 10a). Gene ontology (GO) term analysis of differentially expressed genes showed enrichment of oxidative stress related biological processes (P values ≤0.001). Surprisingly, we also found oxidative

response related genes to be differentially expressed in the *trm4b* mutant compared to wild type under non-stressed conditions (P values ≤0.001) (Supplementary Fig. 10b). Consistent with this observation, GO terms related to oxidative responses were also enriched in the subset of differentially expressed genes common to both paraquat treated wild type compared to untreated wild type and in *trm4b* mutant compared to wild type (Fig. 6a). Closer examination of this common subset of differentially expressed genes, revealed 77% of up-regulated and 78% of down-regulated genes in the *trmb4* mutant compared to wild type were also differentially expressed in the paraquat treated wild type plants compared to control condition (Fig. 6b). Together, these results support the role of TR4MB in regulating oxidative stress responses and suggest that the *trm4b* short root phenotype may be mediated by partial activation of oxidative stress pathways.

Discussion

The discovery of m^5C in cellular RNAs of a wide range of eukaryotic and prokaryotic organisms underscores its importance as a key regulator in RNA metabolism[4, 8, 9, 10, 11]. While previous studies have shown m^5C to be required for tRNA stability[16, 17, 38, 41] and rRNA processing[22], its function in other RNA classes such as mRNAs and the enzymes mediating m^5C methylation are currently being investigated. As an important step in determining biological function, here we describe m^5C sites in diverse classes of coding and non-coding RNAs in *Arabidopsis* shoots, roots and siliques using RBS-seq. To confirm accuracy and to test the range of sensitivity of our transcriptome-wide RBS-seq approach, we independently tested candidate m^5C sites and previously known methylated sites using RBS-amp-seq as well as conventional Sanger sequencing. We found that methylation rates of m^5C sites were in close agreement with our RBS-seq data and were highly reproducible across a wide methylation range (Fig. 1g, 2d and Supplementary Tables 4, 7). Together, these data demonstrate robust and sensitive detection of m^5C sites using our approach. As several other modifications other than m^5C may inhibit conversion of C to T during bisulfite treatment[9, 45], we confirmed that many of these modified sites are indeed m^5C sites, as they are dependent on the RNA methyltransferase TRM4B.

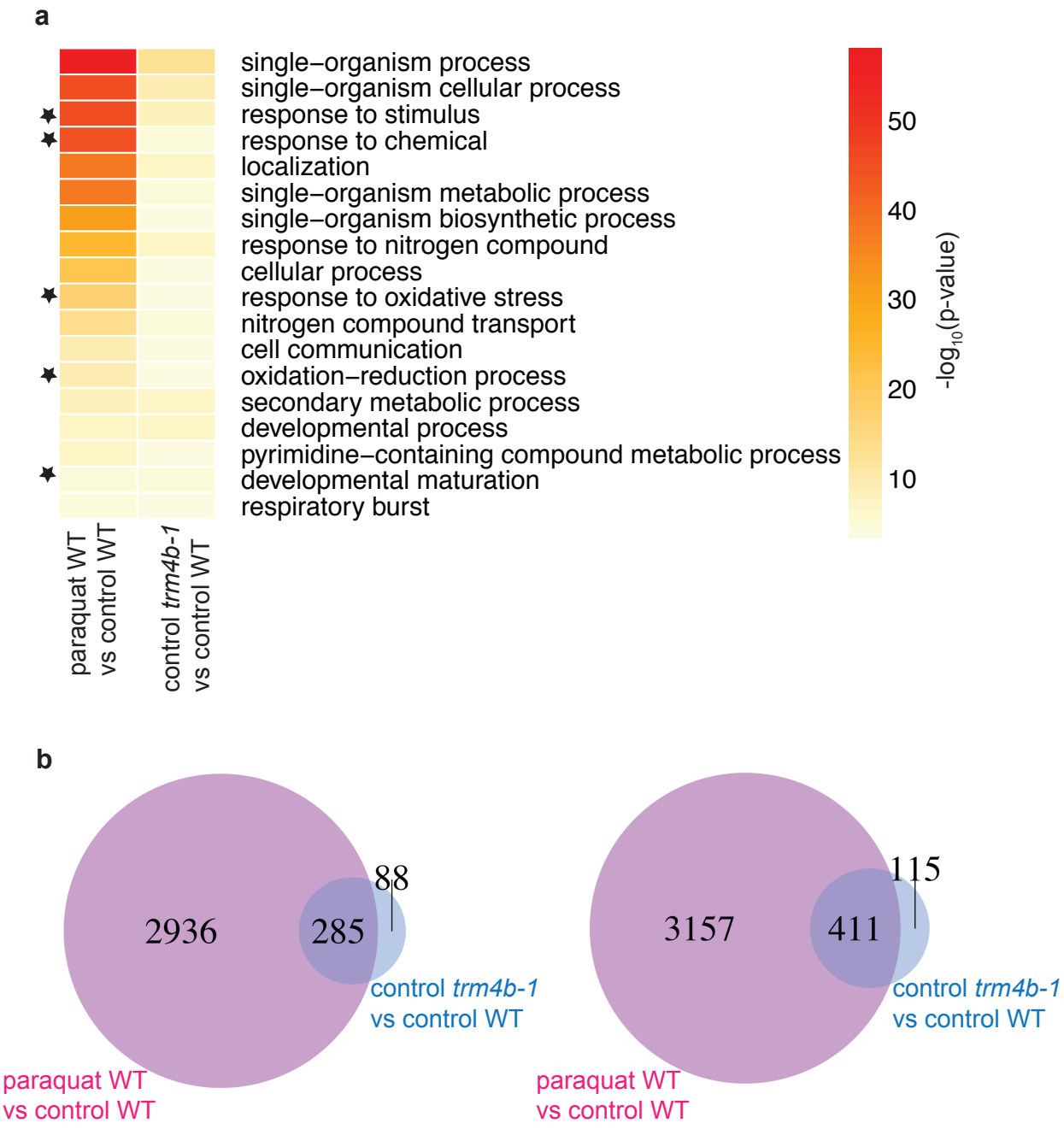

a
single−organism process
single−organism cellular process
response to stimulus
response to chemical
localization
single−organism metabolic process
single−organism biosynthetic process
response to nitrogen compound
cellular process
response to oxidative stress
nitrogen compound transport
cell communication
oxidation−reduction process
secondary metabolic process
developmental process
pyrimidine−containing compound metabolic process
developmental maturation
respiratory burst
paraquat WT vs control WT
control trm4b-1 vs control WT
50
40
30
20
10
-log10(p-value)
b
2936
285
88
paraquat WT vs control WT
control trm4b-1 vs control WT
3157
411
115
paraquat WT vs control WT
control trm4b-1 vs control WT

Figure 6. Oxidative stress responsive pathways are constitutively activated in *trm4b* mutants. Whole roots at 6 DAG were harvested from wild type (WT), *trm4b-1* and paraquat treated wild type seedlings and RNA-seq analysis was performed (n=3). Most genes differentially expressed in *trm4b-1* mutants compared to wild type, are also differentially expressed in paraquat treated wild type plants compared to control conditions. (**a**) Biological process GO terms involved in the oxidative stress response are significantly enriched (FDR<0.05) in *trm4b-1* mutants compared to wild type and in paraquat treated wild type plants compared to untreated wild type plants. GO terms significantly enriched in both comparisons are depicted with black stars indicating GO terms relating to oxidative stress responses. Heat map shows the significance level using the negative log of the p-value, where red = very significant and yellow = significant (GO term FDR <0.05). Analyses were performed using the GOrilla package[46] and REVIGO[47] to create non-redundant GO term sets. (**b**) Venn diagrams show the overlap of differentially up-regulated (Left) and down-regulated (right) genes (FDR <0.05).

We detected 201, 859 and 128 m^5C sites in *Arabidopsis* shoots, roots and siliques, respectively (FDR≤0.3, ≥1% methylation) (Fig. 1a, Supplementary Tables 2, 3). Only a small proportion of these sites were commonly methylated across all three tissue types, suggesting possible tissue-specific functions. Methylated sites were identified predominately in mRNAs and were also found in diverse non-coding RNA types such as snRNA, snoRNAs, lncRNAs and natural antisense transcripts (Fig 1c). The presence of m^5C in many different RNA types is a conserved feature of RNA methylation across mammals and plants[9, 10, 11]. Within *Arabidopsis* mRNAs, m^5C sites most commonly occur within the CDS and fewer numbers of sites were observed in UTRs (Fig. 1f). When normalized for sequence length and read coverage, this revealed enrichment of m^5C sites within *Arabidopsis* 3'UTRs (Fig. 1e), similar to results obtained from human cancer cell lines[9]. To probe for functions of m^5C in mRNAs, we asked if methylation levels affected mRNA abundance in *Arabidopsis* roots. We observed a small negative correlation of mRNA abundance with higher levels of methylation, suggesting that in contrast to tRNAs[16] and other ncRNAs[10], m^5C in mRNA does not generally increase RNA stability, and may in some cases reduce RNA stability through undetermined mechanisms (Fig. 1d). In addition, no correlations were observed for mRNA abundance and m^5C levels in mammals, as no global changes in mRNA abundance were observed in mouse or human *nsun2* mutants[10, 16, 33], suggesting that m^5C may play alternate roles in specific mRNAs and may affect translatability of transcripts.

The RBS-seq approach provides both single base pair resolution and quantitative measurements of methylation percentages of candidate m^5C sites. We used this information to compare m^5C methylation in three tissue types in *Arabidopsis*; seedling shoots, roots and siliques. We found all three tissue types displayed a distinct methylation profile indicating that m^5C deposition is dynamically controlled and may contribute to tissue specific functions and regulation (Fig. 1a, 1b, Supplementary Fig. 2, Supplementary Tables 2, 3). In our previous study, we observed m^5C levels at specific cytosines in *Arabidopsis* tRNAs show little or no variation in methylation levels between floral and seedling shoot tissues[8]. In comparison, greater variation in methylation levels of specific m^5C sites in mRNAs between different tissue types was observed (Fig. 1b, Supplementary Fig. 2,

Supplementary Tables 2, 3). These differences agree with the proposed function for m^5C in tRNAs to maintain stability across all tissues equally, while mRNAs may be differentially methylated in a tissue specific manner to perform tissue specific functions.

We previously demonstrated TRM4B methylated 39 nuclear derived tRNA isodecoders in *Arabidopsis*[8]. Here, we demonstrate an expanded role for TRM4B to mediate methylation of over 100 m^5C sites in mRNAs and snoRNAs across three different plant organs. In the *trm4b* mutant, methylation at these m^5C sites was completely lost or reduced to background levels, whereas over-expression of TRM4B resulted in increased methylation of these sites, in some cases over 124 fold, compared to wild type levels (Fig. 2e). In comparison, no TRM4A or TRDMT1 dependent sites were identified in our dataset, suggesting that the remaining TRM4B independent m^5C sites are mediated by another RNA methyltransferase, such as *Arabidopsis* RCMT9. Together, this data indicates that TRM4B plays a major role in mediating transcriptome-wide methylation in *Arabidopsis*. A similar role was demonstrated for NSUN2 in humans which was shown to have broad substrate range including mRNAs, tRNAs and other non-coding RNAs such as vault ncRNAs, suggesting conserved functions and targeting mechanisms with the *Arabidopsis* TRM4B [9, 10, 11]. Furthermore, as with NSUN2, no consensus target sequence was identified in our RBS-seq dataset for TRM4B dependent m^5C sites. Human NSUN2 was found to have a bias towards C/G flanking nucleotides or to be in a CpG context [9, 10]. However, no further consensus target sequence was identified for human NSUN2.

To determine the sequence requirements for targeting TRM4B, we performed a motif analysis to identify potential consensus sequence for RNA methylation. Similar to human NSUN2, no consensus target sequence was identified in our RBS-seq dataset for TRM4B dependent m^5C sites (Supplementary Fig. 5). To gain further insights into the substrate specificity, we used a transgene-based approach to test the requirements for TRM4B m^5C site recognition in *N. benthamiana* (Fig. 3). Using the internal *MAG5* mRNA m^5C site as a candidate sensor, we found TRM4B requires longer sequences of 50 or 100 nucleotides flanking the m^5C site for effective methylation ((27.7±2%-53.2±1%), while shorter flanking sequences of 25 nucleotides

reduces methylation efficiency (4.6±0.4%.). It remains to be studied whether properties relating to sequence or structure are required for conferring cytosine methylation. Future experiments involving site-directed mutagenesis of nucleotides within the sensor fragments will provide clues to the nature of TRM4B targeting mechanism. Intriguingly, we could not identify a *MAG5* orthologue or sequences that shared significant similarity surrounding the MAG5 m^5C site in the *N. benthamiana* genome. This raises the question of how the endogenous *N. benthamiana* TRM4B is able to effectively methylate m^5C sites in sensor 2 and sensor 3 in our transient expression system and suggests local RNA structure may influence TRM4B specificity. The importance of RNA structure in targeting of m^5C sites by RNA methyltransferases is clearly illustrated by TRDMT1[27, 29, 48].

Importantly, our results demonstrate a novel link between the components required for m^5C methylation and cell proliferation capacity in plants. We show that the *trm4b* mutant exhibits a short-root phenotype and display fewer meristematic cells in the primary RAM (Fig. 4). The cells in the RAM are characterized by their high proliferative capacity, suggesting TRM4B controls the extent of cell cycle progression in the root meristem ultimately affecting root growth. Consistent with this hypothesis, we observed reduced expression of the cell cycle marker CYCB1;1-GUS[44] in the *trm4b* mutant (Fig. 4c). Indeed, NSUN2 has also been shown to be cell cycle regulated and varies its intercellular localization depending on cell cycle stage in human cell lines[49]. In addition, loss of NSUN2 in mice disrupts the balance between stem cell self-renewal and differentiation by delaying cell cycle progression[32, 33] further mirroring TRM4B functions in plants. Future experiments that combine *trm4b* with mutants in the cell cycle pathway as well as identification of protein co-factors will help in understanding the role of TRM4B in plant cell division.

Previous studies have demonstrated m^5C methyltransferases are important regulators of oxidative stress responses in several species[17, 19, 20, 24, 38]. Lack of TRDMT1/DNMT2[17] and TRM4[19, 20] -mediated methylation confers hypersensitivity to oxidative stress in *Drosophila* and yeast, respectively. In each of these cases, the loss of tRNA m^5C methylation and reduced tRNA stability has been suggested to contribute to the increased oxidative stress sensitivity. We earlier reported the

requirement of TRM4B to methylate m^5C sites in a number of nuclear derived tRNAs in *Arabidopsis*[8]. Here we demonstrate *trm4b* mutant plants are more sensitive to paraquat and hydrogen peroxide and show reduced stability of $tRNA^{Asp(GTC)}$ targeted by TRM4B (Fig. 5). We propose that the loss of TRM4B-mediatied tRNA methylation reduces tRNA stability, thereby reducing the pool of available tRNAs and possibly sensitising the cell to oxidative stress. Indeed, emerging evidence indicates tRNAs as important mediators of oxidative stress responses in several species[20, 38, 50]. Notably, in NSUN2-deficient mice, loss of tRNA methylation leads to accumulation of 5' tRNA-derived fragments, which is sufficient to activate cellular stress responses[38]. Whether the loss of TRM4B-secific methylation induces similar downstream responses in *Arabidopsis* remains to be investigated. Interestingly however, we observed genes in the oxidative stress pathway to be constitutively activated in the *trm4b* mutant even in the absence of stress, perhaps reflecting a similar tRNA activated stress pathway that may be conserved in *Arabidopsis* (Fig. 6). Given the broad substrate range demonstrated for TRM4B, it is difficult to tease apart the contributions of tRNA and mRNA methylations to the phenotypes observed in this study. Further analysis of the regulatory functions of m^5C in different RNA contexts will help clarify the biological significance of these modifications in plant development and stress responses.

The dynamic regulation of m^5C in *Arabidopsis* and its impact on gene regulation is just beginning to be elucidated. Here, we provide the first description of m^5C distribution in a plant transcriptome and identify crucial links for this modification to cell division and stress pathways. The parallel roles of TRM4B and NSUN2 in m^5C deposition patterns as well phenotypic similarities of the corresponding mutants provide a fascinating insight into the conserved posttranscriptional regulatory mechanism in plants and animals.

Methods

Plant material and root growth experiments

Arabidopsis thaliana (Columbia ecotype) plants were grown in Phoenix Biosystems controlled environment rooms at 21°C under metal halide lights that provided a level of PAR (photosynthetic active radiation) of 110 µmol of photos/m^2/s. For plate

experiments, *A. thaliana* seeds were first surface sterilized in solution containing 1 part of 10% sodium hypochlorite and 9 parts 100% ethanol and plated on ½ MS medium supplemented with 1% sucrose. Plates were sealed with porous tape (3M) and placed in the dark for three days at 4^{o}C for the seeds to imbibe before being transferred to the growth chamber. All plants were grown under long day photoperiod conditions of 16 h light and 8 h darkness. In order to monitor root growth, seedlings were grown on plates oriented vertically with primary root tips marked daily until 8 days post germination (dpg). For oxidative stress experiments, four-day-old seedlings growing vertically on ½ MS media supplemented with 1% sucrose were transferred to control or media supplemented with specified concentrations of paraquat, hydrogen peroxide or salt. Primary root growth was recorded for three subsequent days after transfer. Seedlings were scanned (Epson flat-bed scanner) at 600 pixels inch^{-1} and primary root length measured in millimetres using ImageJ software.

Characterization of the mutant alleles; *trdmt1* (SALK_136635), *trm4a* (SALK_121111), *trm4b-1* (SAIL_318_G04) and *trm4b-2* (SAIL_667_D03) and the derived double mutant *trdmt1 trm4b* are as described previously[8, 27, 51]. Nucleotide sequence data for the following genes are available from The *Arabidopsis* Information Resource (TAIR) database under the following accession numbers: TRDMT1 (At5g25480), TRM4A (At4g40000) and TRM4B (At2g22400).

Plasmid construction and generation of transgenic plants

For TRM4B overexpression construct, the full-length genomic region of TRM4B including the 5'UTR and 3'UTR was amplified using Col-0 genomic DNA template with primers provided in (Supplementary Table 9) and cloned into Gateway entry vector PCR8 TOPO-TA (Invitrogen). The insert was sequenced and then cloned into the destination vector pMDC32, using the Gateway cloning system[52] following the manufacturers protocol (Invitrogen), resulting in the 35S::TRM4B construct. The construct was transformed into *A. thaliana* wild type Col-0 plants by Agrobacterium mediated floral dip method[53]. Transgenic plants were selected on ½ MS media supplemented with 15 μg ml^{-1} Hygromycin B. TRM4B transcript abundance was assessed in at least five independent T_1 plants using qRT-PCR and two lines

showing the highest TRM4B transcript levels were carried through to homozygous T_3 generation for phenotypic analysis. cDNA synthesis for qRT-PCR was performed using an Invitrogen SuperScript III kit as per the manufacturer's recommendations from 2 µg of total RNA and oligo-dT primed cDNA synthesis. q-RT-PCR detection of *TRM4B* and the housekeeping gene *PDF2A* mRNA was performed using a LightCycler480 (Roche) and SYBER green (Roche). Primers are provided (Supplementary Table 9).

For m^5C sensor constructs, synthesized oligos corresponding to the three sensor fragments flanking *MAG5* C3349 were annealed and cloned into pGreen[54] using the restriction enzyme sites ApaI and SpeI. The ApaI site was inserted into the pGreen multiple cloning site using an oligo adapter. Synthesized oligos are provided (Supplementary Table 9). The constructs were transiently expressed in *N. benthamiana* using Agrobacterium infiltrated into leaves as described previously[55].

For Sanger sequencing of $tRNA^{Asp(GTC)}$ clones, $tRNA^{Asp(GTC)}$ was PCR amplified from bisulfite treated RNA and cloned into the pGEM-T Easy Vector system (Promega) and individual clones sequenced to determine bisulfite conversion efficiency.

Propidium iodide staining, GUS staining and imaging

Eight-day-old *Arabidopsis* seedlings were transferred to 10 µM propidium iodide (Sigma-Aldrich) solution and incubated for 2-3 minutes, rinsed and mounted in MilliQ water. Longitudinal optical sections of the different tissues were visualized using Leica confocal microscopy (Leica, SP5 spectral scanning confocal microscope) with the following settings: excitation wave-length 488 nm; emission 550-800 nm; beam splitter: 488/543/633 nm triple dichroic; objective: HC PL APO CS 20.0x0.70 IMM/COR. For GUS staining, six-day-old *Arabidopsis* seedlings were stained for 16 hours at 37^0C with staining solution (50mM $NaPO_4$, pH7; 2mM potassium-ferrocyanide; 2mM potassium-ferricyanide; 0.2% Triton X-100; 2mM X-Gluc). Roots were then cleared with serial ethanol and methanol washes and rehydrated in water. Roots were visualized using bright field microscopy. Images were processed using ImageJ.

Northern Blot analyses

Total RNA was extracted from 10-day-old seedlings, and 20-25 μg was loaded in each lane of a 12.5% polyacrylamide-urea gel, blotted onto a Hybond-N+ membrane, detected using [γ^{32P}]-ATP labelled oligonucleotide probes and scanned using Phosphorimager (Typhoon 3410). Blot signal intensity was quantified using ImageJ. tRNA and loading control miR156 probe sequences are provided (Supplementary Table 9).

RNA isolation and bisulfite conversion of RNA for RBS-seq

Total RNA was isolated from wild type and *trm4b-1* 10-day-old *Arabidopsis* seedling shoots and 6-day-old seedling roots and 1-day post-fertilization siliques and *trdmt1* 10-day-old seedling shoots using the Spectrum Plant total RNA kit (SIGMA-ALDRICH) and contaminating DNA removed using DNase I (SIGMA-ALDRICH). To enrich for mRNA fraction, 5 μg of RNA isolated was subjected to ribosomal RNA depletion using the Ribo-Zero kit (Illumina, MRZPL116) and successful depletion was monitored using Bioanalyzer 2001 (Agilent Technologies). For bisulfite conversion, 150-200 ng of ribosomal depleted RNA was converted with sodium metabisulfite (SIGMA-ALDRICH) as previously described[8, 9, 45]. As a control, 200 pg of *in vitro* transcribed Renilla luciferase (SIGMA ALDRICH) was added to each RNA sample prior to conversion with sodium metabisulfite. Bisulfite converted RNA was used as a template for strand-specific RNA-seq library preparation using the NEB Ultra directional RNA library kit. As bisulfite treated RNA is sheared, the fragmentation step of the library preparation was omitted and samples were quickly processed for first strand cDNA synthesis after addition of the fragmentation buffer. The remaining steps of library construction were performed as per the manufacturers' instructions. Bisulfite treated RNA library samples were sequenced on the Illumina HiSeq2500 (2 x 100 nt paired end) platform at the ACRF, Adelaide. For each tissue type, a minimum of two - three biological replicate library samples were prepared and sequenced.

RNA isolation and library construction for RNA-seq

Total RNA was isolated from wild type and *trm4b-1* 6-day-old *Arabidopsis* seedling roots grown on ½ MS media containing no paraquat, or supplemented with 0.01 μM

paraquat, using the Spectrum Plant total RNA kit (SIGMA-ALDRICH). DNA was removed using DNase I (SIGMA-ALDRICH). Ribosomal RNA was removed using the Ribo-Zero kit (Illumina, MRZPL116) and successful depletion was monitored using Bioanalyzer 2001 (Agilent Technologies). The resulting mRNA enriched RNA was used as a template for strand-specific RNA-seq library preparation using the NEB Ultra directional RNA library kit. RNA-seq library construction was performed as per the manufacturers' instructions. RNA-seq library samples were sequenced on the Illumina HiSeq2000 (2 x 100 nt paired end) platform at the ACRF, Adelaide. Three biological replicates were prepared and sequenced for each condition.

Bioinformatic analysis of RBS-seq and RNA-seq

Global Mapping: Reads were first adapter trimmed using Trimmomatic in palindromic mode, allowing single base precision, with stringent 3' quality filtering (if any 4 base window in the read had a Phred quality score below 20 then the 3' remained of the read was trimmed)[56]. Reads were then globally mapped to *in silico* bisulfite converted *Arabidopsis* reference genomes (TAIR10)[57], based on the method and implementation of B-Solana[58], but adapted to RNA ("B-Solana RNA" in the sequel). In particular, TopHat[59] was used rather than Bowtie[60] to first map junction reads to a refseq transcriptome, to reliably detect known transcript isoforms, and then to the reference genome. To ensure maximal specificity of predicted sites, only uniquely mapped reads were retained ("uniquely-mapped" was defined by a Bowtie2 mapping quality (mapq) score of ≥ 20).

Differential and overall methylation calls: The method was based on the statistical model and method described in Parker, B.J., Sibbritt, T., Wen, J., et al, Preiss, T. "Role of NSUN2 and TRDMT1 m^5C RNA methylation in mRNA and ncRNA post-transcriptional regulation". (In preparation).

Proportion statistic: The primary statistic used for detection of methylation was a proportion statistic p=(C+psi)/(T+C), where psi is added pseudo counts (1/8 counts), which was transformed to an approximate Gaussian distribution by an arc sin of square root transformation[61], followed by a log transformation, as empirically the distribution is weighted toward lower methylation levels, with transformed proportion p' defined by the following equation:

$$p' = \frac{360}{2\pi}.arcsin\sqrt{p}$$

Transformed silique data was additionally normalized between samples by applying cyclic loess normalization[62]

Non-specific filtering: Non-specific filtering (i.e. blinded to experimental/control labels) to initially exclude low expressed sites and clearly non-methylated C bases was performed by requiring ≥ 5 reads across all samples in experimental and control samples. In addition, for the differential methylation models, we required half the samples to show a significantly (p-value < 0.05; binomial test) greater than 5% methylation fraction. For overall methylation calls we required all samples to have > 0% methylation fraction.

Statistical model: A weighted moderated linear model based on transformed proportion data, p', was used for differential methylation (DM) estimation using Limma[63] the moderation utilizes global variance estimates across methylation sites to improve the variance estimates of the individual methylation sites.

The weights corrected for mean-variance trend, estimated by quadratic fit of variance of the transformed data versus total read counts over each site, using the Voom method[62]. Contrasts appropriate to the particular analyses discussed in the results section were used. For differential methylation site calling, balanced contrasts between methyltransferase mutant and wild type controls were used. For overall methylation calls, wild type samples were compared against a fixed value of 0% methylation.

RNA-seq data: Root RNA-seq data was adaptor- trimmed using Trimmomatic[56] and mapped to the *Arabidopsis* TAIR10[57] reference genome using TopHat[59]. Data was normalized between samples using TMM normalization[64]. Differentially expressed genes were detected using Limma[63].

False discovery rate (FDR) was estimated by the Storey method[65].

Post-filtering: Called methylation sites were required to be homogeneous (C + T ≥ 95% of reads) to exclude calls in bases of high read error rate or unannotated SNPs. A minimum mean methylation fraction of 1% across WT samples was required to exclude very low methylation calls that may be enriched for conversion artifacts due to RNA structure, or other systematic biases.

Sequencing and analysis of RNA bisulfite amplicon sequencing (RBS-amp-seq)

Total RNA was isolated from wild type and *trm4b-1* 10-day-old *Arabidopsis* seedling shoots and 6-day-old seedling roots and 1-day post-fertilization siliques and *trm4a* 10-day-old seedling shoots using the Spectrum Plant total RNA kit (SIGMA-ALDRICH) and contaminating DNA removed using DNase I (SIGMA-ALDRICH). For bisulfite conversion, 2-5 µg of total RNA was converted with sodium metabisulfite (SIGMA-ALDRICH) as described above. cDNA synthesis for RBS-amp-seq was performed using an Invitrogen SuperScript III kit as per the manufacturer's instructions using 1-2 µg of bisulfite treated total RNA and bulked gene specific reverse transcription (RT) primers (Supplementary Table 9). Amplicons were PCR amplified manually, or on an Access Array™ (Fluidigm) using the primers provided (Supplementary Table 9) as per the manufacturers' instructions. PCR amplicons derived from the same biological replicate were bulked and indexed as recommended by the manufacturer. Illumina sequencing was performed on a MiSeq platform (2 x 160 nt paired end) at the ACRF, Adelaide. Three biological replicates were prepared and sequenced for each condition.

Analysis of RBS-amp-seq libraries was performed using CLC Genomics Workbench (Qiagen). Briefly, sequences were trimmed for adapters and filtered for low quality reads. RBS-amp-seq reads were then mapped to *in silico* converted reference sequences corresponding to the expected PCR amplicons. In order to identify methylated cytosines, non-conversion of a cytosine in aligned read sequences was taken to indicate the presence of m^5C. Percentage methylation at specific positions was calculated as the number of mapped cytosines divided by the combined total number of mapped cytosines and mapped thymines.

Accession numbers

The data sets supporting the results of this article will be available in NCBI's GEO database repository.

Acknowledgements

We thank Wendy Parker and Joel Geoghegan at the ACRF (Adelaide) for their help and expertise for deep sequencing and optimizing experiments utilizing the Fluidigm Access Array. This research was supported by ARC grants DP110103805 and FT130100525 awarded to I.S. and an APA and a GRDC PhD top-up scholarship awarded to A.B.

Author's contributions

Experiments were designed by R.D., A.B., T. S., T.P. and I.S. Experiments were performed by R.D., A.B. and K.P. Data analysis was performed by B.P., R.D., A.B., K.P. and I.S. The manuscript was equally prepared and edited by R.D., A.B. and I.S. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References:

1. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476 (2008).

2. Korlach J, Turner SW. Going beyond five bases in DNA sequencing. *Curr Opin Struct Biol* **22**, 251-261 (2012).

3. Motorin Y, Lyko F, Helm M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic acids research* **38**, 1415-1430 (2010).

4. Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS genetics* **9**, e1003602 (2013).

5. Jackman JE, Alfonzo JD. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wires Rna* **4**, 35-48 (2013).

6. Machnicka MA*, et al.* MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic acids research* **41**, D262-267 (2013).

7. Hussain S, Aleksic J, Blanco S, Dietmann S, Frye M. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome biology* **14**, 215 (2013).

8. Burgess AL, David R, Searle IR. Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. *BMC plant biology* **15**, (2015).

9. Squires JE*, et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic acids research* **40**, 5023-5033 (2012).

10. Hussain S*, et al.* NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. *Cell reports* **4**, 255-261 (2013).

11. Khoddami V, Cairns BR. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat Biotechnol* **31**, 458-464 (2013).

12. Motorin Y, Helm M. tRNA Stabilization by modified nucleotides. *Biochemistry* **49**, 4934-4944 (2010).

13. Alexandrov A*, et al.* Rapid tRNA decay can result from lack of nonessential modifications. *Mol Cell* **21**, 87-96 (2006).

14. Agris PF. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *Embo Rep* **9**, 629-635 (2008).

15. Helm M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic acids research* **34**, 721-733 (2006).

16. Tuorto F*, et al.* RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nature structural & molecular biology* **19**, 900-905 (2012).

17. Schaefer M*, et al.* RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes & development* **24**, 1590-1595 (2010).

18. Tuorto F*, et al.* The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis. *The EMBO journal* **34**, 2350-2362 (2015).

19. Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ. A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS genetics* **6**, e1001247 (2010).

20. Chan CTY*, et al.* Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nature communications* **3**, (2012).

21. Gigova A, Duggimpudi S, Pollex T, Schaefer M, Kos M. A cluster of methylations in the domain IV of 25S rRNA is required for ribosome stability. *RNA* **20**, 1632-1644 (2014).

22. Hong B, Brockenbrough JS, Wu P, Aris JP. Nop2p is required for pre-rRNA processing and 60S ribosome subunit synthesis in yeast. *Mol Cell Biol* **17**, 378-388 (1997).

23. Sharma S, Yang J, Watzinger P, Kotter P, Entian KD. Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively. *Nucleic acids research* **41**, 9062-9076 (2013).

24. Schosserer M*, et al.* Methylation of ribosomal RNA by NSUN5 is a conserved mechanism modulating organismal lifespan. *Nature communications* **6**, 6158 (2015).

25. Amort T*, et al.* Long non-coding RNAs as targets for cytosine methylation. *RNA biology* **10**, 1003-1008 (2013).

26. Warren L*, et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618-630 (2010).

27. Goll MG*, et al.* Methylation of tRNA(AsP) by the DNA methyltransferase homolog Dnmt2. *Science* **311**, 395-398 (2006).

28. Becker M, Muller S, Nellen W, Jurkowski TP, Jeltsch A, Ehrenhofer-Murray AE. Pmt1, a Dnmt2 homolog in Schizosaccharomyces pombe, mediates tRNA methylation in response to nutrient signaling. *Nucleic acids research* **40**, 11648-11658 (2012).

29. Jurkowski TP*, et al.* Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *Rna-a Publication of the Rna Society* **14**, 1663-1670 (2008).

30. Motorin Y, Grosjean H. Multisite-specific tRNA : m(5)C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: Identification of the gene and substrate specificity of the enzyme. *RNA* **5**, 1105-1118 (1999).

31. Auxilien S, Guerineau V, Szweykowska-Kulinska Z, Golinelli-Pimpaneau B. The Human tRNA m(5)C methyltransferase Misu is multisite-specific. *RNA biology* **9**, 1331-1338 (2012).

32. Blanco S, Kurowski A, Nichols J, Watt FM, Benitah SA, Frye M. The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate. *PLoS genetics* **7**, e1002403 (2011).

33. Hussain S*, et al.* The Mouse Cytosine-5 RNA Methyltransferase NSun2 Is a Component of the Chromatoid Body and Required for Testis Differentiation. *Mol Cell Biol* **33**, 1561-1570 (2013).

34. Khan MA*, et al.* Mutation in NSUN2, which Encodes an RNA Methyltransferase, Causes Autosomal-Recessive Intellectual Disability. *Am J Hum Genet* **90**, 856-863 (2012).

35. Abbasi-Moheb L*, et al.* Mutations in NSUN2 Cause Autosomal-Recessive Intellectual Disability. *Am J Hum Genet* **90**, 847-855 (2012).

36. Martinez FJ*, et al.* Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. *J Med Genet* **49**, 380-385 (2012).

37. Fahiminiya S*, et al.* Whole exome sequencing unravels disease-causing genes in consanguineous families in Qatar. *Clinical genetics* **86**, 134-141 (2014).

38. Blanco S*, et al.* Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *The EMBO journal* **33**, 2020-2039 (2014).

39. Pavlopoulou A, Kossida S. Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. *Genomics* **93**, 350-357 (2009).

40. Chen P, Jager G, Zheng B. Transfer RNA modifications and genes for modifying enzymes in Arabidopsis thaliana. *BMC plant biology* **10**, 201 (2010).

41. Chen Y, Sierzputowskagracz H, Guenther R, Everett K, Agris PF. 5-Methylcytidine Is Required for Cooperative Binding of Mg2+ and a Conformational Transition at the Anticodon Stem-Loop of Yeast Phenylalanine Transfer-RNA. *Biochemistry* **32**, 10249-10253 (1993).

42. Stahl Y, Simon R. Plant primary meristems: shared functions and regulatory mechanisms. *Curr Opin Plant Biol* **13**, 53-58 (2010).

43. Beemster GT, Baskin TI. Analysis of cell division and elongation underlying the developmental acceleration of root growth in Arabidopsis thaliana. *Plant Physiol* **116**, 1515-1526 (1998).

44. Doerner P, Jorgensen JE, You R, Steppuhn J, Lamb C. Control of root growth and development by cyclin expression. *Nature* **380**, 520-523 (1996).

45. Schaefer M, Pollex T, Hanna K, Lyko F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic acids research* **37**, e12 (2009).

46. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics* **10**, 48 (2009).

47. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *Plos One* **6**, e21800 (2011).

48. Shanmugam R*, et al.* The Dnmt2 RNA methyltransferase homolog of Geobacter sulfurreducens specifically methylates tRNA-Glu. *Nucleic acids research* **42**, 6487-6496 (2014).

49. Sakita-Suto S, Kanda A, Suzuki F, Sato S, Takata T, Tatsuka M. Aurora-B regulates RNA methyltransferase NSUN2. *Molecular biology of the cell* **18**, 1107-1117 (2007).

50. Thompson DM, Lu C, Green PJ, Parker R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **14**, 2095-2103 (2008).

51. Alonso JM*, et al.* Genome-wide Insertional mutagenesis of Arabidopsis thaliana. *Science* **301**, 653-657 (2003).

52. Curtis MD, Grossniklaus U. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* **133**, 462-469 (2003).

53. Davis AM, Hall A, Millar AJ, Darrah C, Davis SJ. Protocol: Streamlined sub-protocols for floral-dip transformation and selection of transformants in Arabidopsis thaliana. *Plant Methods* **5**, 3 (2009).

54. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol Biol* **42**, 819-832 (2000).

55. Sparkes IA, Runions J, Kearns A, Hawes C. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc* **1**, 2019-2025 (2006).

56. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

57. Lamesch P*, et al.* The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids research* **40**, D1202-1210 (2012).

58. Kreck B, Marnellos G, Richter J, Krueger F, Siebert R, Franke A. B-SOLANA: an approach for the analysis of two-base encoding bisulfite sequencing data. *Bioinformatics* **28**, 428-429 (2012).

59. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).

60. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25 (2009).

61. Snedecor GW, Cochran WG. *Statistical methods*, 7th edn. Iowa State University Press (1980).

62. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* **15**, R29 (2014).

63. Smyth GK. Limma: Linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (ed^(eds Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W). Springer (2005).

64. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology* **11**, R25 (2010).

65. Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498 (2002).

Chapter 4 Discussion, Conclusions and Future Directions

Modification of the four ribonucleic acid bases allows expansion of the functions of RNA. These modifications can affect RNA structure, stability, processing and inhibit or facilitate binding of RNA binding proteins (RBPs) (reviewed in Chapter 1). The importance of RNA modifications is demonstrated as defects in modifying enzymes or 'writers' are linked to diseases and cancer in humans and various defects in other organisms [1-3]. Furthermore, RNA modifications such as 5-methylcytosine (m^5C) are required for adaptive translational responses to abiotic and biotic stresses [4-6]. While m^5C sites have been identified transcriptome-wide in human, mouse, yeast, bacteria and archaea utilizing different detection methods [7-10], the m^5C landscape has not been previously investigated in plants.

Here, we utilize RNA Bisulfite treatment coupled with deep sequencing (RBS-seq) to investigate m^5C modifications transcriptome-wide in *Arabidopsis*. While there are many detection methods available for m^5C modifications in RNA, RBS-seq was the preferred method, as RBS-seq allows high-throughput, quantitative, site-specific detection of m^5C sites in both abundant and lowly abundant transcripts such as tRNAs and mRNAs, respectively [7, 11]. As with any method, there are pros and cons and like-wise, appropriate controls are required when using RBS-seq to interpret our results. It has been previously reported that incomplete denaturation of highly structured regions of RNA, such as those found in rRNAs, can lead to incomplete bisulfite conversion rates of unmodified cytosines, leading to false methylation positives [7, 12]. Consequently, high conversion rates are necessary when using RBS-seq to minimize the number of false positives. Controls such as *in vitro* transcribed transcripts, analysis of rRNA conversion rates and global conversion rates can be used to determine the background levels of false positives [7, 11, 13]. Bisulfite treatment of RNA converts unmodified cytosine residues to uracil, while methylated cytosines are protected and not converted. Unfortunately, resistance to RNA Bisulfite treatment is not confined to m^5C, and hence, any non-converted sites are candidate m^5C sites. Other modifications that are resistant to conversion include 3-methylcytosine (m^3C), N4-methylcytosine (m^4C), N4, 2'-O-dimethylcytosine (m^4Cm) and other N4-acetylated variants [7, 12]. Furthermore, 5-

hydroxymethylcytosine (hm^5C) is converted to cytosine-5-methylenesulfonate (CMS), instead of uracil after bisulfite treatment, so it is also indistinguishable from non-converted m^5C sites [14]. To determine *bona fide* m^5C sites and overcome ambiguity, we identified candidate m^5C sites which were present in wild type plants and absent or reduced in RNA 5-methylcytosine transferase mutants as true m^5C sites (Chapter 2 and Chapter 3).

Along with RBS-seq, three other high-throughput methods used for identifying m^5C transcriptome-wide across different organisms are 1.) RNA Immunoprecipitation (RIP) [8], 2.) 5-azacytidine Immunoprecipitation (Aza-IP) [9] and 3.) Methylation individual-nucleotide-resolution crosslinking and Immunoprecipitation (miCLIP) [10]. These methods, along with RBS-seq are reviewed in [15, 16]. The first method, RIP uses anti-m^5C antibodies which recognise the m^5C modification on both DNA and RNA. This method can be used to specifically detect regions in RNA containing m^5C, however, this technique is not site-specific. The second high-throughput method, Aza-IP, requires over-expression of an affinity-tagged RMTase in cells or an organism, grown in the presence of the suicide inhibitor 5-azacytidine. 5-azacytidine is a cytosine analogue and can be incorporated into newly synthesized RNA. When an RMTase methylates 5-azacytidine in RNA, the enzyme becomes covalently bound to the RNA. This allows the affinity-tagged RMTase to be immunoprecipitated with an appropriate antibody to enrich for methylated RNA. As there is a high percentage of C-G transversions at the site where the RMTase was covalently bound, this enables site-specific identification of m^5C sites. The third transcriptome-wide method, known as miCLIP requires over-expression of a mutated affinity-tagged RMTase for immunoprecipitation. The RMTase mutation results in irreversible binding to the RNA target. This RMTase binding causes reverse transcription to stall, enabling site-specific detection of m^5C sites.

RBS-seq was chosen over these other high-throughput methods for a number of reasons. RBS-seq requires much less input RNA to perform experiments, is site-specific (unlike RIP) and does not require the use of affinity-tagged over-expression constructs (unlike Aza-IP and miCLIP), which would not have allowed us to examine tissue-specific differences in methylation. Using RBS-seq does however require increased sequencing depth compared to the other three high-throughput methods

to identify lowly methylated and/or expressed transcripts, as there is no enrichment for methylated transcripts (reviewed in [15, 16]).

We required high-throughput methods of detecting m^5C sites, however there are several other, non-high-throughput methods of m^5C detection including Thin Layer Chromatography (TLC), High Pressure Liquid Chromatography (HPLC) and mass spectrometry (MS) (reviewed in [17]). These techniques can identify and quantify multiple RNA modifications simultaneously, and are better suited to observing changes in multiple RNA modifications in total RNA, or in fractions of RNA, such as tRNAs, or all mRNAs. It would be technically difficult and laborious to purify sufficient amounts of specific, low abundant transcripts such as a mRNA species like *MAG5* for analysis using these methods. Therefore, for our experimental goals, using RBS-seq was the best method for high-throughput identification of specific m^5C sites transcriptome-wide in different tissues of *Arabidopsis*. In the future, other methods such as RIP using anti-m^5C antibodies or HPLC may help to confirm and supplement our findings presented here.

In this thesis, we report over a thousand m^5C sites transcriptome-wide in *Arabidopsis thaliana* seedling shoots, roots and siliques (Chapter 3). In *Arabidopsis* seedling shoots, roots and siliques, 201, 859 and 128 m^5C sites were identified with methylation levels ranging from 1% to 92%, respectively. We show that this methylation is dynamic, as many m^5C methylation sites change in methylation amount or are unique to one tissue type, suggesting tissue-specific functions and regulation. Of those sites which are common to multiple tissue types, some of these are differentially methylated between tissue types, while others are constitutively methylated at the same level across all tissue types. While the overall methylation levels at many of these m^5C sites are extremely low, and raises questions about the physiological relevance of such low methylation, it is possible that for some of these lowly methylated sites there may be cell type or organelle specific methylation of RNA, so that in some compartments of the cell, or in specific cell types, methylation at that site may be very high, but the signal may be diluted as we tested whole tissues. We may obtain a much more complex and nuanced picture of the methylation landscape by examining individual cell types and organelle or nuclear

fractions. Furthermore, a proportion of these very lowly methylated sites are expected to be false positives and hence have no physiological relevance as a false discovery rate of 30% was used as the cut off. The additional use of complementary techniques to detect m^5C sites such as RIP using anti-m^5C antibodies will be required to confirm these sites.

In addition to identifying dynamic, tissue-specific m^5C sites, we also discovered methylation sites in most classes of RNAs including mRNAs, long non-coding RNAs, snoRNAs, rRNAs and nuclear tRNAs (Chapter 2 and Chapter 3). Within mRNAs we discovered that m^5C sites are uniformly distributed along coding sequences and slightly enriched at the beginning of 3'UTRs. This mirrors the deposition pattern identified in human HeLa cells, suggesting conservation of functions and targeting mechanisms between humans and plants [7]. Furthermore, we demonstrate conservation of rRNA and tRNA m^5C sites across six species in the kingdom *Plantae*, suggesting important and highly conserved roles of this post-transcriptional modification in non-coding RNAs.

To further investigate the role of m^5C in plant development, we identified and characterized T-DNA insertion lines in 5 predicted m^5C 'writers' and extend the functional characterization of TRDMT1 and NOP2A/OLI2 (Chapter 2). The *Arabidopsis thaliana* genome encodes 5 predicted rRNA RNA methyltransferases (RMTases); NSUN5, NOP2A, NOP2B, NOP2C and FMU/RNMT [18, 19]. Here we show that *Arabidopsis* NSUN5 is required for methylation of C2278 in nuclear large subunit 25S rRNA and that *nsun5* mutants appear wild type under standard laboratory conditions. Furthermore, *nop2a*, *nop2b* and *nop2c* single mutants failed to show changes in methylation of any rRNA sites when compared to wild type controls (Chapter 2). This suggests possible functional redundancies among these three genes. In support of this, *nop2a nop2b* and *nop2a nop2c* double mutants were unable to be identified and segregation ratios suggest that haploid-insufficiency is lethal or occurs very rarely (Chapter 2, data not shown). *nop2b nop2c* double mutants were identified and appeared wild type, suggesting that NOP2B and NOP2C can interchangeably co-operate with NOP2A (data not shown). The lethality of *nop2a*

nop2b and *nop2a nop2c* double mutants may be due to additional functions for these enzymes in rRNA maturation, independent of their methylation activity as discussed previously in Chapter 1.

Four other RMTase genes in the *Arabidopsis thaliana* genome are predicted to methylate tRNAs and potentially other RNAs such as mRNAs and snoRNAs; TRDMT1, RCMT9, TRM4A and TRM4B [18, 19]. TRDMT1 was previously shown to methylate *Arabidopsis* $tRNA^{Asp(GTC)}$ at position C38 [20]. In Chapter 2, we independently confirmed this finding and extend the known target sites to position C38 of $tRNA^{Gly(GCC)}$ and $tRNA^{Gly(CCC)}$. We identified *Arabidopsis* mutants for the two TRM4 homologs, TRM4A and TRM4B. TRM4A is predicted to be non-functional, as TRM4A contains a deletion in motif I which is required for target cytosine binding [21](Chapter 2). In Chapter 2, methylation of tRNAs at structural positions C48, C49 and C50 was shown to be mediated by TRM4B and no methylated sites were altered in *trm4a* mutants, supporting the prediction of TRM4A being catalytically inactive.
Furthermore in Chapter 3, we identified over 100 additional m^5C sites dependent on TRM4B in other classes of RNAs such as mRNAs and snoRNAs across three different tissue types in *Arabidopsis*. In comparison, we could not identify any TRDMT1 dependent sites in seedling shoots, suggesting that TRDMT1 is tRNA specific (Chapter 3). However, we may have missed potential TRDMT1 targeting sites due to the limits of sequencing depth, as two mRNA m^5C sites have been shown to be dependent of TRDMT1 in human cells [9]. Perhaps TRDMT1 may target cytosines in other classes of RNA when they mimic the secondary structure of TRDMT1 target sites in tRNAs. While we reported thousands of m^5C sites across three different *Arabidopsis* tissue types, only a little over 100 of these sites were found to be dependent on TRM4B. This may suggest that many of these candidate m^5C sites which are independent of both TRM4B and TRDMT1 may not be m^5C sites, but may be other modifications such as m^3C, which are also resistant to bisulfite treatment, as discussed above. Using RIP with anti-m^5C antibodies could help to confirm if these TRM4B independent sites are m^5C sites, or another, bisulfite resistant RNA modification. Another alternative is that many of these sites may be mediated by another, related RNA methyltransferase, named RCMT9. RCMT9 has greater sequence similarity with TRM4B and TRM4A than with the other *Arabidopsis* rRNA methyltransferases (data not shown, [18]), suggesting that RCMT9 may have

a similar function to TRM4B and may also methylate coding RNAs. Future RBS-seq experiments are required to identify RCMT9 dependent m^5C target sites. Together, these data suggest that TRM4B and potentially also RCMT9, play a major role in mediating m^5C methylation transcriptome-wide in plants.

To determine the role of TRM4B in plant development, we investigated primary root growth (Chapter 3). We demonstrated *trm4b* mutants have a short root phenotype as a consequence of reduced cell division when compared to wild type. Furthermore, *trm4b* mutants are more sensitive to oxidative stress and have reduced RNA stability of tRNAs targeted by TRM4B. We propose that alterations in the tRNA pool and possibly in translation rates due to reduced tRNA abundance may be causing reduced cell division in the root meristem. These results agree will roles for TRM4 in other organisms, such as regulating stem cell proliferation and differentiation and also in oxidative stress tolerance [22-24].

One of the target sites of TRM4B is position C3349 in *MAG5* mRNA. This site is methylated at 55%, 33% and 26% in siliques, shoots and roots respectively in *Arabidopsis* wild type and methylation is completely lost in *trm4b* mutants [11]. The physiological relevance of *MAG5* mRNA methylation in particular is yet to be elucidated. We tested if the methylation affected *MAG5* RNA stability by examining *MAG5* transcript abundance in wild type and *trm4b* mutant plants. *MAG5* transcript levels were unchanged, suggesting that RNA stability was not affected (data not shown). Alternatively, it is possible that non-methylated *MAG5* mRNA is degraded faster, and that transcription is increased in *trm4b* mutants to compensate for increased degradation rates, so that the steady-state levels of *MAG5* mRNA are the same in wild type and *trm4b*. Further experiments to specifically examine transcription rates and degradation rates of *MAG5* mRNA are required to delve deeper into answering this question. Examples include nuclear run-on assays or RNA Immunoprecipitation using antibodies against Pol II to determine transcription rates and utilization of transcription inhibitors such as Actinomycin D to examine rates of degradation (methods reviewed in[25]). Furthermore, in a preliminary experiment, the amount of translated MAG5 protein was indirectly investigated. MAG5 is involved in export of precursors of the storage proteins 12S globulin and 2S

albumin precursors from endoplasmic reticulum exit sites. Consequently, *mag5* mutants have increased accumulation of these storage protein precursors and decreased levels of mature storage proteins in mature, dry seeds [26]. We hypothesised that if methylation was required for efficient translation of MAG5, that in the absence of methylation in *trm4b* mutants, less MAG5 protein would be produced and that there would be increased levels of storage protein precursors and decreased abundance of mature storage proteins in *trm4b* mutants compared to wild type. The levels of 12S globulin were unchanged between wild type and *trm4b*, suggesting that methylation is not required for efficient translation of MAG5 (data not shown, experiment performed by Trung Do). It is possible that this indirect technique was not sensitive enough to determine if methylation of MAG5 by TRM4B has a small effect on MAG5 translation, but not enough to reduce MAG5 protein abundance to a level where storage protein abundance is affected. The methylation of *MAG5* mRNA has thus far not been able to help to explain the observed phenotype of *trm4b* mutants, as no shared phenotypes were observed in *mag5* and *trm4b* mutant plants [11, 13, 26]. *MAG5* mRNA is only one of over 100 TRM4B dependent m^5C sites that we have discovered and other TRM4B dependent sites may have direct roles in mediating the phenotype of *trm4b* mutants. It is possible that subtle effects of loss of *MAG5* mRNA methylation are yet to be uncovered under abiotic or biotic stress conditions. Future experiments may uncover a function for *MAG5* mRNA methylation by TRM4B.

We tested the determinants for targeting of TRM4B in Chapter 3, by performing a motif analysis and were unable to identify a consensus sequence targeting TRM4B methylation. We subsequently demonstrated a 50nt sequence flanking the m^5C site C3349 in *MAG5* mRNA is sufficient to confer methylation of a transgene reporter in *Nicotiana benthamiana*. We propose that local RNA structure or other complex (protein and/or RNA) components may modulate the methylation activity of TRM4B in *A. thaliana* and *N. benthamiana.* These results suggest conservation of the targeting mechanism of TRM4B across different plant species and provide insights into determinants required for targeting m^5C methylation.

While the functional relevance of m^5C in tRNAs and the respective RNA modifying enzymes is remarkably conserved, there are also differences between different

organisms, and there may be plant-specific functions for this modification (reviewed in [18, 19, 27-30]). Plants provide a unique comparative system for investigating the origin and evolution of m^5C as they contain three different genomes derived from the nucleus, mitochondria and chloroplast. We discovered that in *Arabidopsis*, many nuclear tRNAs contained m^5C sites, while no methylated sites were detected in mitochondrial or chloroplast tRNAs. This trend was generally conserved across the other five plant species we examined in Chapter 2. This contrasts with studies performed in mammals, where methylated sites were detected in mitochondrial tRNAs [7, 31, 32]. This suggests independent evolution of organelle methylation in animals and plants.

While there is strong conservation of nuclear tRNA methylation in the variable region and in the anticodon loop across different organisms [7, 13, 23, 33], there are also kingdom and species-specific differences in methylation sites and also differences for functions of tRNA methylation by TRDMT1/DNMT2 and TRM4.

For example, methylation mediated by TRDMT1/DNMT2 appears to have different functions in different species as *trdmt1* deficiency is lethal for zebra fish [34], while *Arabidopsis*, mouse and *Drosophila* appear wild type under standard laboratory conditions [13, 20]. In some species, the functions of *trdmt1* are only apparent under stress conditions, for example *trdmt1* deficient *Drosophila* are sensitive to oxidative and heat stress [35, 36], whereas *Arabidopsis trdmt1* mutants did not show increased sensitivity to oxidative stress in Chapter 3. These differences in functions may be explained by either different methylated tRNA targets of TRDMT1 in different species, or different functions for the same tRNA targets in different species. TRDMT1 is responsible for methylating cytosine 38 in tRNAs $Asp^{(GTC)}$, $Gly^{(GCC)}$, and $Val^{(ACC)}$ in mouse, human and *Drosophila* [9, 35, 37]. In contrast, while *Arabidopsis* TRDMT1 also methylates cytosine 38 of tRNAs $Asp^{(GTC)}$ and $Gly^{(GCC)}$, *Arabidopsis* TRDMT1 does not methylate $tRNA^{Val(ACC)}$, but instead methylates $tRNA^{Gly(CCC)}$ [13, 20]. One possibility for why there is animal and plant specific methylation of tRNAs $Val^{(AAC)}$ and $Gly^{(CCC)}$ by TRDMT1 is that methylation of these different tRNAs may be required for species specific translation of subsets of proteins with biased codon usage. Furthermore, even though tRNAs $Gly^{(GCC)}$ and $Asp^{(GTC)}$ are methylated in both animals and plants, their functions may be species specific due to species

specific codon usage biases in subsets of proteins. An example of this was recently shown when a subtle, previously over-looked function was discovered for TRDMT1 in the mouse [38]. *trdmt1* deficient mice have reduced proliferative capacity of haematopoietic stem cells. While global protein translation rates are unchanged in *trdmt1* mutants, TRDMT1 methylation of $tRNA^{Asp(GTC)}$ was found to be required for efficient and accurate translation of a subset of proteins in bone marrow, where haematopoietic stem cells are produced.

Likewise, there are also similarities and differences in functions for TRM4/NSUN2 in mammals and plants. TRM4 has been shown to be required for methylation of cytosine positions 34, 48, 49 and 50 in tRNAs of both animals, yeast and plants [7-10, 13, 17, 23, 39]. While TRM4 has been shown to be involved in regulating stem cell proliferation and differentiation in both plants and mammals, there are differences in the exact function of TRM4 [11, 22, 23, 37]. More specifically, TRM4 deficiency was shown to delay differentiation of epidermal stem cells, resulting in accumulation of proliferating cells in the mouse [22, 37], while TRM4 deficiency in *Arabidopsis* resulted in reduced numbers of root apical meristem cells, caused by reduced ability of the meristem to divide [11]. These differences in the regulation of stem cells in mammals and plants may be explained by many factors. While loss of TRM4 methylation of tRNAs in mammals leads to increased abundance of tRNA fragments from tRNAs which require TRM4 for methylation, steady-state levels of these TRM4 dependent tRNAs are unchanged [23, 37]. In contrast, TRM4 deficiency in plants causes decreased levels of mature $tRNA^{Asp(GTC)}$ and decreased levels of tRNA fragments ([11], data not shown). These very different outcomes for tRNAs in TRM4 deficient mammals and plants may explain their differences in function. The increased abundance of tRNA fragments in mammals is thought to induce inhibition of translation initiation [40], while decreased abundance of mature tRNAs in *Arabidopsis trm4* mutants may lead to reduced translation efficiencies for a subset of proteins, which are biased for those codons. The reduction in potentially translation inhibiting tRNA fragments in *Arabidopsis* may also have effects on translation regulation, and may lead to an increase in translation of a subset of proteins. Furthermore, the differences in TRM4 function between mammals and plants may also be explained in part by methylation of different mRNA targets, which may have different functions in regulating stem cell renewal and differentiation. The very low

conservation of m^5C mRNA target sites between human and mouse [10, 15], suggests that most of the m^5C mRNA sites in plants will be different from m^5C mRNA target sites in mammals.

Together, these data shown in this thesis, represent the first transcriptome-wide high-resolution view of m^5C in plants and provides links to biological function. Moreover, the identification and characterization of *Arabidopsis* loss of function genetic mutants provides a valuable resource for the research community to further probe the functions of these m^5C 'writers'. Significant discussion has already been undertaken (Chapters 2 and 3) of these research findings and below I discuss broader implications of RNA modifications in Nature.

New Functions for Old RNA Modifications

Although the presence of RNA modifications in mRNAs such as m^6A [41, 42] and the 5' cap structure ([43-46] and reviewed in [47, 48]) were discovered decades ago, new functions for these modifications are still being discovered and the mechanisms elucidated (reviewed in [27]). In the following sections, I discuss potential new functions for RNA modifications and review the supporting evidence. In particular, I discuss the recent discovery of RNA modification 'erasers', uncovering new layers for the regulation of RNA modifications. Furthermore, the broader functions of RNA modifications are discussed in relation to plant immune responses and defence against pathogens. Possible contributions of RNA modifications to hybridization barriers and hybrid vigour are discussed, as understanding these processes is important for development and improvement of agricultural crops.

Dynamic, Reversible Regulation of RNA Modifications – 'Yin and Yang'

One intriguing mechanism for regulating RNA structure and function is through the active removal of RNA modifications. This balancing act of removal and addition of modifications can be likened to the concept of yin and yang in Chinese philosophy, where yin and yang symbolize two opposing forces which complement each other and are interdependent. While dynamic regulation of both m^6A and very recently

m^5C, has been shown by the opposing actions of RNA modification ‘writers’ and ‘erasers’ in animals, it is unclear if this also occurs for other RNA modifications [49-52]. Investigation of potential ‘erasers’ will allow insights into mechanisms and functions of RNA modifications in development, stress responses and potentially human diseases.

Many potential RNA modification ‘erasers’, with diverse substrate specificities have been predicted, suggesting that many other RNA modifications could also be dynamically regulated [53]. Further research is required to match candidate ‘erasers’ with the modifications they remove. Both RNA and DNA modification ‘erasers’ have been identified in the superfamily of Fe(II)- and 2-oxoglutarate-dependent oxygenases (Fe(II)/2-OG-dependent oxygenases) [54]. In particular, the ALKB dioxygenase homologues (ALKBH) family and Ten-Eleven Translocation (TET) subfamilies have both been shown to remove RNA and DNA modifications [49, 50, 52, 55, 56]. Dynamic regulation of m^6A has been demonstrated by the ALKBH family proteins Fat mass and obesity associated protein (FTO) and Alkylation repair homologue protein 5 (ALKBH5) [49, 50] and discussed in (Chapter 1). Similarly to other ALKBH family proteins, FTO is not restricted to one substrate. FTO preferentially removes m^6A in RNA, but also demethylates 3-methyluracil (m^3U) in RNA and 3-methylthymidine (m^3T) in DNA [50, 57]. Meanwhile, mammalian TET enzymes have been demonstrated to demethylate m^5C to 5-hydroxymethylcytosine (hm^5C) in DNA, and more recently in RNA [52, 55, 56]. For modifications which are not predicted to be removed by Fe(II)/2-OG-dependent oxygenases, the mechanism may require DNA/RNA glycosylases, or other, as yet unidentified enzymes. In plants, RNA m^5C may also be removed by DNA/RNA glycosylases or other enzymes, as no potential homologs of TET enzymes were identified [53]. Possible plant specific RNA modification ‘erasers’ deserve further investigation to answer these questions. Future experiments to identify potential ‘erasers’ could utilize HPLC analysis of rRNA depleted RNA from mutants in candidate genes to identify mRNA modifications which are increased in the absence of ‘erasers’.

A possible function of RNA modification ‘erasers’ may be to ensure the specificity of ‘writers’ by proof-reading or removing modifications which may be deleterious. This may help to partially explain why the targeting mechanisms of many RNA modifications are not fully understood. To illustrate this point, although consensus sequences were identified for the m^6A ‘writer’ complex (RR**A**CH), and the Ψ ‘writers’ PUS4 (GU**Ψ**C/NANNC) and PUS7 (UG**Ψ**A/R), not all sites with these consensus sequences are modified, suggesting that other factors such as ‘erasers’ are editing the activity of the RNA modification ‘writers’ [58-61]. This may also be the case for RNA modifications which are not found in consensus sequences, such as m^5C [7, 10] and (Chapter 3). Removal of m^5C sites by demethylases may be modulated by the local RNA structure to control not only whether a cytosine is methylated, but also perhaps the percentage of RNA molecules which are methylated at that position. ‘Erasers’ could also potentially regulate RNA modifications in response to various stresses and control tissue and even cell compartment specific levels of modifications in plants.

This seemingly simple picture of yin and yang RNA modification ‘erasers’ and ‘writers’ is not as black and white as it seems. It’s not just an on or off mechanism as some ‘eraser’ reactions have multiple intermediates. For example, TET enzymes mediate iterative oxidative demethylation of m^5C to produce hm^5C, then 5-formylcytosine (f^5C) and then subsequently 5-carboxylcytosine (ca^5C) in DNA and have also been shown to mediate oxidative demethylation of m^5C to hm^5C in RNA [52, 56]. The intermediate hm^5C in DNA is associated with transcriptional activation and cell differentiation, while RNA hm^5C has recently been implicated in increasing translation of mRNAs [62-64]. It is unknown whether other ‘eraser’ intermediates play a functional role in RNA, or if they are passive steps in the removal of other modifications. If other intermediates do have roles in RNA function, or attract RNA binding proteins or ‘readers’, this will add a further layer to the complexity of dynamic regulation of the epitranscriptome.

RNA Modifications – Instruments of War

Several mechanisms for RNA modifications in mediating immune responses against pathogens have already been elucidated in animals, lending support to the possibility of additional functions for RNA modifications in plant immune responses [65]. For example, bacterial tRNA modifications can inhibit immune responses in human cells [66]. Furthermore, modification of *in vitro* transcribed mRNAs can also inhibit immune responses, leading to increased translation efficiency [67, 68]. Modified yeast tRNAs can also be targets for microbe produced toxins, resulting in tRNA cleavage and reduced growth [69]. More recently, a role for tRNA modifications in mediating plant immunity against bacterial pathogens was discovered [6]. In *Arabidopsis*, the tRNA methyltransferase SUPPRESSOR OF CSB3 9 (SCS9) was shown to be required for resistance to *Pseudomonas syringe.* SCS9 was required for increased 2'-O-ribose methylation of tRNA anticodons in response to infection, and this is thought to increase translation of proteins required for an effective immune response [6]. Additional roles for RNA modifications in plant immunity await discovery.

Unlike animals, plants largely lack adaptive immunity (except viral derived siRNA priming) and must rely solely on innate immunity to detect conserved molecules common in pathogens such as flagellin or chitin to trigger immune responses (reviewed in [70, 71]). These molecules are also known as pathogen associated molecular patterns (PAMPs). In order to suppress plant immune responses to PAMPs, pathogens have evolved effector proteins. In response, plants have evolved resistance (R) proteins to fight back by recognising or inhibiting pathogen effector proteins. In addition, both plants and their pathogens utilize non-coding small RNAs such as miRNAs and siRNAs as weapons in this evolutionary struggle [71]. RNA modifications may be required for increased stability and effectiveness of these small RNAs in defence and virulence. Furthermore, plant hosts and pathogens could target each other's transcripts for modifications to alter abundance and translation of mRNAs. As discussed above, plant 'erasers' could also act in this war, removing RNA modifications which increase pathogen virulence. In effect, RNA modification 'writers' and 'erasers' could act as additional effector proteins and R proteins in this

evolutionary battle. ALKBH family ‘erasers’ have been identified in several plant viruses, but their roles in virulence are yet to be determined [72].

RNA Modifications in Parental Conflict and Hybrid Vigour

A major goal in plant science is to understand and break through hybridization barriers to allow the creation of new crop plants with hybrid vigour and to increase the genetic diversity and disease resistance of agriculturally important crops. Bypassing hybridization barriers is not an easy task, as the mechanisms are varied and there are both pre- and post- fertilization barriers to overcome. Examples of pre-fertilization barriers include pollen tube guidance and pollen-stigma incompatibilities such as self-incompatibility [73-75]. While post-fertilization barriers include differences in genome dosages (ploidy levels), activation of transposable elements and interspecific differences in genomic imprinting [76-78]. The potential functions of RNA modifications in enabling and inhibiting hybridization are yet to be explored.

At least three different mechanisms for self-incompatibility have evolved independently in flowering plants to promote out-crossing (reviewed in [74, 75]). These mechanisms inhibit self-fertilization and may also inhibit interspecific crosses [74]. One mechanism in *Brassicaceae* involves pollen derived proteins, which are recognized by maternal receptor proteins and enable ‘self-recognition’. Only certain combinations of paternal proteins and maternal receptors allow fertilization, decreasing the rate of self-pollination. This mechanism can be likened to using specific pollen keys to unlock maternal receptors. When crossing different species, the maternal receptors can act as locks to inhibit pollination, as the appropriate pollen protein, or key is not encoded in the other species genome. For example, *Arabidopsis thaliana* has lost both the paternal and maternal components for self-incompatibility, allowing self-pollination. *Arabidopsis arenosa* on the other hand has these components intact. As a result, *A. arenosa* can pollinate *A. thaliana*, as *A. thaliana* has no maternal receptor to block pollen tube growth, while *A. thaliana* pollen cannot fertilize *A. arenosa*, as it does not contain the appropriate pollen key [77]. Likewise, another potential mechanism for self-incompatibility and recognition of foreign pollen could be RNA based, instead of protein based. The pattern or absence of diverse RNA modifications on specific RNAs from pollen may be

deciphered by RNA modification 'reader' proteins, to differentiate self- from non-self-pollination events and to identify pollen from other species. Furthermore, pollen derived RNAs may also be modified to alter their function and may be required to complete fertilization. Similarly in animals, thousands of diverse RNAs from sperm including both mRNAs and ncRNAs are hypothesized to play varied roles in fertilization [79]. Modification of these RNAs may be required for their activity and increase stability [80]. Potentially species specific male gamete RNA modifications may create pre-fertilization hybridization barriers.

Furthermore, RNA modifications may also play roles in post-fertilization barriers. Here, we propose a new layer of genetic imprinting – RNA imprinting. Similar to genomic imprinting [81], RNA molecules may be differentially modified, based on whether they were transcribed from the maternal or the paternal genomic alleles. As the triploid endosperm of flowering plants contains twice as much genetic material from the mother, compared to the father, this would result in regulating the percentage of a pool of RNA transcripts which contained the modification. This could potentially be a mechanism for subtle regulation of the function and abundance of the imprinted transcripts, allowing fine-tuning of seed development. Like genomic imprinting, RNA imprinting may also cause hybridization barriers. Genomic imprinting can cause post-fertilization hybridization barriers, as many imprinted genes are not conserved across species or even across different stains, leading to imbalances and miss-expression of genes [76-78]. This loss of genomic imprinting and de-regulation of genes leads to defects in endosperm development, resulting in hybrid failure [76, 77, 82].

Another post-fertilization barrier that RNA modifications may be involved in, is the regulation of transposable elements. In *Drosophila*, piRNAs are maternally derived and silence transposable elements in developing eggs. If the paternal genome contains a transposable element not present in the maternal parent, it becomes reactivated. This process is known as hybrid dysgenesis and results in sterile male offspring [83, 84]. I hypothesized that RNA modifications such as m^5C may be required for effective silencing of transposons by increasing stability of maternally

derived piRNAs (animals) or siRNAs (plants) and mediate hybridization barriers. In support of this, DNMT2 was found to be required for transposon silencing in *Drosophila*, however due to conflicting reports, it is unknown whether this is due to its RNA or DNA methylation activity [85, 86].

RNA modifications may also modulate another agriculturally important trait, hybrid vigour. Hybrid vigour, or heterosis, is the production of superior offspring, with improved qualities compared to both parents, such as increased biomass and stress tolerance. The underlying mechanisms of hybrid vigour are still being explored, such as DNA methylation, small RNAs and differential gene regulation [87, 88]. Beneficial combinations of RNA modifications on transcripts might be produced in F_1 hybrids, producing improved crop traits due to new combinations of strain specific modifications inherited from the parents. Strain-specific m^6A modifications in plants have already been demonstrated for the *Arabidopsis* ecotypes Hen-16 and Can-0 [89], supporting this possibility. In addition, hybrid vigour mediated by new deposition patterns of RNA modifications may be inherited, leading to hybrid vigour being retained in offspring of F_1 hybrids. RNA modifications have already been shown to be inheritable and necessary in epigenetic events such as paramutation and inheritance of an acquired metabolic disorder in mice [80, 90]. The possible functions of RNA modifications in hybridization and hybrid vigour in animals and plants merits further investigation.

Conclusions and Future Directions

In summary, this thesis presents the first transcriptome-wide high-resolution maps of m^5C sites which occur in *Arabidopsis thaliana* mRNAs, tRNAs, rRNAs and other non-coding RNAs. These methylated sites are dynamically regulated between tissue types and many of these sites are dependent on TRM4B. Furthermore, we report identification and/or characterization of T-DNA mutants for predicted m^5C RMTases in *Arabidopsis*, providing a means to further investigate the functions of these genes in RNA metabolism, plant development and stress responses.

As many of these RMTases methylate, or are predicted to methylate tRNAs and rRNAs, mutations in these genes are expected to perturb translation, abundance, and processing of tRNAs and rRNAs. Future experiments examining protein synthesis and ribosome activity such as polysome profiling, and the use of reporter constructs to analyse specific defects in translation such as stop codon read through will shed light on the functions of these RMTases [91, 92]. In order to determine the role of rRNA methyltransferases in rRNA biogenesis and processing, northern blots, RT-PCR and assessment of poly-adenylation of rRNA can be performed to detect alterations in rRNA RMTase mutant compared to wildtype [93, 94].

These data, combined with recent developments in uncovering the *Arabidopsis* epitranscriptome (reviewed in Chapter 1) and the potential functions of dynamic, reversible RNA modifications in plant immunity, hybrid vigour and hybridization barriers, open up perspectives for further analysis and deepen our understanding of the biological functions of RNA modifications in *Plantae*.

References

1. Torres AG, Batlle E, de Pouplana LR: **Role of tRNA modifications in human diseases**. *Trends Mol Med* 2014, **20**(6):306-314.
2. Blanco S, Frye M: **Role of RNA methyltransferases in tissue renewal and pathology**. *Current opinion in cell biology* 2014, **31**:1-7.
3. Charette M, Gray MW: **Pseudouridine in RNA: what, where, how, and why**. *IUBMB life* 2000, **49**(5):341-351.
4. Chan CTY, Pang YLJ, Deng WJ, Babu IR, Dyavaiah M, Begley TJ, Dedon PC: **Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins**. *Nature communications* 2012, **3**:937.
5. Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ: **A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress**. *PLoS genetics* 2010, **6**(12):e1001247.
6. Ramirez V, Gonzalez B, Lopez A, Castello MJ, Gil MJ, Zheng B, Chen P, Vera P: **Loss of a Conserved tRNA Anticodon Modification Perturbs Plant Immunity**. *PLoS genetics* 2015, **11**(10):e1005586.
7. Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T: **Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA**. *Nucleic acids research* 2012, **40**(11):5023-5033.
8. Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R: **Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs**. *PLoS genetics* 2013, **9**(6):e1003602.
9. Khoddami V, Cairns BR: **Identification of direct targets and modified bases of RNA cytosine methyltransferases**. *Nat Biotechnol* 2013, **31**(5):458-464.

10. Hussain S, Sajini AA, Blanco S, Dietmann S, Lombard P, Sugimoto Y, Paramor M, Gleeson JG, Odom DT, Ule J *et al*: **NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs**. *Cell reports* 2013, **4**(2):255-261.
11. David R, Burgess AL, Parker BJ, Pulsford KE, Sibbritt T, Preiss T, Searle IR: **Transcriptome-wide mapping of RNA 5-methylcytosine in Arabidopsis mRNAs and ncRNAs**. *Under submission to Nature Commun* 2016.
12. Schaefer M, Pollex T, Hanna K, Lyko F: **RNA cytosine methylation analysis by bisulfite sequencing**. *Nucleic acids research* 2009, **37**(2):e12.
13. Burgess AL, David R, Searle IR: **Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae**. *BMC plant biology* 2015, **15**:199.
14. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A: **The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing**. *Plos One* 2010, **5**(1):e8888.
15. Hussain S, Aleksic J, Blanco S, Dietmann S, Frye M: **Characterizing 5-methylcytosine in the mammalian epitranscriptome**. *Genome biology* 2013, **14**(11):215.
16. Shafik A, Schumann U, Evers M, Sibbritt T, Preiss T: **The emerging epitranscriptomics of long noncoding RNAs**. *Biochimica et biophysica acta* 2016, **1859**(1):59-70.
17. Motorin Y, Lyko F, Helm M: **5-methylcytosine in RNA: detection, enzymatic formation and biological functions**. *Nucleic acids research* 2010, **38**(5):1415-1430.
18. Pavlopoulou A, Kossida S: **Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases**. *Genomics* 2009, **93**(4):350-357.
19. Chen P, Jager G, Zheng B: **Transfer RNA modifications and genes for modifying enzymes in Arabidopsis thaliana**. *BMC plant biology* 2010, **10**:201.
20. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang XY, Golic KG, Jacobsen SE, Bestor TH: **Methylation of tRNA(AsP) by the DNA methyltransferase homolog Dnmt2**. *Science* 2006, **311**(5759):395-398.
21. Bujnicki JM, Feder M, Ayres CL, Redman KL: **Sequence-structure-function studies of tRNA : m(5)C methyltransferase Trm4p and its relationship to DNA : m(5)C and RNA : m(5)U methyltransferases**. *Nucleic acids research* 2004, **32**(8):2453-2463.
22. Blanco S, Kurowski A, Nichols J, Watt FM, Benitah SA, Frye M: **The RNA-methyltransferase Misu (NSun2) poises epidermal stem cells to differentiate**. *PLoS genetics* 2011, **7**(12):e1002403.
23. Blanco S, Dietmann S, Flores JV, Hussain S, Kutter C, Humphreys P, Lukk M, Lombard P, Treps L, Popis M *et al*: **Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders**. *The EMBO journal* 2014, **33**(18):2020-2039.
24. Hussain S, Tuorto F, Menon S, Blanco S, Cox C, Flores JV, Watt S, Kudo NR, Lyko F, Frye M: **The Mouse Cytosine-5 RNA Methyltransferase NSun2 Is a Component of the Chromatoid Body and Required for Testis Differentiation**. *Mol Cell Biol* 2013, **33**(8):1561-1570.
25. Perez-Ortin JE, de Miguel-Jimenez L, Chavez S: **Genome-wide studies of mRNA synthesis and degradation in eukaryotes**. *Biochimica et biophysica acta* 2012, **1819**(6):604-615.
26. Takagi J, Renna L, Takahashi H, Koumoto Y, Tamura K, Stefano G, Fukao Y, Kondo M, Nishimura M, Shimada T *et al*: **MAIGO5 functions in protein export from Golgi-associated endoplasmic reticulum exit sites in Arabidopsis**. *Plant Cell* 2013, **25**(11):4658-4675.
27. Burgess A, David R, Searle IR: **Deciphering the epitranscriptome: A green perspective**. *J Integr Plant Biol* 2016.
28. Jackman JE, Alfonzo JD: **Transfer RNA modifications: nature's combinatorial chemistry playground**. *Wires Rna* 2013, **4**(1):35-48.
29. Towns WL, Begley TJ: **Transfer RNA methytransferases and their corresponding modifications in budding yeast and humans: activities, predications, and potential roles in human health**. *DNA Cell Biol* 2012, **31**(4):434-454.
30. Swinehart WE, Jackman JE: **Diversity in mechanism and function of tRNA methyltransferases**. *RNA biology* 2015, **12**(4):398-411.

31. Suzuki T, Suzuki T: **A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs**. *Nucleic acids research* 2014, **42**(11):7346-7357.
32. Suzuki T, Nagao A, Suzuki T: **Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases**. *Annu Rev Genet* 2011, **45**:299-329.
33. Motorin Y, Helm M: **RNA nucleotide methylation**. *Wiley interdisciplinary reviews RNA* 2011, **2**(5):611-631.
34. Rai K, Chidester S, Zavala CV, Manos EJ, James SR, Karpf AR, Jones DA, Cairns BR: **Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish**. *Genes & development* 2007, **21**(3):261-266.
35. Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M, Lyko F: **RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage**. *Genes & development* 2010, **24**(15):1590-1595.
36. Lin MJ, Tang LY, Reddy MN, Shen CK: **DNA methyltransferase gene dDnmt2 and longevity of Drosophila**. *The Journal of biological chemistry* 2005, **280**(2):861-864.
37. Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M, Stoecklin G, Lyko F: **RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis**. *Nature structural & molecular biology* 2012, **19**(9):900-905.
38. Tuorto F, Herbst F, Alerasool N, Bender S, Popp O, Federico G, Reitter S, Liebers R, Stoecklin G, Grone HJ *et al*: **The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis**. *The EMBO journal* 2015, **34**(18):2350-2362.
39. Motorin Y, Grosjean H: **Multisite-specific tRNA : m(5)C-methyltransferase (Trm4) in yeast Saccharomyces cerevisiae: Identification of the gene and substrate specificity of the enzyme**. *RNA* 1999, **5**(8):1105-1118.
40. Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P: **Angiogenin-induced tRNA fragments inhibit translation initiation**. *Mol Cell* 2011, **43**(4):613-623.
41. Perry RP, Kelley DE: **Existence of methylated messenger RNA in mouse L cells**. *Cell* 1974, **1**:37-42.
42. Desrosiers R, Friderici K, Rottman F: **Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells**. *Proc Natl Acad Sci USA* 1974, **71**(10):3971-3975.
43. Furuichi Y, Morgan M, Muthukrishnan S, Shatkin AJ: **Reovirus messenger RNA contains a methylated, blocked 5'-terminal structure: m-7G(5')ppp(5')G-MpCp**. *Proc Natl Acad Sci U S A* 1975, **72**(1):362-366.
44. Perry RP, Kelley DE, Friderici K, Rottman F: **The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus**. *Cell* 1975, **4**(4):387-394.
45. Wei CM, Moss B: **Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA**. *Proc Natl Acad Sci U S A* 1975, **72**(1):318-322.
46. Wei CM, Gershowitz A, Moss B: **Methylated nucleotides block 5' terminus of HeLa cell messenger RNA**. *Cell* 1975, **4**(4):379-386.
47. Shatkin AJ: **Capping of eucaryotic mRNAs**. *Cell* 1976, **9**(4 PT 2):645-653.
48. Furuichi Y: **Discovery of m(7)G-cap in eukaryotic mRNAs**. *Proc Jpn Acad Ser B Phys Biol Sci* 2015, **91**(8):394-409.
49. Zheng GQ, Dahl JA, Niu YM, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH *et al*: **ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility**. *Mol Cell* 2013, **49**(1):18-29.
50. Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG *et al*: **N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO**. *Nat Chem Biol* 2011, **7**(12):885-887.
51. Zhao X, Yang Y, Sun BF, Shi Y, Yang X, Xiao W, Hao YJ, Ping XL, Chen YS, Wang WJ *et al*: **FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis**. *Cell Res* 2014, **24**(12):1403-1419.

52. Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, Cai Q, Ji D, Jin SG, Niedernhofer LJ *et al*: **Tet-mediated formation of 5-hydroxymethylcytosine in RNA**. *J Am Chem Soc* 2014, **136**(33):11582-11585.
53. Iyer LM, Tahiliani M, Rao A, Aravind L: **Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids**. *Cell Cycle* 2009, **8**(11):1698-1710.
54. Fedeles BI, Singh V, Delaney JC, Li D, Essigmann JM: **The AlkB Family of Fe(II)/alpha-Ketoglutarate-dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond**. *The Journal of biological chemistry* 2015, **290**(34):20734-20742.
55. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L *et al*: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1**. *Science* 2009, **324**(5929):930-935.
56. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y: **Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine**. *Science* 2011, **333**(6047):1300-1303.
57. Jia G, Yang CG, Yang S, Jian X, Yi C, Zhou Z, He C: **Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO**. *FEBS letters* 2008, **582**(23-24):3313-3319.
58. Wei CM, Moss B: **Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid**. *Biochemistry* 1977, **16**(8):1672-1676.
59. Csepany T, Lin A, Baldick CJ, Jr., Beemon K: **Sequence specificity of mRNA N6-adenosine methyltransferase**. *The Journal of biological chemistry* 1990, **265**(33):20117-20122.
60. Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, Leon-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES *et al*: **Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA**. *Cell* 2014, **159**(1):148-162.
61. Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV: **Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells**. *Nature* 2014, **515**(7525):143-146.
62. Delatte B, Wang F, Ngoc LV, Collignon E, Bonvin E, Deplus R, Calonne E, Hassabi B, Putmans P, Awe S *et al*: **RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine**. *Science* 2016, **351**(6270):282-285.
63. Gross JA, Pacis A, Chen GG, Barreiro LB, Ernst C, Turecki G: **Characterizing 5-hydroxymethylcytosine in human prefrontal cortex at single base resolution**. *Bmc Genomics* 2015, **16**:672.
64. Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G *et al*: **Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells**. *Cell Stem Cell* 2011, **8**(2):200-213.
65. Dalpke AH, Helm M: **RNA mediated Toll-like receptor stimulation in health and disease**. *RNA biology* 2012, **9**(6):828-842.
66. Gehrig S, Eberle ME, Botschen F, Rimbach K, Eberle F, Eigenbrod T, Kaiser S, Holmes WM, Erdmann VA, Sprinzl M *et al*: **Identification of modifications in microbial, native tRNA that suppress immunostimulatory activity**. *J Exp Med* 2012, **209**(2):225-233.
67. Kariko K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, Weissman D: **Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability**. *Mol Ther* 2008, **16**(11):1833-1840.
68. Kariko K, Weissman D: **Naturally occurring nucleoside modifications suppress the immunostimulatory activity of RNA: implication for therapeutic RNA development**. *Curr Opin Drug Discov Devel* 2007, **10**(5):523-532.
69. Lu J, Huang B, Esberg A, Johansson MJ, Bystrom AS: **The Kluyveromyces lactis gamma-toxin targets tRNA anticodons**. *RNA* 2005, **11**(11):1648-1654.
70. Jones JD, Dangl JL: **The plant immune system**. *Nature* 2006, **444**(7117):323-329.

71. Weiberg A, Wang M, Bellinger M, Jin H: **Small RNAs: a new paradigm in plant-microbe interactions**. *Annu Rev Phytopathol* 2014, **52**:495-516.
72. van den Born E, Omelchenko MV, Bekkelund A, Leihne V, Koonin EV, Dolja VV, Falnes PO: **Viral AlkB proteins repair RNA damage by oxidative demethylation**. *Nucleic acids research* 2008, **36**(17):5451-5461.
73. Marton ML, Fastner A, Uebler S, Dresselhaus T: **Overcoming hybridization barriers by the secretion of the maize pollen tube attractant ZmEA1 from Arabidopsis ovules**. *Curr Biol* 2012, **22**(13):1194-1198.
74. Kitashiba H, Nasrallah JB: **Self-incompatibility in Brassicaceae crops: lessons for interspecific incompatibility**. *Breed Sci* 2014, **64**(1):23-37.
75. Silva NF, Goring DR: **Mechanisms of self-incompatibility in flowering plants**. *Cell Mol Life Sci* 2001, **58**(14):1988-2007.
76. Lafon-Placette C, Kohler C: **Epigenetic mechanisms of postzygotic reproductive isolation in plants**. *Curr Opin Plant Biol* 2015, **23**:39-44.
77. Josefsson C, Dilkes B, Comai L: **Parent-dependent loss of gene silencing during interspecies hybridization**. *Curr Biol* 2006, **16**(13):1322-1328.
78. Wolff P, Jiang H, Wang G, Santos-Gonzalez J, Kohler C: **Paternally expressed imprinted genes establish postzygotic hybridization barriers in Arabidopsis thaliana**. *Elife* 2015, **4**.
79. Hosken DJ, Hodgson DJ: **Why do sperm carry RNA? Relatedness, conflict, and control**. *Trends Ecol Evol* 2014, **29**(8):451-455.
80. Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, Feng GH, Peng H, Zhang X, Zhang Y *et al*: **Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder**. *Science* 2016, **351**(6271):397-400.
81. Gehring M: **Genomic imprinting: insights from plants**. *Annu Rev Genet* 2013, **47**:187-208.
82. Bushell C, Spielman M, Scott RJ: **The basis of natural and artificial postzygotic hybridization barriers in Arabidopsis species**. *Plant Cell* 2003, **15**(6):1430-1442.
83. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ: **Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila**. *Cell* 2007, **128**(6):1089-1103.
84. Blumenstiel JP, Hartl DL: **Evidence for maternally transmitted small interfering RNA in the repression of transposition in Drosophila virilis**. *Proc Natl Acad Sci U S A* 2005, **102**(44):15965-15970.
85. Schaefer M, Lyko F: **Lack of evidence for DNA methylation of Invader4 retroelements in Drosophila and implications for Dnmt2-mediated epigenetic regulation**. *Nature genetics* 2010, **42**(11):920-921; author reply 921.
86. Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G: **Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2**. *Nature genetics* 2009, **41**(6):696-702.
87. Ng DW, Lu J, Chen ZJ: **Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility**. *Curr Opin Plant Biol* 2012, **15**(2):154-161.
88. Groszmann M, Greaves IK, Fujimoto R, Peacock WJ, Dennis ES: **The role of epigenetics in hybrid vigour**. *Trends Genet* 2013, **29**(12):684-690.
89. Luo GZ, MacQueen A, Zheng G, Duan H, Dore LC, Lu Z, Liu J, Chen K, Jia G, Bergelson J *et al*: **Unique features of the m6A methylome in Arabidopsis thaliana**. *Nature communications* 2014, **5**:5630.
90. Kiani J, Grandjean V, Liebers R, Tuorto F, Ghanbarian H, Lyko F, Cuzin F, Rassoulzadegan M: **RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2**. *PLoS genetics* 2013, **9**(5):e1003498.
91. Sharma S, Yang J, Watzinger P, Kotter P, Entian KD: **Yeast Nop2 and Rcm1 methylate C2870 and C2278 of the 25S rRNA, respectively**. *Nucleic acids research* 2013, **41**(19):9062-9076.

92. Lioy VS, Goussard S, Guerineau V, Yoon EJ, Courvalin P, Galimand M, Grillot-Courvalin C: **Aminoglycoside resistance 16S rRNA methyltransferases block endogenous methylation, affect translation efficiency and fitness of the host**. *RNA* 2014, **20**(3):382-391.
93. Lange H, Sement FM, Gagliardi D: **MTR4, a putative RNA helicase and exosome co-factor, is required for proper rRNA biogenesis and development in Arabidopsis thaliana**. *The Plant journal : for cell and molecular biology* 2011, **68**(1):51-63.
94. Abbasi N, Kim HB, Park NI, Kim HS, Kim YK, Park YI, Choi SB: **APUM23, a nucleolar Puf domain protein, is involved in pre-ribosomal RNA processing and normal growth patterning in Arabidopsis**. *The Plant journal : for cell and molecular biology* 2010, **64**(6):960-976.

Supplementary Materials

The attached CD-ROM contains supplementary tables in excel for Chapter 2 and Chapter 3.

For Chapter 2, in Additional file 1, there are 7 supplementary tables (Table S1 – Table S2)

For Chapter 3, in Additional file 2, there are 9 supplementary tables (Supplementary Table 1 – Supplementary Table 9)

The following pages contain the supplementary figures for Chapter 2 and Chapter 3.

For Chapter 2, there are 4 supplementary figures (Figure S1 – Figure S4)

For Chapter 3, there are 10 supplementary figures (Supplementary Figure 1 – Supplementary Figure 10)

Chapter 2

Conservation of tRNA and rRNA 5-methylcytosine in the kingdom *Plantae*

Supplementary Figures (Figure S1 – Figure S4)

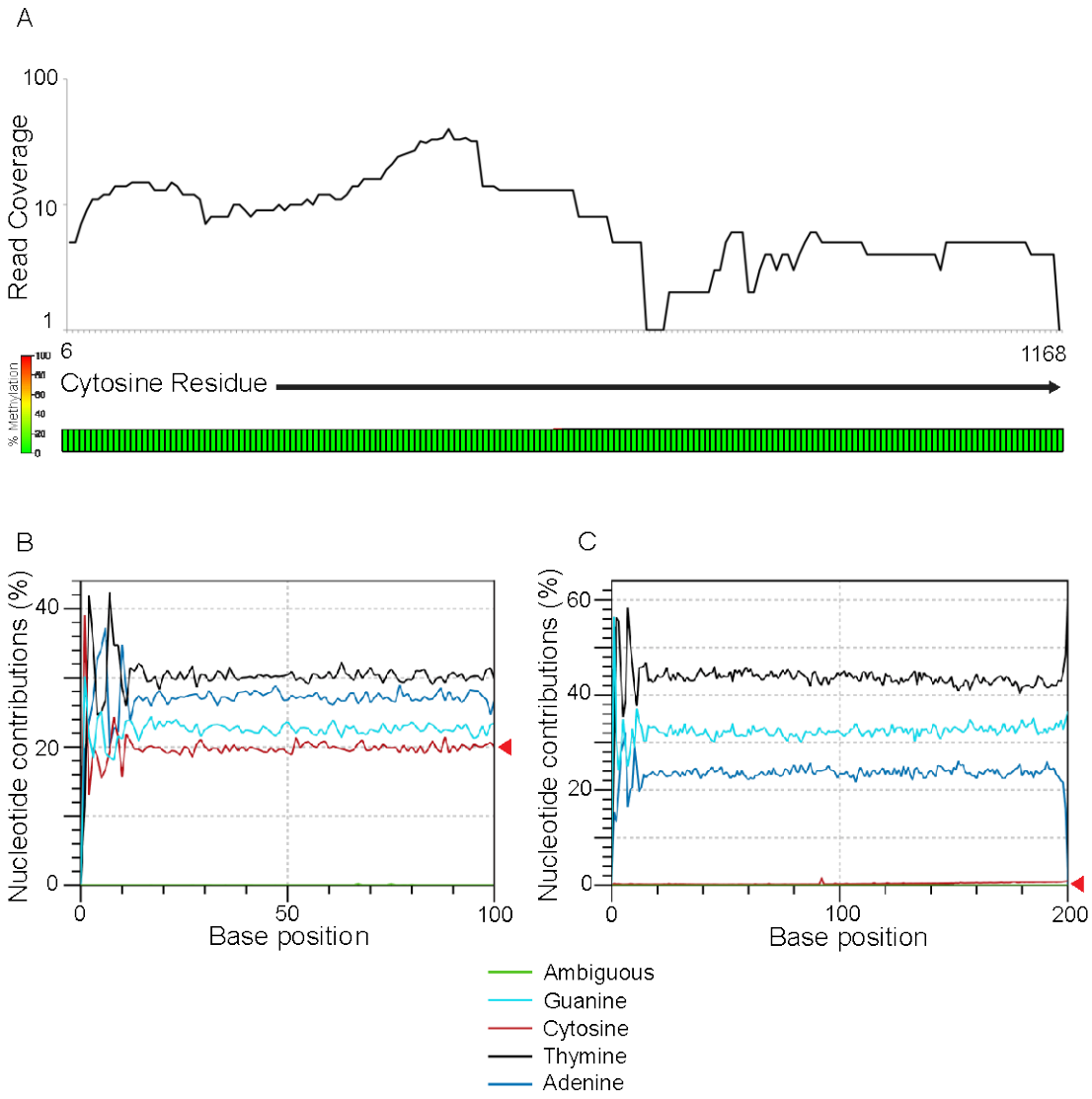


Figure S1. Efficient bisulfite conversion of non-methylated cytosine residues. (**A**) Above: Read coverage across cytosine residues from the Renilla Luciferase (R-Luc) mRNA *in vitro* transcribed BS conversion control. Below: Bisulfite conversion rates of cytosine residues in the R-Luc mRNA *in vitro* transcribed BS conversion control. Results shown are from a representative total RNA RBS-seq library. Nucleotide (A, T, C and G) contributions in representative total RNA RNA-seq (**B**) and in stranded RBS-seq libraries (**C**). The red triangles indicate the change in cytosine abundance.

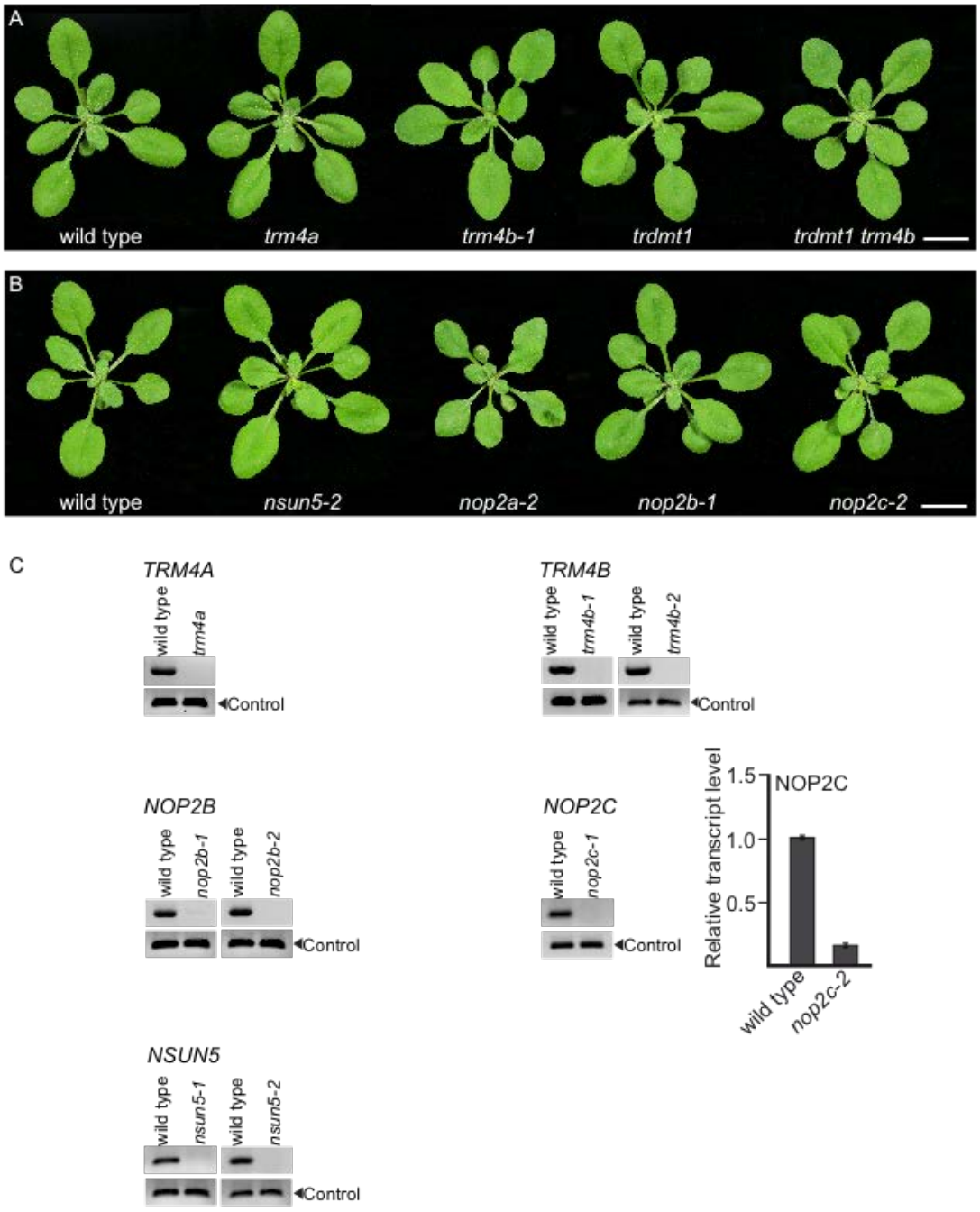


Figure S2. Characterization of *Arabidopsis thaliana* T-DNA mutants. Phenotypes of (**A**) tRNA and (**B**) rRNA RMTase mutants grown on soil. Photos are of 3 week old plants grown under long day photoperiods of 16 hr light and 8 hr darkness. Scale bar = 1 cm. (**C**) Semi-quantitative RT-PCR of *TRM4A*, *TRM4B*, *NOP2B*, *NOP2C* and *NSUN5* transcripts in wild type, *trm4a*, *trm4b-1*, *trm4b-2*, *nop2b-1, nop2b-2, nop2c-1*, *nsun5-1 and nsun5-2*. *PDF2A* is used as a loading control. Due to the low abundance of *NOP2C* mRNA transcripts, it was difficult to obtain a bright band in the quantitative range. Consequently, quantitative RT-PCR was performed for *nop2c-2.* Expression was normalized to *PDF2A* mRNA. Error bars represent standard error of 3 technical replicates.

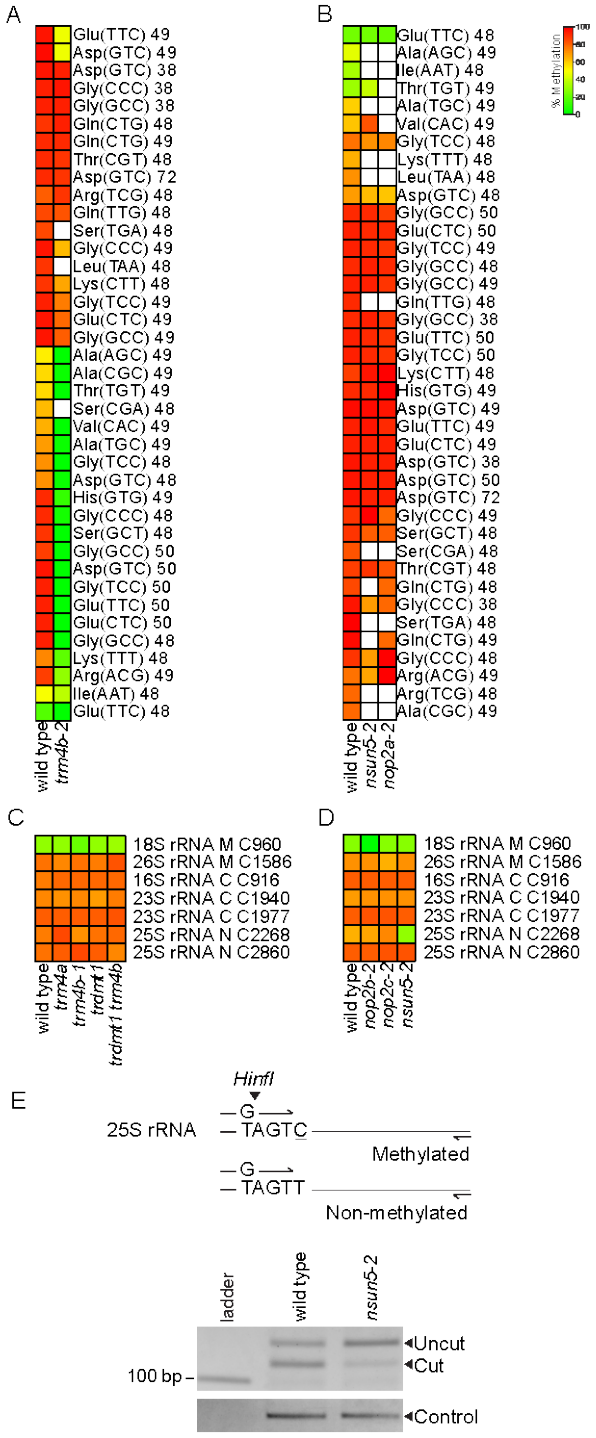


Figure S3. Specificity of tRNA and rRNA MTases. Heatmaps showing percentage methylation of all cytosines detected in nuclear tRNAs of wild type (n=1) and trm4b-2 (n=1) (**A**), and wild type (n=3 biological replicates), *nsun5-2* (n=1) and *nop2a-2/oli2-2* (n=1) (**B**). Cytosine positions are indicated next to tRNAs. White boxes represent cytosine positions with coverage less than five reads. Heatmaps showing percentage methylation of cytosines in nuclear (N), chloroplast (C) and mitochondrial (M) rRNA sequences in wild type, *trm4a*, *trm4b-1*, *trdmt1* and *trdmt1 trm4b* (n=1) (**C**) and in wild type (n=3), *nop2b-2* (n=1), *nop2c-2* (n=2) and *nsun5-2* (n=2) (**D**). Cytosine positions are indicated next to rRNAs. Analysis shown in (**A**), (**B**), (**C**) and (**D**) are from RBS-seq.

(**E**) Analysis of RNA methylation by NSUN5 at position C2268 on BS treated nuclear LSU 25S rRNA template. Above: Restriction maps of dCAPS amplified products showing the expected digest patterns of methylated and non-methylated template. The 25S_rRNA_F dCAPS primer contains a G mismatch at position four to generate a HinfI restriction site when C2268 is methylated. Below: Cleavage of PCR amplified product by HinfI confirms C2268 methylation in wild type as opposed to non-methylated C2268 in *nsun5-2* results in loss of HinfI restriction site. Loading control is undigested PCR product.

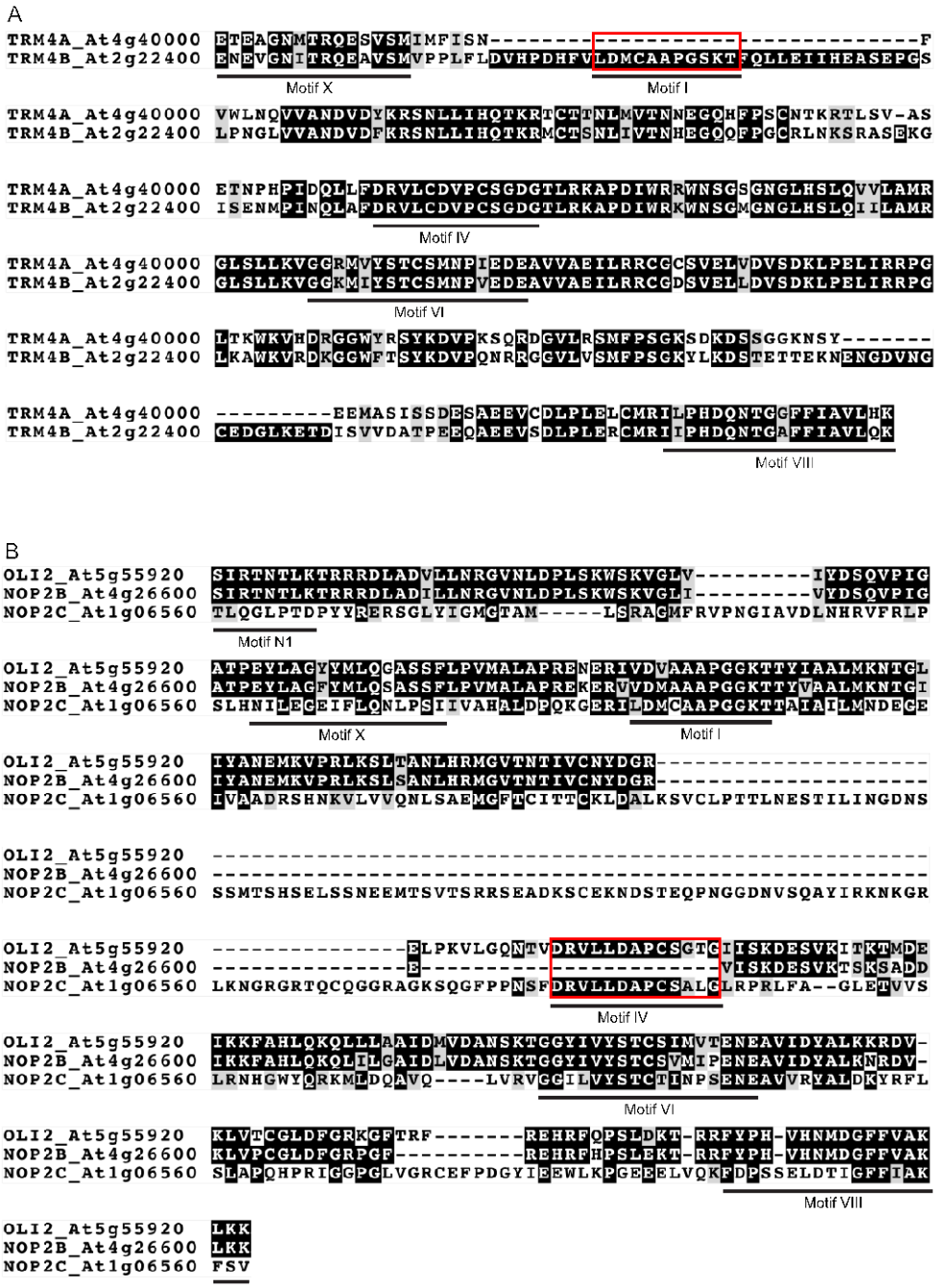
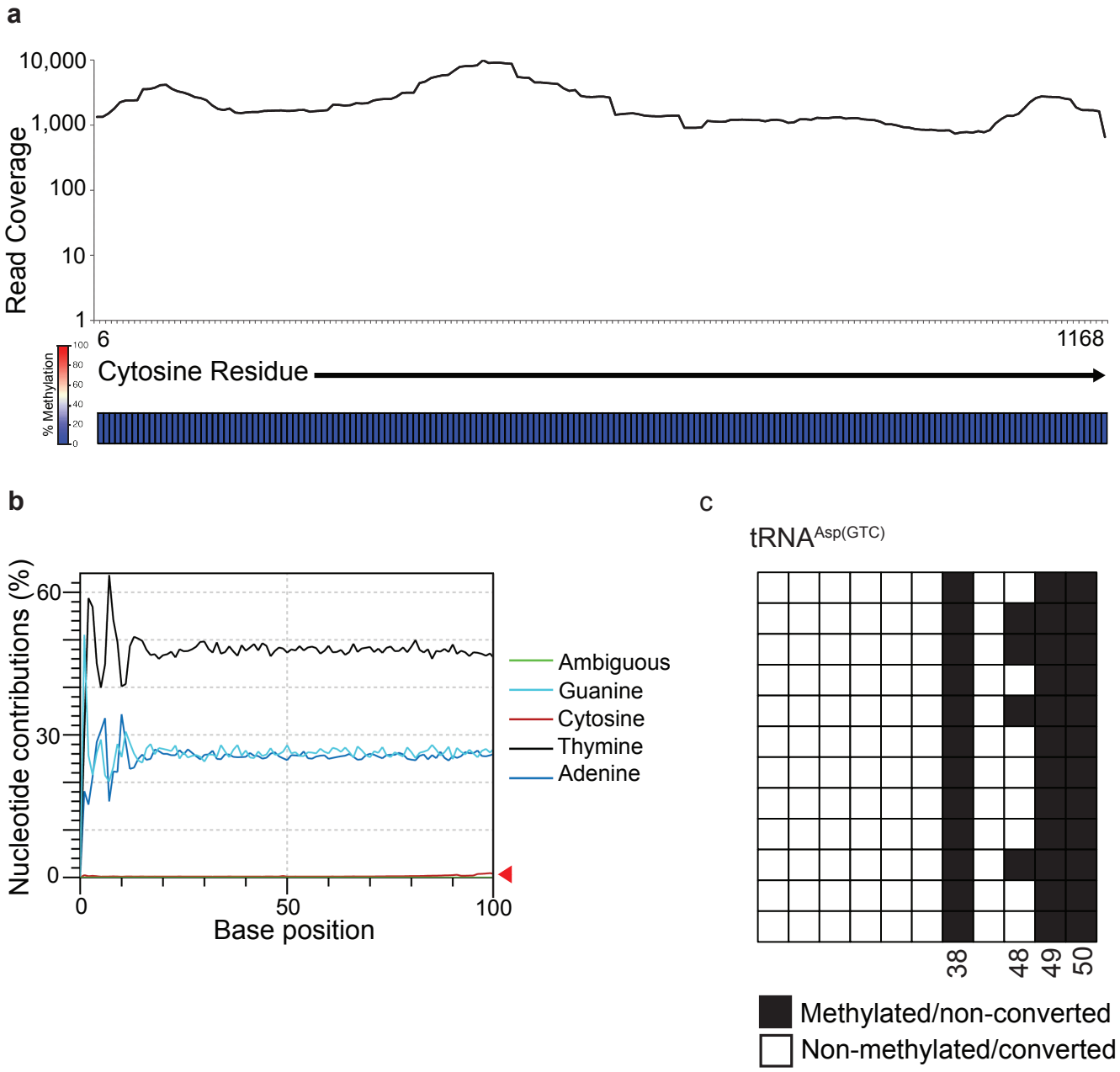


Figure S4. Multiple sequence alignment of methyltransferase motifs from *Arabidopsis thaliana* RMTases. The amino acid sequences of two subfamilies of RMTases in Arabidopsis; (**A**) TRM4A and TRM4B; (**B**) NOP2A/OLI2, NOP2B and NOP2C were aligned using Clustal Omega [72] and the MTase motifs were identified. Amino acids are shaded black if they are identical, and grey if they are similar in at least half of the aligned sequences. In panel (**A**), the red rectangle highlights a deletion of Motif-I in TRM4A, which has previously been shown to be critical for RMTase activity through AdoMet binding/catalysis. In panel (**B**), the red rectangle highlights a deletion in NOP2B, which corresponds to Motif-IV in NOP2A/OLI2 and NOP2C. Amino acid residues 86-362 and 161-487 are shown for TRM4A and TRM4B respectively. Amino acid residues 285-563, 268-520 and 218-595 are shown for NOP2A/OLI2, NOP2B and NOP2C respectively.

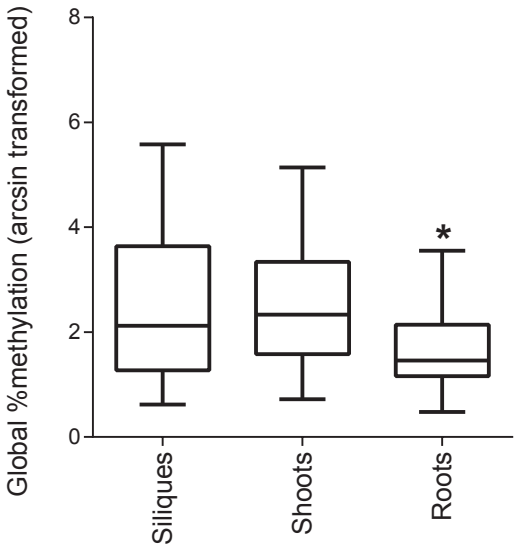
Chapter 3

Transcriptome-wide mapping of RNA 5-methylcytosine in *Arabidopsis* mRNAs and ncRNAS

Supplementary Figures (Supplementary Figure 1 – Supplementary Figure 10)

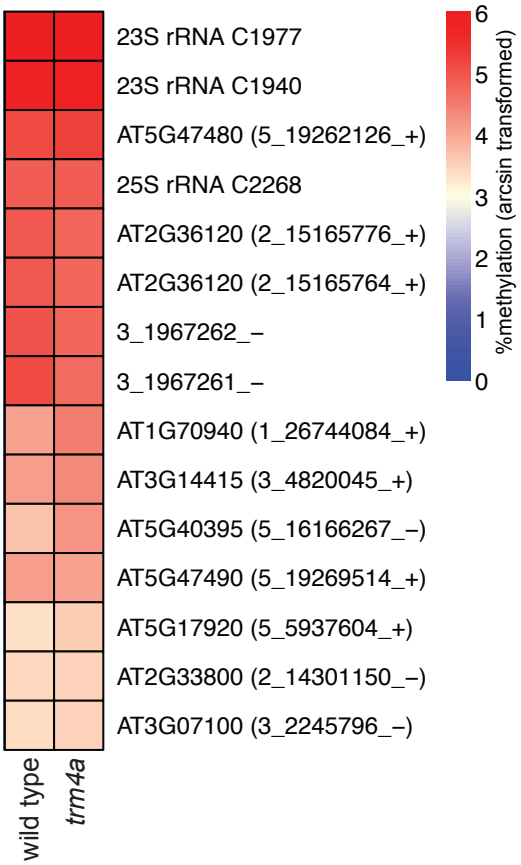


Supplementary Figure 1: Efficient bisulfite conversion of non-methylated cytosine residues. (**a**) Above: Read coverage across cytosine residues from the Renilla Luciferase (R-Luc) mRNA *in vitro* transcribed BS conversion control. Below: Bisulfite conversion rates of cytosine residues in the R-Luc mRNA *in vitro* transcribed BS conversion control. (**b**) Global endogenous cytosine abundance is reduced after bisulfite treatment. Results show nucleotide (A, T, C and G) contributions from the RBS-seq libraries of the two biological replicates of wild type seedling shoots RBS-seq libraries combined. The red triangle indicates the change in cytosine abundance to less than 0.5% after bisulfite conversion transcriptome-wide. (**c**) Endogenous $tRNA^{Asp(GTC)}$ was used as a positive control to confirm previously identified m^5C sites at cytosine positions C38, C48, C49 and C50 and to also confirm efficient bisulfite conversion of highly structured RNAs, such as tRNAs. Results shown are from 12 individual Sanger sequenced PCR clones. Each row represents an individual clone. Methylated cytosines are shaded black, while non-methylated cytosines are white.

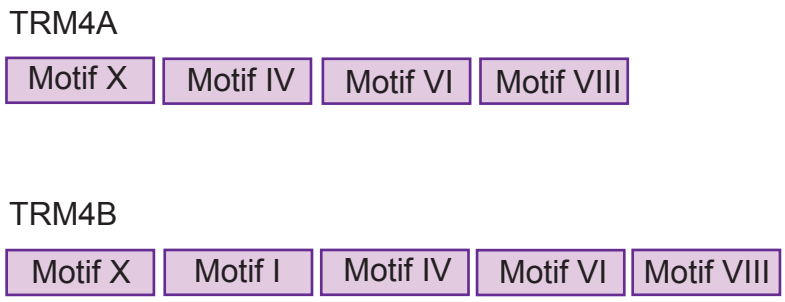


Supplementary Figure 2: Differences in global methylation levels of m^5C sites in *Arabidopsis* siliques, shoots and roots. Box-and-whisker plots (Tukey method) show the distributions of log2 arcsin transformed methylation percentages of transcriptome-wide m^5C sites called in *Arabidopsis* siliques, shoots and roots (FDR≤ 0.3, methylation ≥1%). The bar in the box-and-whisker plots represents the median value, while the box encompasses the range of data between the first and third quartile (interquartile range). The whiskers, or error bars, show the range of data points within 1.5 times the interquartile range. Differences between global methylation levels in the three tissue types were subjected to one way ANOVA and Tukey's multiple comparisons test (*P-value ≤0.0001).

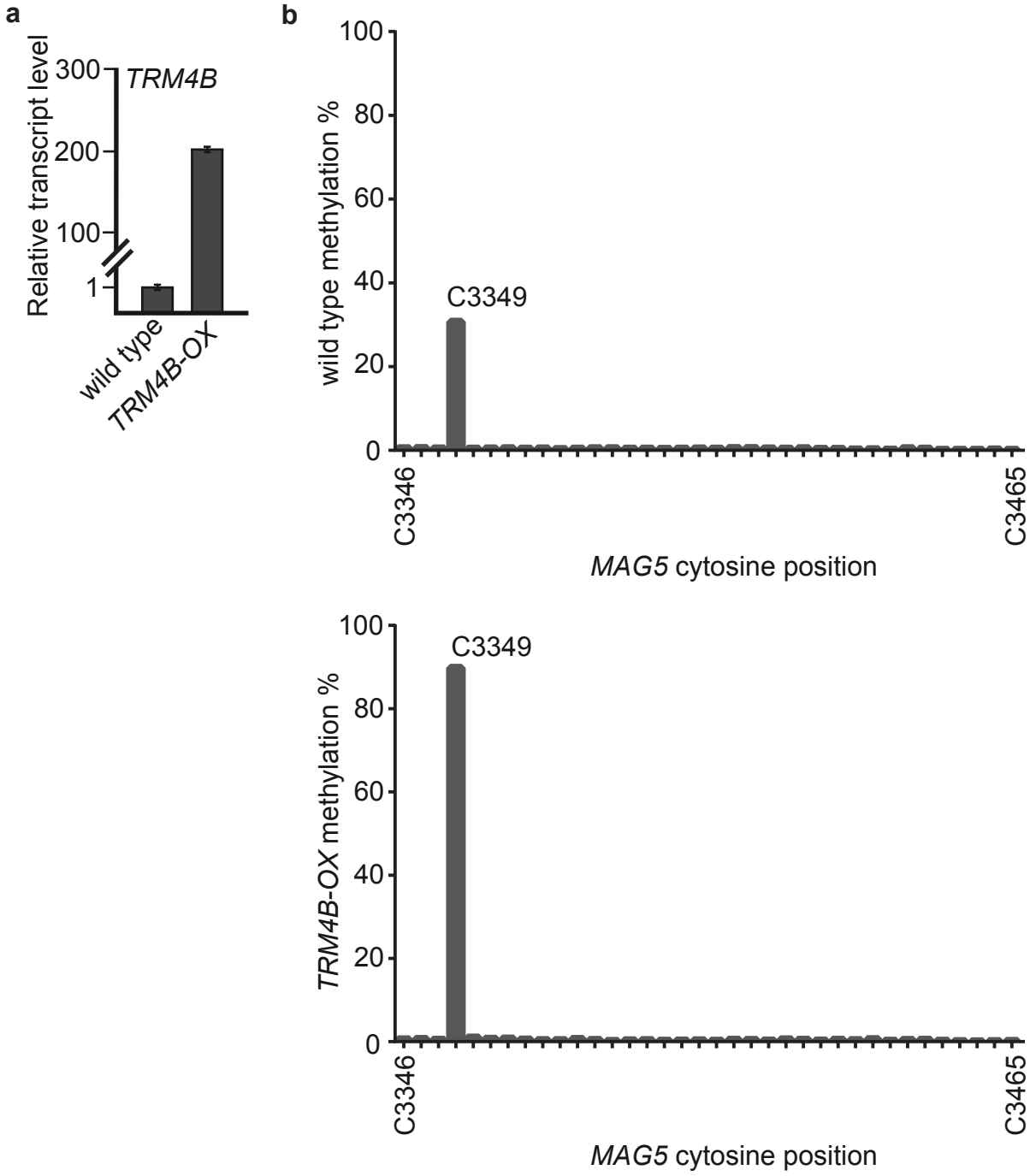
a



b

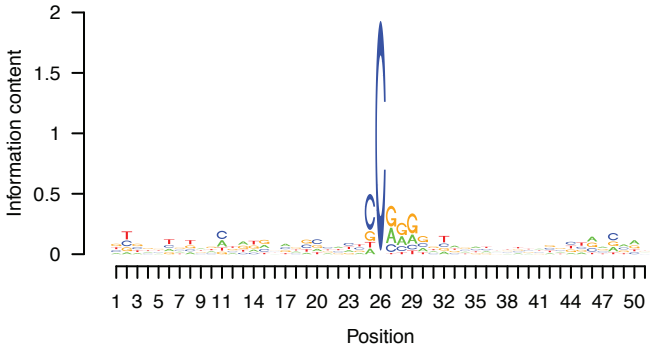


Supplementary Figure 3: Methylation of selected m^{5}C sites analyzed using RBS-amp-seq are not perturbed in *trm4a* mutants. (**a**) Heat map showing percentage methylation of cytosines derived from targeted RBS-amp-sequencing in wild type and trm4a seedling shoots. Cytosine positions are indicated on the left (3 biological replicates). (**b**) The amino acid sequences of *Arabidopsis* RNA 5-methylcytosine methyltransferases TRM4A and TRM4B were compared and the methyltransferase motifs were identified and are shown as labeled purple boxes in the schematic. TRM4A is missing Motif I, which has previously been shown to be critical for RMTase activity.

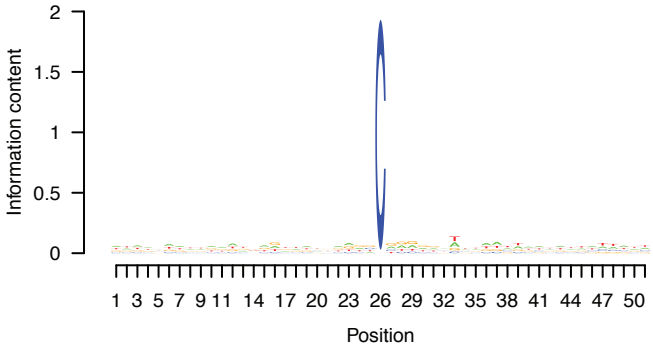


Supplementary Figure 4: qRT-PCR expression of *TRM4B-OX* line and BS conversion efficiency of RBS-amplicon-seq. (**a**) Quantitative RT-PCR of *TRM4B* transcripts in wild type and *TRM4B-OX*. Expression was normalized to *PDF2A* mRNA abundance. Error bars represent s.e. (n=3 technical replicates). (**b**) Methylation percentages across the 36 cytosine residues from the *MAG5* region amplified using RBS-amp-seq in wild type (above) and *TRM4B-OX* (below). The TRM4B dependent m^5C site at position C3349 shows increased methylation in plants overexpressing *TRM4B* compared to wildtype and no increases in methylation at neighboring cytosines.

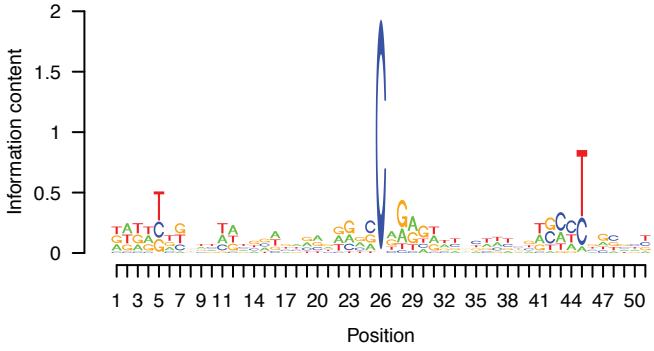
a



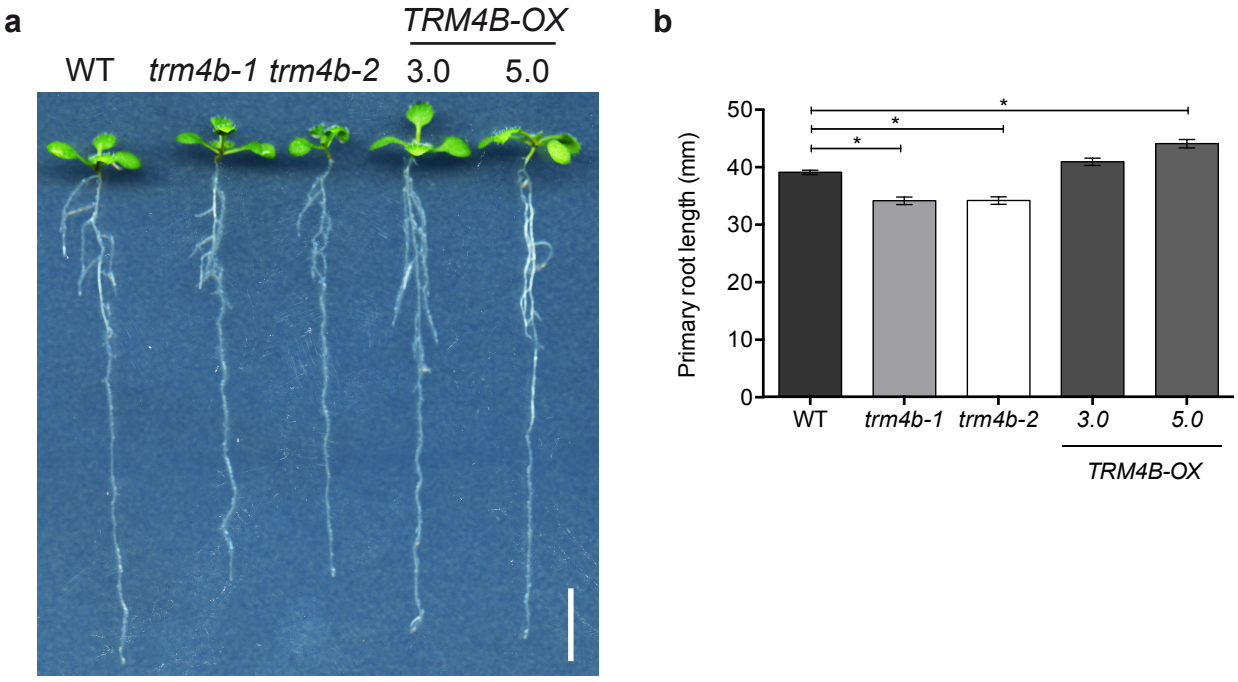
b



c

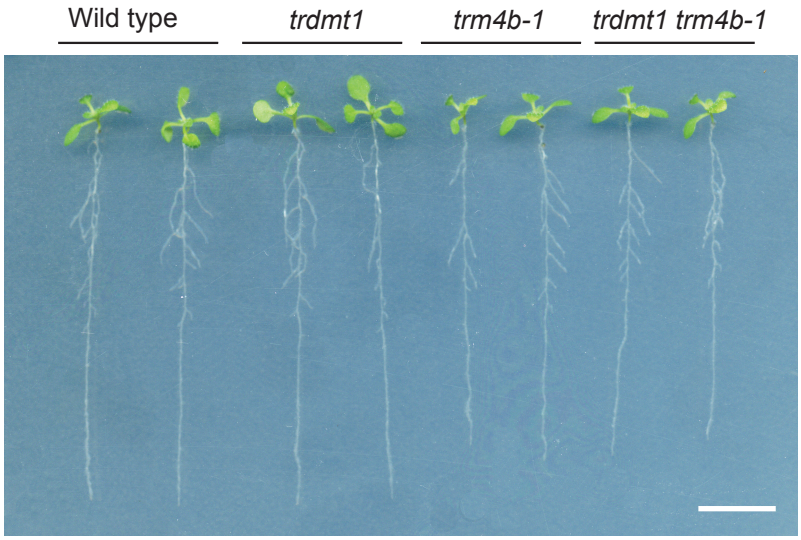


Supplementary Figure 5: LOGO motif analysis of TRM4B dependent m^{5}C sites. LOGO motif analysis of TRM4B dependent sites from *Arabidopsis* (**a**) siliques (**b**) shoots and (**c**) roots.

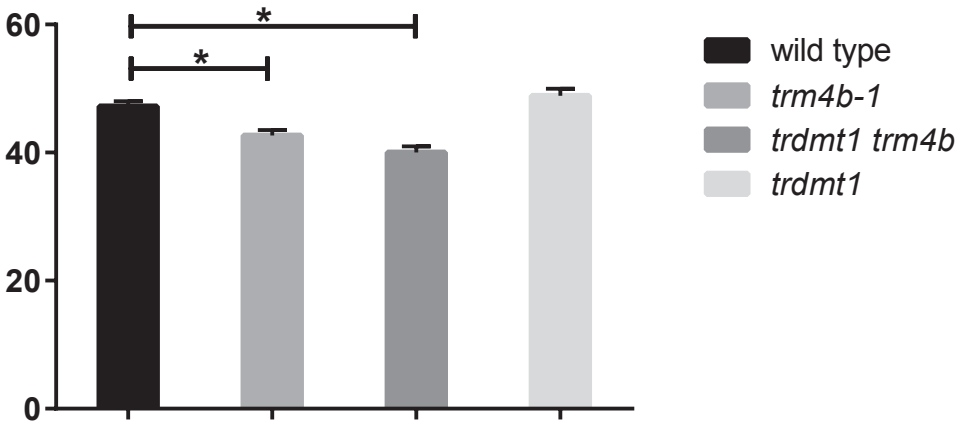


Supplementary Figure 6: Complementation of *trm4b* short-root phenotype. (**a**) Root elongation of seven day after germination seedlings of wild type (Col-0); *trm4b-1* and *trm4b-2* T-DNA mutants and two independent TRM4B over-expression transgenic lines; 3.0 and 5.0, grown on grown on half-strength MS medium with 1% sucrose. Scale bar, 1 cm. (**b**) Quantification of primary root length (7 DAG) of wild type, *trm4b* mutants and *TRM4B-OX* lines. (*P<0.001, Students t-test; n≥10-40 seedlings).

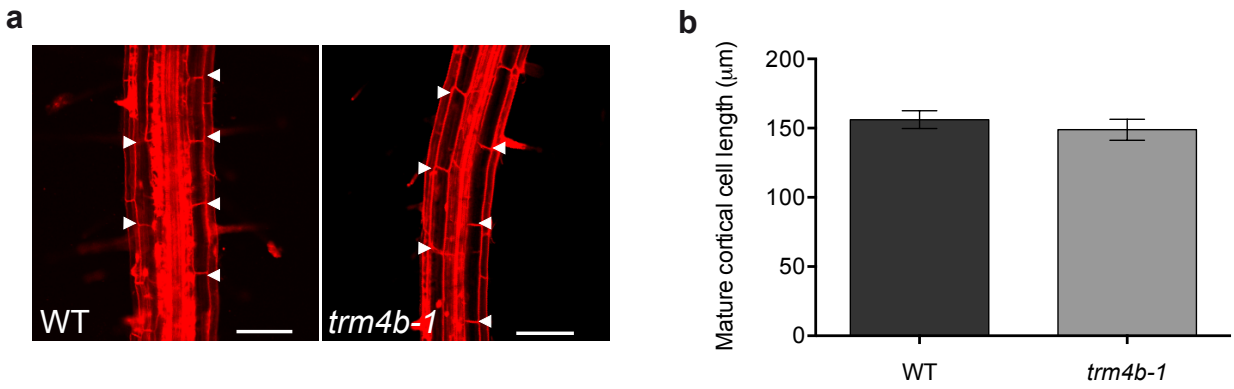
a



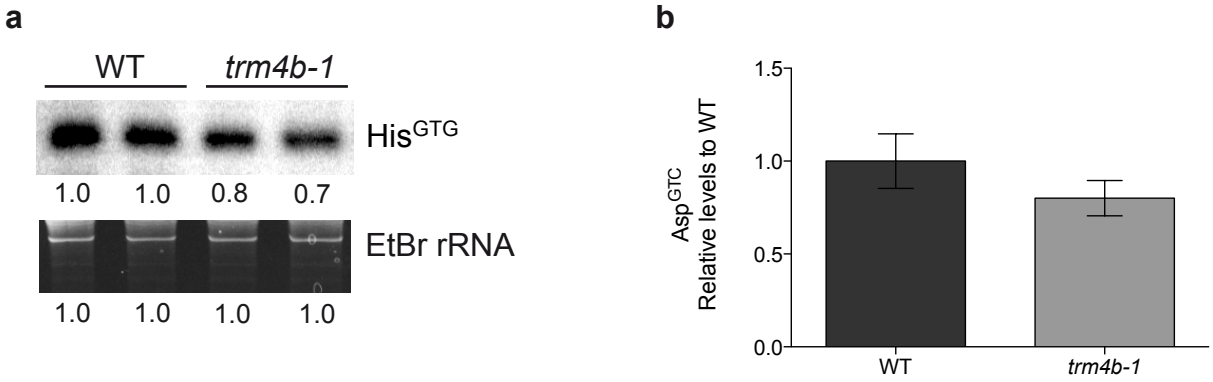
b



Supplementary Figure 7: Primary root length of *trdmt1* mutants is similar to wild type plants. (**a**) Root elongation of 10 days after germination (DAG) wild type, *trdmt1*, *trm4b-1* and *trdmt1 trm4b* double mutants, grown on half-strength MS media with 1% sucrose. Scale bar, 1 cm. (**b**) Primary root growth measurements at 10 DAG for wild type, *trm4b-1, trdmt1 trm4b* and *trdmt1* mutants. Error bars indicate s.e. (*P<0.05, Students t-test; n≥14 seedlings)



Supplementary Figure 8: Mature cortical cell length of wild type and *trm4b-1* mutant roots. (**a**) Propidium iodide stained confocal images of cortical cells in the root differentiatiation zone of wild type and *trm4b-1* mutant seedlings (7 DAG). White arrowheads indicate cortical cell wall boundaries. Scale bar, 100 μm. (**b**) Average mature root cortical cell length for wild type and *trm4b-1* mutant. Error bars indicate the s.e. (*P<0.05, Students t-test; n = 9 seedlings).



Supplementary Figure 9: Northern blot analysis of tRNA$^{His(GTG)}$. (**a**) Northern blot detection of TRM4B substrate tRNA$^{His(GTG)}$ in wild type and *trm4b-1*. Normalised intensity values are given beneath each lane. (b) Average signal intensity of tRNA$^{His(GTG)}$ in wild type and *trm4b-1*. The tRNA signal intensities were normalized to rRNA loading control and then compared to wild type.Error bars represent s.d; n=2 replicate RNA samples from one Northern blot experiment.


a
single−organism process
single−organism cellular process
response to stimulus
response to chemical
localization
single−organism metabolic process
single−organism biosynthetic process
cellular protein localization
response to nitrogen compound
small molecule metabolic process
cellular process
response to endogenous stimulus
immune system process
response to abiotic stimulus
primary metabolic process
innate immune response
response to oxidative stress
organic hydroxy compound biosynthetic process
cell differentiation
phosphorus metabolic process
response to external stimulus
organic hydroxy compound metabolic process
organic cyclic compound biosynthetic process
metabolic process
organic cyclic compound metabolic process
ncRNA processing
nitrogen compound transport
benzene−containing compound metabolic process
organophosphate biosynthetic process
response to ethylene
hydrogen peroxide catabolic process
biological regulation
ncRNA metabolic process
aromatic compound biosynthetic process
cellular aromatic compound metabolic process
response to biotic stimulus
regulation of biological process
organonitrogen compound metabolic process
monocarboxylic acid metabolic process
carbohydrate derivative metabolic process
response to external biotic stimulus
reactive oxygen species metabolic process
cell communication
regulation of response to stimulus
ion homeostasis
nitrogen compound metabolic process
oxidation−reduction process
multi−organism process
macromolecule modification
single−organism membrane organization
membrane organization
cellular metabolic process
response to carbohydrate
organic substance metabolic process
secondary metabolic process
RNA processing
response to light stimulus
lipid biosynthetic process
cellular component organization or biogenesis
rhythmic process
developmental process
lipid metabolic process
catabolic process
secondary metabolite biosynthetic process
carbohydrate derivative biosynthetic process
pyrimidine−containing compound metabolic process
pigment metabolic process
circadian rhythm
protein metabolic process
heterocycle metabolic process
cellular nitrogen compound metabolic process
oligopeptide transport
response to wounding
carbohydrate metabolic process
chlorophyll metabolic process
negative regulation of programmed cell death
response to fructose
developmental maturation
negative regulation of biological process
pigment biosynthetic process
photosynthesis, light reaction
signal transduction by phosphorylation
multicellular organismal process
cellular response to hypoxia
respiratory burst
positive regulation of metabolic process
MAPK cascade
cellular carbohydrate metabolic process
organic acid transport
cellular biogenic amine catabolic process
regulation of multicellular organismal process
flavonoid biosynthetic process
generation of precursor metabolites and energy
regulation of phosphate metabolic process
regulation of phosphorus metabolic process
dephosphorylation
positive regulation of biological process
plastid organization
positive regulation of catalytic activity
flavonoid metabolic process
RNA modification
regulation of protein modification process
response to transition metal nanoparticle
paraquat WT
vs control WT
50
40
30
20
10
-log10(p-value)
b
root epidermal cell differentiation
single−organism process
single−organism cellular process
response to stimulus
response to nitrogen compound
phenylpropanoid metabolic process
developmental process
response to nitrate
secondary metabolic process
localization
cell wall organization or biogenesis
anion transport
response to chemical
developmental maturation
single−organism metabolic process
cell wall organization
cellular response to iron ion starvation
protein localization to nucleus
nitrogen compound transport
pyrimidine−containing compound metabolic process
pyrimidine ribonucleotide biosynthetic process
respiratory burst
respiratory burst involved in defense response
cullin deneddylation
polysaccharide metabolic process
cellular localization
cellular process
response to red or far red light
oxidation−reduction process
cell communication
response to extracellular stimulus
cell junction organization
cell−cell junction organization
single−organism biosynthetic process
response to oxidative stress
control trm4b-1
vs control WT
16
14
12
10
8
6
4
-log10(p-value)

Supplementary Figure 10: Oxidative stress responsive biological process GO terms are constitutively activated in *trm4b* mutants. Whole roots at 6 DAG were harvested from wild type (WT), *trm4b-1* and paraquat treated wild type seedlings and RNA-seq analysis was performed (n=3). Biological process GO terms involved in the oxidative stress response are significantly enriched (FDR<0.05) in (**a**) paraquat treated wild type plants compared to wild type controls and in (**b**) *trm4b-1* mutants compared to wild type plants. Black stars indicate GO terms relating to oxidative stress responses. The heat map shows the significance level using the negative log of the p-value, where red = very significant and yellow = significant (FDR <0.05). Analyses were performed using the GOrilla package[64] and REVIGO[65] to create non-redundant GO term sets.