

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

Table of Contents.....	I
Abstract.....	III
Declaration.....	V
Acknowledgements.....	VII
Abbreviations.....	VIII
Chapter 1 Introduction	
Deciphering the epitranscriptome: a green perspective	1
Statement of Authorship.....	2
Chapter 2 Conservation of tRNA and rRNA 5-methylcytosine in the kingdom <i>Plantae</i>	3
Statement of Authorship.....	4
Chapter 3 Transcriptome-wide mapping of RNA 5-methylcytosine in <i>Arabidopsis</i> mRNAs and ncRNAs.....	6
Statement of Authorship.....	7
Chapter 4 Discussion, Conclusions and Future Directions.....	51
New Functions for Old RNA Modifications.....	60
Dynamic, Reversible Regulation of RNA Modifications – ‘Yin and Yang’	60
RNA Modifications - Instruments of War.....	63
RNA Modifications in Parental Conflict and Hybrid Vigour.....	64
Conclusions and Future Directions.....	66
References.....	67

Supplementary Materials.....	73
 Chapter 2 Supplementary Figures.....	74
 Chapter 3 Supplementary Figures.....	80

Abstract

Post-transcriptional methylation of RNA cytosine residues to 5-methylcytosine (m⁵C) 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 m⁵C sites has been undertaken in plants. Here, we present over a thousand m⁵C 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 m⁵C methylation sites can be tissue-specific, or shared among the tissue types investigated. Among the shared m⁵C 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 m⁵C sites are located within coding sequences. A small, significant enrichment of m⁵C 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 m⁵C sites across six species in the kingdom *Plantae*, suggesting important and highly conserved roles of this post-transcriptional modification.

We identified over 100 m⁵C 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 m⁵C sites in tRNAs mediated by Transfer RNA aspartic acid methyltransferase 1 (TRDMT1) and find no evidence of m⁵C 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* m⁵C 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

Copyright: © 2016 The Authors. *Journal of Integrative Plant Biology* Published by John Wiley & Sons Australia, Ltd on behalf of Institute of Botany, Chinese Academy of Sciences. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

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

Copyright: © 2015 Burgess et al. Open Access. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

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

Ca⁵C: 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
F₁: First Generation Hybrid
f⁵C: 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
H₂O₂: Hydrogen Peroxide
HAMR: High-throughput Annotation of Modified Ribonucleotides
HeLa: Henrietta Lacks Human Cervical Cancer Cells
Hen-16: Henriksjfall-16
hm⁵C: 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
i⁶A: N⁶-Isopentenyladenosylation

IGV: Integrative Genomics viewer
lncRNA: Long Non-coding RNA
m¹G: 1-Methylguanosine
m²G: 2-Methylguanosine
m²⁶A: N-6 Dimethylation
m³C: 3-Methylcytosine
m³T: 3-Methylthymidine
m³U: 3-Methyluridine
m⁴C: N4-Methylcytosine
m⁴Cm: N4, 2'-O-Dimethylcytosine
m⁵C: 5-Methylcytosine
m⁶A: N⁶-Methyladenosine
m⁷G: 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
t⁶A: 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